SUPERINTENDENCIA DE INDUSTRIA Y COMERCIO

BT

DELEGATURA PROPIEDAD INDUSTRIAL

División de Nuevas Creaciones

PCT

SOLICITUD

PATENTE DE INVENCION

ENTRADA EN FASE NACIONAL

(21) **EXPEDIENTE No.** 04 117741

(54) **TITULO** POLIPEPTIDOS Y RUTAS BIOSINTETICAS

(51) **CLASIFICACION INTERNACIONAL**

(71) **SOLICITANTE** CARGILL, INCORPORATED

DOMICILIO MINNESOTA, 55440-5624, ESTADOS UNIDOS DE AMERICA

(74) **APODERADO** JUAN PABLO CADENA SARMIENTO

(22) **SANTAFE DE BOGOTA, D.C.**

PETITORIO

1

AS: 95.162

Industria y Comercio
SUPERINTENDENCIA

SUPERINTENDENCIA Y COMERCIO
Radicacion : 04117741 Folios: 385
Fecha (AMD): 2004-11-23 17:12:51
Tramite : 011 PCT-PATENT 379 FASE NACI 411 PRESENTA
Dependencia: 2020 DIVISION DE NUEVAS CREACIONES

FORMULARIO ÚNICO DE SOLICITUD DE PATENTE

PCT

ENTRADA EN FASE NACIONAL

1 SOLICITUD DE:

☒ Patente de Invención. ☐ Patente de Modelo de Utilidad.

☐ Capítulo I. ☒ Capítulo II.

2 SOLICITANTE (71)	Nombre: CARGILL, INCORPORATED Dirección: P.O. Box 5624, Minneapolis, Minnesota 55440-5624, Estados Unidos de América Nacionalidad o Domicilio: Estados Unidos de América Lugar de Constitución: Estados Unidos de América Teléfono: Fax: E-mail	IDENTIFICACION C.C. ☐ NIT ☐ C.E. ☐ Otro ☐ Cual____ Número ____
3 REPRESENTANTE O APODERADO	Nombre: JUAN PABLO CADENA SARMIENTO Dirección: Cra. 7 No. 16-56 Piso 9 Teléfono: 380-2200 Fax: 380-2700 E-mail:	IDENTIFICACION C.C. ☒ NIT ☐ C.E. ☐ Otro ☐ Cual____ Número 80.410.337 TP 57492
4 INVENTOR (ES) (72)	Nombre: Paula M. Hicks, Sara C. Mcfarlan, Timothy W. Abraham, Russell J. Hobson, Douglas C. Cameron, Joseph Dalluge, Jim Mills, John Rosazza, Lishan Zhao, David Paul Weiner Dirección: 11601 Tanglewood Drive, Eden Prairie, Minnesota 55347, US; 2151 Bayard Ave. St. Paul, Minnesota 55116, US; 18507 Apple Tree Court, Minnetonka, Minnesota 55345, US; 316 Vinewood Lane North, Plymouth, Minnesota, 55441, US; 3590 Ranier Lane North Plymouth, Minnesota 55447, US; 3516 County Road 30 S.W., Waverly, Minnesota 55390, US; 2360 Yuma Lane North, Plymouth, Minnesota 55447, US; 453 Linder Road N.E. Iowa City, Iowa 52240, US; 6564 Red Knot Street Carlsbad, California 92009, US; 13416 Portofino Drive Del Mar, California 92014, Estados Unidos de América (Respectivamente) Nacionalidad o Domicilio: Estados Unidos de América (Respectivamente)	

5 PCT (86)
Solicitud Internacional No. PCT/US03/12588 Fecha 23-04-2003
Publicación Internacional No. WO 03/091396 A2 Fecha 06-11-2003

Título (54): POLIPÉPTIDOS Y RUTAS BIOSINTÉTICAS

7 Clasificación Internacional (51) C07D 209/10 , C12P 17/10

8 Prioridad SI ☒ NO ☐	(33) País de Origen	(32) Fecha	(31) Número de Solicitud
	US	23-04-2002	60/374.831

9 Comprobante de pago No. 04-564780 **Fecha** 23-11-2004

Instrucciones para el diligenciamiento del presente formulario, ver reverso de esta página

2

10 ANEXOS

- [x] Comprobante de pago de la tasa de presentación de la solicitud. $422.000
- [ ] Documento que acredita la existencia y representación legal cuando el solicitante sea persona jurídica.
- [ ] Poderes, si fuere el caso.
- [ ] Documento de cesión del inventor al solicitante o a su causante.
- [ ] Declaraciones
- [x] Copia de la solicitud internacional. (Resumen, descripción, reivindicaciones, dibujos)
- [x] Traducción de la solicitud internacional al castellano. (Resumen, descripción, reivindicaciones, dibujos)
- [ ] Copia del examen preliminar efectuado por la IPEA.
- [ ] Traducción del examen preliminar efectuado por la IPEA.
- [ ] De ser el caso, copia del contrato de acceso.
- [ ] De ser el caso, documento que acredite la licencia o autorización de uso de conocimientos tradicionales de las comunidades indígenas.
- [ ] De ser el caso, certificado de depósito del material biológico.
- [ ] Arte final 12x12.
- [ ] De ser el caso, información sobre otras solicitudes de patente o títulos obtenidos en el extranjero por el mismo titular o su causante, relacionadas parcial o totalmente con la invención de esta solicitud.
- [x] Otros: Tarjeta de propietarios.

11 FIGURA CARACTERÍSTICA:

12 Solicito la concesión de la patente

NOMBRE: JUAN PABLO SALDEMA SARMIENTO

FIRMA:

C.C. 80.418.637 de Usaquén T.P. 57492 del C.S.J.

Instrucciones para la presentación de los anexos, ver reverso de esta página.

SUPERINDUSTRIA Y COMERCIO

Radicacion : 04117741 03023080 Folios: 386

Fecha (AMD): 2004-11-23 17:12:51

Tramite : 811 PCT-PATENT 379 FASE NACI 411 PRESENTA

Dependencia: 2220 DIVISION DE NUEVAS CREACIONES

NIT : 900176099-2

RECIBO OFICIAL DE CAJA : 04 - 55,786

FECHA : AGOSTO 24 DE 2004

C O N S I G N A C I O N

DEPOSITANTE	TIPO PAGO	BANCO	CUENTA	No. PAGO	FECHA PAG	Vr. PAGO
BRIGARD & CASTRO LTDA	CONSIGNACION	BANCO POPULAR	050-00110-6	12685625	24/08/2004	4,220,000.00

C O N C E P T O

CANT.	RENTISTICO	CONCEPTO	TOTAL CONCEPTO
1	50005-01-01 SOLICITUDES	1 TRAMITES DE SOL. DE PATENTE DE IN	422,000.00
		TOTAL : $	422,000.00

SON: CUATROCIENTOS VEINTIDOS MIL PESOS

RESPONSABLE : _____

RECIBO DE CAJA APLICADO AL EXPEDIENTE No._____

AS: 95.162

4

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
6 November 2003 (06.11.2003)

PCT

(10) International Publication Number
WO 03/091396 A2

(51) International Patent Classification⁷: C12N

(21) International Application Number: PCT/US03/12588

(22) International Filing Date: 23 April 2003 (23.04.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/374,831 23 April 2002 (23.04.2002) US

(71) Applicant: CARGILL, INCORPORATED [US/US]; P.O. Box 5624, Minneapolis, MN 55440-5624 (US).

(72) Inventors: ABRAHAM, Timothy, W.; 18507 Apple Tree Court, Minnetonka, MN 55345 (US). CAMERON, Douglas, C.; 3590 Ranier Lane, N., Plymouth, MN 55447 (US). DALLUGE, Joseph; 3516 County Road 30 S.W., Waverly, MN 55390 (US). HICKS, Paula, M.; 11601 Timglewood Drive, Eden Prairie, MN 55347 (US). HOBSON, Russell, J.; 316 Vinewood Lane North, Plymouth, MN 55441 (US). MCFARLAN, Sara, C.; 2151 Bayard Avenue, St Paul, MN 55116 (US). MILLIS, Jim; 2360 Yuma Lane North, Plymouth, MN 55447 (US). ROSAZZA, John; 453 Linder Road N.E., Iowa City, IA 52240 (US).

(74) Agent: SKELTON, Jeffrey, J.; Cargill, Incorporated, P.O. Box 5624, Minneapolis, MN 55440-5624 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 03/091396 A2

(54) Title: POLYPEPTIDES AND BIOSYNTHETIC PATHWAYS

(57) Abstract: Methods and compositions that can be used to make monatin from glucose, tryptophan, indole-3-lactic acid, indole-3-pyruvate, and 2-hydroxy 2-(iridol-3-ylmethyl)-4-keto glutaric acid, are provided. Methods are also disclosed for producing the indole-3-pyruvate and 2-hydroxy 2-(indol-3ylmethyl)-4-keto glutaric acid intermediates. Compositions provided include nucleic acid molecules, polypeptides, chemical structures, and cells. Methods include *in vitro* and *in vivo* processes, and the *in vitro* methods include chemical reactions.

POLYPEPTIDES AND BIOSYNTHETIC PATHWAYS

CROSS-REFERENCE TO RELATED APPLICATION

This application claims priority to U.S. Patent Application 60/374,831 filed April 23, 2002.

FIELD

This disclosure provides polypeptides and biosynthetic pathways that are useful in the production of indole-3-pyruvate, 2-hydroxy 2-(indol-3ylmethyl)-4-keto glutaric acid (MP) and/or monatin.

BACKGROUND

Indole pyruvate.

Indole-3-pyruvate is a strong antioxidant that is believed to counter act oxidative stress in tissues with high oxygen concentrations (Politi *et al.* "Recent advances in Tryptophan Research", edited by G. A. Filippini *et al.* Plenum Press, New York, 1996, pp 291-8). Indole pyruvate also is an intermediate in a pathway to produce indole-acetic acid (IAA), the primary plant growth hormone auxin (diffusible growth promoting factor). IAA is active in submicrogram amounts in a range of physiological processes including apical dominance, tropisms, shoot elongation, induction of cambial cell division, and root initiation. Synthetic auxins are used in horticulture to induce rooting and to promote the set and development of fruit. At high concentrations the synthetic auxins are effective herbicides against broad-leafed plants. Natural auxins produced by fermentation may be considered more environmentally friendly than chemically produced herbicides. Growth regulators had world sales in 1999 of 0.4 billion pounds (1.4 billion U.S. dollars).

Some examples of patents on indole acetic acid and derivatives thereof include: US Patent No. 5,843,782 Micropropagation of rose plants, auxin used in culture medium and US 5,952,231 Patent No. Micropropagation of rose plants.

In addition to plant related utilities, indole acetic acid is useful in pharmaceutical applications. For example, US Patent No. 5,173,497 "Method of preparing alpha-oxopyrrolo[2,3-B]indole acetic acids and derivatives" proposes the use of these compounds in the treatment of memory impairment such as that associated with Alzheimer's disease and senile dementia. The mechanism proposed in US Patent No. 5,173,497 is that these compounds inhibit the polypeptide acetylcholinesterase and increase acetylcholine levels in the brain.

Indole-3-carbinol is produced from indole-3-acetic acid by peroxidase-catalyzed oxidation, and can easily be converted into diindolylmethane. Both compounds are reported to eliminate toxins and promote the production of hormones beneficial to women's health.

Tryptophan Derivatives

Chlorinated D-tryptophan has been identified as a nonnutritive sweetener, and there is increasing interest in pursuing other derivatives as well. Monatin is a natural sweetener that is similar

in composition to the amino acid tryptophan. It can be extracted from the bark of the roots of the South African shrub, *Sclerochiton ilicifolius*, and has promise in the food and beverage industry as a high-intensity sweetener. Some examples of patents on monatin include: US Patent No. 5994559 Synthesis of monatin-A high intensity natural sweetener, US Patent No. 4975298 3-(1-amino-1,3-dicarboxy-3-hydroxy-but-4-yl)-indole compounds, US Patent No. 5128164 Composition for human consumption containing 3-(1-amino-1,3-dicarboxy-3-hydroxy-but-4-yl)-indole compounds; and US Patent No. 5128482 Process for the production of 3-1(1-amino-1,3-dicarboxy-3-hydroxy-but-4-yl) indole.

Some of the precursors of monatin described here can also be useful as synthetic sweeteners or as intermediates in the synthesis of monatin derivatives.

SUMMARY

The disclosure provides several biosynthetic routes for making monatin from glucose, tryptophan, indole-3-lactic acid, and/or through monatin precursors such as indole-3-pyruvate and 2-hydroxy 2-(indole-3-ylmethyl)-4-keto glutaric acid. Polypeptides and nucleic acid sequences that can be used to make monatin, indole-3-pyruvate, and 2-hydroxy 2-(indole-3-ylmethyl)-4-keto glutaric acid are disclosed.

Monatin can be produced through indole-3-pyruvate, 2-hydroxy 2-(indole-3-ylmethyl)-4-keto glutaric acid (monatin precursor, MP, the alpha-keto form of monatin), indole-3-lactic acid, tryptophan, and/or glucose (FIG. 1). Methods of producing or making monatin or its intermediates shown in FIGS. 1-3 and 11-13 that involve converting a substrate to a first product, and then converting the first product to a second product, and so on, until the desired end product is created, are disclosed.

FIGS. 1-3 and 11-13 show potential intermediate products and end products in boxes. For example, a conversion from one product to another, such as glucose to tryptophan, tryptophan to indole-3-pyruvate, indole-3-pyruvate to MP, MP to monatin, or indole-3-lactic acid (indole-lactate) to indole-3-pyruvate, can be performed by using these methods. These conversions can be facilitated chemically or biologically. The term "convert" refers to the use of either chemical means or polypeptides in a reaction which changes a first intermediate to a second intermediate. The term "chemical conversion" refers to reactions that are not actively facilitated by polypeptides. The term "biological conversion" refers to reactions that are actively facilitated by polypeptides. Conversions can take place *in vivo* or *in vitro*. When biological conversions are used the polypeptides and/or cells can be immobilized on supports such as by chemical attachment on polymer supports. The conversion can be accomplished using any reactor known to one of ordinary skill in the art, for example in a batch or a continuous reactor.

Methods are also provided that include contacting a first polypeptide with a substrate and making a first product, and then contacting the first product created with a second polypeptide and creating a second product, and then contacting the second product created with a third polypeptide

and creating a third product, for example monatin. The polypeptides used and the products produced are shown in FIGS. 1-3 and 11-13.

Polypeptides, and their coding sequences, that can be used to perform the conversions shown in FIGS. 1-3 and 11-13 are disclosed. In some examples, polypeptides having one or more point mutations that allow the substrate specificity and/or activity of the polypeptides to be modified, are used to make monatin.

Isolated and recombinant cells that produce monatin are disclosed. These cells can be any cell, such as a plant, animal, bacterial, yeast, algal, archaeal, or fungal cell.

In a particular example, the disclosed cells include one or more of the following activities, for example two or more or three or more of the following activities: tryptophan aminotransferase (EC 2.6.1.27), tyrosine (aromatic) aminotransferase (EC 2.6.1.5), multiple substrate aminotransferase (EC 2.6.1.-), aspartate aminotransferase (EC 2.6.1.1), tryptophan dehydrogenase (EC 1.4.1.19), tryptophan-phenylpyruvate transaminase (EC 2.6.1.28), L-amino acid oxidase (EC 1.4.3.2), tryptophan oxidase (no EC number, Hadar *et al.*, *J. Bacteriol* 125:1096-1104, 1976 and Furuya *et al.*, *Biosci Biotechnol Biochem* 64:1486-93, 2000), D-amino acid dehydrogenase (EC 1.4.99.1), D-amino acid oxidase (EC 1.4.3.3), D-alanine aminotransferase (EC 2.6.1.21), synthase/lyase (EC 4.1.3.-), such as 4-hydroxy-4-methyl-2-oxoglutarate aldolase (EC 4.1.3.17) or 4-hydroxy-2-oxoglutarate aldolase (EC 4.1.3.16), synthase/lyase (4.1.2.-), D-tryptophan aminotransferase (Kohiba and Mito, Proceedings of the 8th International Symposium on Vitamin B_6 and Carbonyl Catalysis, Osaka, Japan 1990), phenylalanine dehydrogenase (EC 1.4.1.20) and/or glutamate dehydrogenase (EC 1.4.1.2, 1.4.1.3, 1.4.1.4).

In another example, cells include one or more, for example two or more, or three or more, of the following activities: indolelactate dehydrogenase (EC 1.1.1.110), R-4-hydroxyphenyllactate dehydrogenase (EC 1.1.1.222), 3-(4)-hydroxyphenylpyruvate reductase (EC 1.1.1.237), lactate dehydrogenase (EC 1.1.1.27, 1.1.1.28, 1.1.2.3), (3-imidazol-5-yl) lactate dehydrogenase (EC 1.1.1.111), lactate oxidase (EC 1.1.3.-), synthase/lyase (4.1.3.-) such as 4-hydroxy-4-methyl-2-oxoglutarate aldolase (EC 4.1.3.17) or 4-hydroxy-2-oxoglutarate aldolase (EC 4.1.3.16), synthase/lyase (4.1.2.-), tryptophan dehydrogenase (EC 1.4.1.19), tryptophan-phenylpyruvate transaminase (EC 2.6.1.28), tryptophan aminotransferase (EC 2.6.1.27), tyrosine (aromatic) aminotransferase (EC 2.6.1.5), multiple substrate aminotransferase (EC 2.6.1.-), aspartate aminotransferase (EC 2.6.1.1), phenylalanine dehydrogenase (EC 1.4.1.20), glutamate dehydrogenase (EC 1.4.1.2, 1.4.1.3, 1.4.1.4), D-amino acid dehydrogenase (EC 1.4.99.1), D-tryptophan aminotransferase, and/or D-alanine aminotransferase (EC 2.6.1.21).

In addition, the disclosed cells can include one or more of the following activities, for example two or more or three or more of the following activities: tryptophan aminotransferase (EC 2.6.1.27), tyrosine (aromatic) aminotransferase (EC 2.6.1.5), multiple substrate aminotransferase (EC 2.6.1.-), aspartate aminotransferase (EC 2.6.1.1), tryptophan dehydrogenase (EC 1.4.1.19), tryptophan-phenylpyruvate transaminase (EC 2.6.1.28), L-amino acid oxidase (EC 1.4.3.2), tryptophan oxidase (no EC number), D-amino acid dehydrogenase (EC 1.4.99.1), D-amino acid

oxidase (EC 1.4.3.3), D-alanine aminotransferase (EC 2.6.1.21), indolelactate dehydrogenase (EC 1.1.1.110), R-4-hydroxyphenyllactate dehydrogenase (EC 1.1.1.222), 3-(4)-hydroxyphenylpyruvate reductase (EC 1.1.1.237), lactate dehydrogenase (EC 1.1.1.27, 1.1.1.28, 1.1.2.3), (3-imidazol-5-yl) lactate dehydrogenase (EC 1.1.1.111), lactate oxidase (EC 1.1.3.-), synthase/lyase (4.1.3.-) such as 4-hydroxy-4-methyl-2-oxoglutarate aldolase (EC 4.1.3.17) or 4-hydroxy-2-oxoglutarate aldolase (EC 4.1.3.16), synthase/lyase (4.1.2.-), glutamate dehydrogenase (EC 1.4.1.2, 1.4.1.3, 1.4.1.4), phenylalanine dehydrogenase (EC 1.4.1.20), and/or D-tryptophan aminotransferase.

Monatin can be produced by a method that includes contacting tryptophan and/or indole-3-lactic acid with a first polypeptide, wherein the first polypeptide converts tryptophan and/or indole-3-lactic acid to indole-3-pyruvate (either the D or the L form of tryptophan or indole-3-lactic acid can be used as the substrate that is converted to indole-3-pyruvate; one of skill in the art will appreciate that the polypeptides chosen for this step ideally exhibit the appropriate specificity), contacting the resulting indole-3-pyruvate with a second polypeptide, wherein the second polypeptide converts the indole-3-pyruvate to 2-hydroxy 2-(indol-3-ylmethyl)-4-keto glutaric acid (MP), and contacting the MP with a third polypeptide, wherein the third polypeptide converts MP to monatin. Exemplary polypeptides that can be used for these conversions are shown in FIGS. 2 and 3.

Another aspect of the invention provides compositions such as MP, cells that contain at least two polypeptides, or sometimes at least three or at least four polypeptides, that are encoded on at least one exogenous nucleic acid sequence.

These and other aspects of the disclosure are apparent from the following detailed description and *illustrative examples*.

BRIEF DESCRIPTION OF THE FIGURES

FIG. 1 shows biosynthetic pathways used to produce monatin and/or indole-3-pyruvate. One pathway produces indole-3-pyruvate via tryptophan, while another produces indole-3-pyruvate via indole-3-lactic acid. Monatin is subsequently produced via a 2-hydroxy 2-(indol-3ylmethyl)-4-keto glutaric acid (MP) intermediate.

Compounds shown in boxes are substrates and products produced in the biosynthetic pathways.

Compositions adjacent to the arrows are cofactors, or reactants that can be used during the conversion of a substrate to a product. The cofactor or reactant used will depend upon the polypeptide used for the particular step of the biosynthetic pathway. The cofactor PLP (pyridoxal 5' -phosphate) can catalyze reactions independent of a polypeptide, and therefore, merely providing PLP can allow for the progression from substrate to product.

FIG. 2 is a more detailed diagram of the biosynthetic pathway that utilizes the MP intermediate. The substrates for each step in the pathways are shown in boxes. The polypeptides allowing for the conversion between substrates are listed adjacent to the arrows between the substrates. Each polypeptide is described by its common name and an enzymatic class (EC) number.

FIG. 3 shows a more detailed diagram of the biosynthetic pathway of the conversion of indole-3-lactic acid to indole-3-pyruvate. The substrates are shown in boxes, and the polypeptides allowing for the conversion between the substrates are listed adjacent to the arrow between the substrates. Each polypeptide is described by its common name and an enzymatic class (EC) number.

FIG. 4 shows one possible reaction for making MP via chemical means.

FIGS. 5A and 5B are chromatograms showing the LC/MS identification of monatin produced enzymatically.

FIG. 6 is an electrospray mass spectrum of enzymatically synthesized monatin.

FIGS. 7A and 7B show chromatograms of the LC/MS/MS daughter ion analyses of monatin produced in an enzymatic mixture.

FIG. 8 is a chromatogram showing the high resolution mass measurement of monatin produced enzymatically.

FIGS. 9A-9C are chromatograms showing the chiral separation of (A) R-tryptophan, (B) S-tryptophan, and (C) monatin produced enzymatically.

FIG. 10 is a bar graph showing the relative amount of monatin produced in bacterial cells following IPTG induction. The (-) indicates a lack of substrate addition (no tryptophan or pyruvate was added).

FIGS. 11-12 are schematic diagrams showing pathways used to increase the yield of monatin produced from tryptophan or indole-3-pyruvate.

FIG. 13 is a schematic diagram showing a pathway which can be used to increase the yield of monatin produced from tryptophan or indole-3-pyruvate.

SEQUENCE LISTING

The nucleic and amino acid sequences listed in the accompanying sequence listing are shown using standard letter abbreviations for nucleotide bases, and three-letter code for amino acids. Only one strand of each nucleic acid sequence is shown, but the complementary strand is understood to be included by any reference to the displayed strand.

SEQ ID NOS: 1 and 2 show the nucleic acid and amino acid sequences of an aminotransferase from *Sinorhizobium meliloti*, respectively (*tatA* gene, called a tyrosine or aromatic aminotransferase in literature).

SEQ ID NOS: 3 and 4 show the nucleic acid and amino acid sequences of a tyrosine aminotransferase from *Rhodobacter sphaeroides* (2.4.1), respectively (by homology with *tatA* (SEQ ID NOS: 1 and 2) predicted to be an "aspartate aminotransferase" by genomics software).

SEQ ID NOS: 5 and 6 show the nucleic acid and amino acid sequences of an aminotransferase from *Rhodobacter sphaeroides* (35053), respectively (novel, cloned based on 2.4.1 sequence SEQ ID NOS 3 and 4).

SEQ ID NOS: 7 and 8 show the nucleic acid and amino acid sequences of a broad substrate aminotransferase (*bsat*) from *Leishmania major*, respectively.

SEQ ID NOS: 9 and 10 show the nucleic acid and amino acid sequences of an aromatic aminotransferase (*araT*) from *Bacillus subtilis*, respectively.

SEQ ID NOS: 11 and 12 show novel nucleic acid and amino acid sequences of an aromatic aminotransferase (*araT*) from *Lactobacillus amylovorus*, respectively (by homology identified as an aromatic aminotransferase).

SEQ ID NOS: 13 and 14 show the nucleic acid and amino acid sequences of a multiple substrate aminotransferase (*mso*) from *R. sphaeroides* (35053), respectively (identified as a multiple substrate aminotransferase by homology to Accession No. AAAE01000093.1, bp 14743-16155 and Accession No. ZP00005082.1).

SEQ ID NOS: 15 and 16 show primers used to clone the *B. subtilis* D-alanine aminotransferase (*dat*) sequence.

SEQ ID NOS: 17 and 18 show primers used to clone the *S. meliloti tatA* sequence.

SEQ ID NOS: 19 and 20 show primers used to clone the *B. subtilis araT* aminotransferase sequence.

SEQ ID NOS: 21 and 22 show primers used to clone the *Rhodobacter sphaeroides* (2.4.1 and 35053) multiple substrate aminotransferase sequences.

SEQ ID NOS: 23 and 24 show primers used to clone the *Leishmania major bsat* sequence.

SEQ ID NOS: 25 and 26 show primers used to clone the *Lactobacillus amylovorus araT* sequence.

SEQ ID NOS: 27 and 28 show primers used to clone the *R. sphaeroides tatA* sequences (both 2.4.1 and 35053).

SEQ ID NOS: 29 and 30 show primers used to clone the *E. coli aspC* sequence (gene sequence Genbank Accession No.: AE000195.1, protein sequence Genbank Accession No.:AAC74014.1).

SEQ ID NOS: 31 and 32 show the nucleic acid and amino acid sequences of aromatic aminotransferase (*tyrB*) from *E. coli*, respectively.

SEQ ID NOS: 33 and 34 show primers used to clone the *E. coli tyrB* sequence.

SEQ ID NOS: 35-40 show primers used to clone polypeptides with 4-hydroxy-2-oxoglutarate aldolase (KHG) (EC 4.1.3.16) activity.

SEQ ID NOS: 41 and 42 show the nucleic acid sequences of tryptophanase (*tna*) from *E. coli* and tyrosine phenol-lyase (*tpl*) from *Citrobacter freundii*, coding for proteins P00913 (GI:401195) and P31013 (GI:401201), respectively.

SEQ ID NOS: 43 - 46 show primers used to clone tryptophanase polypeptides and β-tyrosinase (tyrosine phenol-lyase) polypeptides.

SEQ ID NOS: 47 – 54 show primers used to mutate tryptophanase polypeptides and β-tyrosinase polypeptides.

SEQ ID NOS: 55-64 show primers used to clone polypeptides with 4-hydroxy-4-methyl-2-oxoglutarate aldolase (EC 4.1.3.17) activity.

SEQ ID NOS: 65 and 66 show the nucleic acid and amino acid sequences of 4-hydroxy-4-methyl-2-oxoglutarate aldolase (*proA*) from *C. testosteroni*, respectively.

SEQ ID NOS: 67-68 show primers used to clone *C. testosteroni* 4-hydroxy-4-methyl-2-oxoglutarate aldolase (*proA*) in an operon with *E. coli aspC* in pET30 Xa/LIC.

SEQ ID NOS: 69-72 show primers used to clone *E. coli aspC* and *C. testosteroni proA* in pESC-his.

SEQ ID NOS: 73-74 show sequences added to the 5' end of primers used to clone the genes disclosed herein.

DETAILED DESCRIPTION OF SEVERAL EMBODIMENTS

Abbreviations and Terms

The following explanations of terms and methods are provided to better describe the present disclosure and to guide those of ordinary skill in the art in the practice of the present disclosure. As used herein, "comprising" means "including" and the singular forms "a" or "an" or "the" include plural references unless the context clearly dictates otherwise. For example, reference to "comprising a protein" includes one or a plurality of such proteins, and reference to "comprising the cell" includes reference to one or more cells and equivalents thereof known to those skilled in the art, and so forth.

Unless explained otherwise, all technical and scientific terms used herein have the same meaning as commonly understood to one of ordinary skill in the art to which this disclosure belongs. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present disclosure, suitable methods and materials are described below. The materials, methods, and examples are illustrative only and not intended to be limiting. Other features and advantages of the disclosure are apparent from the following detailed description and the claims.

cDNA (complementary DNA): A piece of DNA lacking internal, non-coding segments (introns) and regulatory sequences which determine transcription. cDNA can be synthesized in the laboratory by reverse transcription from messenger RNA extracted from cells.

Conservative substitution: One or more amino acid substitutions (for example 2, 5, or 10 residues) for amino acid residues having similar biochemical properties. Typically, conservative substitutions have little to no impact on the activity of a resulting polypeptide. For example, ideally, a tryptophan aminotransferase polypeptide including one or more conservative substitutions retains tryptophan aminotransferase activity. A polypeptide can be produced to contain one or more conservative substitutions by manipulating the nucleotide sequence that encodes that polypeptide using, for example, standard procedures such as site-directed mutagenesis or PCR.

Substitutional variants are those in which at least one residue in the amino acid sequence has been removed and a different residue inserted in its place. Examples of amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative substitutions include: Ala substituted with ser or thr; arg substituted with gln, his, or lys; asn substituted with glu, gln, lys, his, asp; asp substituted with asn, glu, or gln; cys substituted with ser or ala; gln substituted with asn, glu, lys, his, asp, or arg; glu substituted with asn, gln lys, or asp; gly

substituted with pro; his substituted with asn, lys, gln, arg, tyr; ile substituted with leu, met, val, phe; leu substituted with ile, met, val, phe; lys substituted with asn, glu, gln, his, arg; met substituted with ile, leu, val, phe; phe substituted with trp, tyr, met, ile, or leu; ser substituted with thr, ala; thr substituted with ser or ala; trp substituted with phe, tyr; tyr substituted with his, phe, or trp; and val substituted with met, ile, leu.

Further information about conservative substitutions can be found in, among other locations, Ben-Bassat *et al.*, (*J. Bacteriol.* 169:751-7, 1987), O'Regan *et al.*, (*Gene* 77:237-51, 1989), Sahin-Toth *et al.*, (*Protein Sci.* 3:240-7, 1994), Hochuli *et al.*, (Bio/Technology 6:1321-5, 1988), WO 00/67796 (Curd *et al.*) and in standard textbooks of genetics and molecular biology.

Exogenous: The term "exogenous" as used herein with reference to nucleic acid and a particular cell refers to any nucleic acid that does not originate from that particular cell as found in nature. Thus, non-naturally-occurring nucleic acid is considered to be exogenous to a cell once introduced into the cell. Nucleic acid that is naturally-occurring also can be exogenous to a particular cell. For example, an entire chromosome isolated from a cell of person X is an exogenous nucleic acid with respect to a cell of person Y once that chromosome is introduced into Y's cell.

Functionally Equivalent: Having an equivalent function. In the context of an enzyme, functionally equivalent molecules include different molecules that retain the function of the enzyme. For example, functional equivalents can be provided by sequence alterations in an enzyme sequence, wherein the peptide with one or more sequence alterations retains a function of the unaltered peptide, such that it retains its enzymatic activity. In a particular example, a tryptophan aminotransferase functional equivalent retains the ability to convert tryptophan to indole-3-pyruvate.

Examples of sequence alterations include, but are not limited to, conservative substitutions, deletions, mutations, frameshifts, and insertions. In one example, a given polypeptide binds an antibody, and a functional equivalent is a polypeptide that binds the same antibody. Thus a functional equivalent includes peptides that have the same binding specificity as a polypeptide, and that can be used as a reagent in place of the polypeptide. In one example a functional equivalent includes a polypeptide wherein the binding sequence is discontinuous, wherein the antibody binds a linear epitope. Thus, if the peptide sequence is MPELANDLGL (amino acids 1-10 of SEQ ID NO: 12) a functional equivalent includes discontinuous epitopes, that can appear as follows (**=any number of intervening amino acids): NH2 -**-M**P**E**L**A**N**D**L**G**L-COOH. In this example, the polypeptide is functionally equivalent to amino acids 1-10 of SEQ ID NO: 12 if the three dimensional structure of the polypeptide is such that it can bind a monoclonal antibody that binds amino acids 1-10 of SEQ ID NO: 12.

Hybridization: The term "hybridization" as used herein refers to a method of testing for complementarity in the nucleotide sequence of two nucleic acid molecules, based on the ability of complementary single-stranded DNA and/or RNA to form a duplex molecule. Nucleic acid hybridization techniques can be used to obtain an isolated nucleic acid within the scope of the disclosure. Briefly, any nucleic acid having some homology to a sequence set forth in SEQ ID NO: 11 can be used as a probe to identify a similar nucleic acid by hybridization under conditions of

moderate to high stringency. Once identified, the nucleic acid then can be purified, sequenced, and analyzed to determine whether it is within the scope of the present disclosure.

Hybridization can be done by Southern or Northern analysis to identify a DNA or RNA sequence, respectively, that hybridizes to a probe. The probe can be labeled with a biotin, digoxygenin, a polypeptide, or a radioisotope such as ^{32}P. The DNA or RNA to be analyzed can be electrophoretically separated on an agarose or polyacrylamide gel, transferred to nitrocellulose, nylon, or other suitable membrane, and hybridized with the probe using standard techniques well known in the art such as those described in sections 7.39-7.52 of Sambrook *et al.*, (1989) Molecular Cloning, second edition, Cold Spring Harbor Laboratory, Plainview, NY. Typically, a probe is at least about 20 nucleotides in length. For example, a probe corresponding to a contiguous 20 nucleotide sequence set forth in SEQ ID NO: 11 can be used to identify an identical or similar nucleic acid. In addition, probes longer or shorter than 20 nucleotides can be used.

The disclosure also provides isolated nucleic acid sequences that are at least about 12 bases in length (e.g., at least about 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 40, 50, 60, 100, 250, 500, 750, 1000, 1500, 2000, 3000, 4000, or 5000 bases in length) and hybridize, under hybridization conditions, to the sense or antisense strand of a nucleic acid having the sequence set forth in SEQ ID NO: 11. The hybridization conditions can be moderately or highly stringent hybridization conditions.

For the purpose of this disclosure, moderately stringent hybridization conditions mean the hybridization is performed at about 42°C in a hybridization solution containing 25 mM KPO_4 (pH 7.4), 5X SSC, 5X Denhart's solution, 50 µg/mL denatured, sonicated salmon sperm DNA, 50% formamide, 10% Dextran sulfate, and 1-15 ng/mL probe (about $5x10^7$ cpm/µg), while the washes are performed at about 50°C with a wash solution containing 2X SSC and 0.1% sodium dodecyl sulfate.

Highly stringent hybridization conditions mean the hybridization is performed at about 42°C in a hybridization solution containing 25 mM KPO_4 (pH 7.4), 5X SSC, 5X Denhart's solution, 50 µg/mL denatured, sonicated salmon sperm DNA, 50% formamide, 10% Dextran sulfate, and 1-15 ng/mL probe (about $5x10^7$ cpm/µg), while the washes are performed at about 65°C with a wash solution containing 0.2X SSC and 0.1% sodium dodecyl sulfate.

Isolated: The term "isolated" as used herein with reference to nucleic acid refers to a naturally-occurring nucleic acid that is not immediately contiguous with both of the sequences with which it is immediately contiguous (one on the 5' end and one on the 3' end) in the naturally-occurring genome of the organism from which it is derived. For example, an isolated nucleic acid can be, without limitation, a recombinant DNA molecule of any length, provided one of the nucleic acid sequences normally found immediately flanking that recombinant DNA molecule in a naturally-occurring genome is removed or absent. Thus, an isolated nucleic acid includes, without limitation, a recombinant DNA that exists as a separate molecule (e.g., a cDNA or a genomic DNA fragment produced by PCR or restriction endonuclease treatment) independent of other sequences as well as recombinant DNA that is incorporated into a vector, an autonomously replicating plasmid, a virus (e.g., a retrovirus, adenovirus, or herpes virus), or into the genomic DNA of a prokaryote or eukaryote. In addition, an isolated nucleic acid can include a recombinant DNA molecule that is part

of a hybrid or fusion nucleic acid sequence.

The term "isolated" as used herein with reference to nucleic acid also includes any non-naturally-occurring nucleic acid since non-naturally-occurring nucleic acid sequences are not found in nature and do not have immediately contiguous sequences in a naturally-occurring genome. For example, non-naturally-occurring nucleic acid such as an engineered nucleic acid is considered to be isolated nucleic acid. Engineered nucleic acid can be made using common molecular cloning or chemical nucleic acid synthesis techniques. Isolated non-naturally-occurring nucleic acid can be independent of other sequences, or incorporated into a vector, an autonomously replicating plasmid, a virus (e.g., a retrovirus, adenovirus, or herpes virus), or the genomic DNA of a prokaryote or eukaryote. In addition, a non-naturally-occurring nucleic acid can include a nucleic acid molecule that is part of a hybrid or fusion nucleic acid sequence.

It will be apparent to those of skill in the art that a nucleic acid existing among hundreds to millions of other nucleic acid molecules within, for example, cDNA or genomic libraries, or gel slices containing a genomic DNA restriction digest is not to be considered an isolated nucleic acid.

Nucleic acid: The term "nucleic acid" as used herein encompasses both RNA and DNA including, without limitation, cDNA, genomic DNA, and synthetic (e.g., chemically synthesized) DNA. The nucleic acid can be double-stranded or single-stranded. Where single-stranded, the nucleic acid can be the sense strand or the antisense strand. In addition, nucleic acid can be circular or linear.

Operably linked: A first nucleic acid sequence is "operably linked" with a second nucleic acid sequence whenever the first nucleic acid sequence is placed in a functional relationship with the second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription of the coding sequence. Generally, operably linked DNA sequences are contiguous and, where necessary to join two polypeptide-coding regions, in the same reading frame.

Peptide Modifications: The present disclosure includes enzymes, as well as synthetic embodiments thereof. In addition, analogues (non-peptide organic molecules), derivatives (chemically functionalized peptide molecules obtained starting with the disclosed peptide sequences) and variants (homologs) having the desired enzymatic activity can be utilized in the methods described herein. The peptides disclosed herein include a sequence of amino acids, that can be either L- and/or D- amino acids, naturally occurring and otherwise.

Peptides can be modified by a variety of chemical techniques to produce derivatives having essentially the same activity as the unmodified peptides, and optionally having other desirable properties. For example, carboxylic acid groups of the protein, whether carboxyl-terminal or side chain, may be provided in the form of a salt of a pharmaceutically-acceptable cation or esterified to form a C1-C16 ester, or converted to an amide of formula $NR1R2$ wherein $R1$ and $R2$ are each independently H or C1-C16 alkyl, or combined to form a heterocyclic ring, such as a 5- or 6-membered ring. Amino groups of the peptide, whether amino-terminal or side chain, may be in the form of a pharmaceutically-acceptable acid addition salt, such as the HCl, HBr, acetic, benzoic,

toluene sulfonic, maleic, tartaric and other organic salts, or may be modified to C1-C16 alkyl or dialkyl amino or further converted to an amide.

Hydroxyl groups of the peptide side chains may be converted to C1-C16 alkoxy or to a C1-C16 ester using well-recognized techniques. Phenyl and phenolic rings of the peptide side chains may be substituted with one or more halogen atoms, such as F, Cl, Br or I, or with C1-C16 alkyl, C1-C16 alkoxy, carboxylic acids and esters thereof, or amides of such carboxylic acids. Methylene groups of the peptide side chains can be extended to homologous C2-C4 alkylenes. Thiols can be protected with any one of a number of well-recognized protecting groups, such as acetamide groups. Those skilled in the art will also recognize methods for introducing cyclic structures into the peptides of this disclosure to select and provide conformational constraints to the structure that result in enhanced stability. For example, a C- or N-terminal cysteine can be added to the peptide, so that when oxidized the peptide will contain a disulfide bond, generating a cyclic peptide. Other peptide cyclizing methods include the formation of thioethers and carboxyl- and amino-terminal amides and esters.

Peptidomimetic and organomimetic embodiments are also within the scope of the present disclosure, whereby the three-dimensional arrangement of the chemical constituents of such peptido- and organomimetics mimic the three-dimensional arrangement of the peptide backbone and component amino acid side chains, resulting in such peptido- and organomimetics of the proteins of this disclosure having detectable enzyme activity. For computer modeling applications, a pharmacophore is an idealized, three-dimensional definition of the structural requirements for biological activity. Peptido- and organomimetics can be designed to fit each pharmacophore with current computer modeling software (using computer assisted drug design or CADD). See Walters, "Computer-Assisted Modeling of Drugs", in Klegerman & Groves (eds.), Pharmaceutical Biotechnology, 1993, Interpharm Press: Buffalo Grove, IL, pp. 165-74 and Ch. 102 in Munson (ed.), Principles of Pharmacology, 1995, Chapman & Hall, for descriptions of techniques used in CADD. Also included within the scope of the disclosure are mimetics prepared using such techniques. In one example, a mimetic mimics the enzyme activity generated by an enzyme or a variant, fragment, or fusion thereof.

Probes and primers: Nucleic acid probes and primers can be prepared readily based on the amino acid sequences and nucleic acid sequences provided herein. A "probe" includes an isolated nucleic acid containing a detectable label or reporter molecule. Exemplary labels include, but are not limited to, radioactive isotopes, ligands, chemiluminescent agents, and polypeptides. Methods for labeling and guidance in the choice of labels appropriate for various purposes are discussed in, for example, Sambrook *et al.* (ed.), Molecular Cloning: A Laboratory Manual 2nd ed., vol. 1-3, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989, and Ausubel *et al.* (ed.) Current Protocols in Molecular Biology, Greene Publishing and Wiley-Interscience, New York (with periodic updates), 1987.

"Primers" are typically nucleic acid molecules having ten or more nucleotides (e.g., nucleic acid molecules having between about 10 nucleotides and about 100 nucleotides). A primer can be

annealed to a complementary target nucleic acid strand by nucleic acid hybridization to form a hybrid between the primer and the target nucleic acid strand, and then extended along the target nucleic acid strand by, for example, a DNA polymerase polypeptide. Primer pairs can be used for amplification of a nucleic acid sequence, for example, by the polymerase chain reaction (PCR) or other nucleic-acid amplification methods known in the art.

Methods for preparing and using probes and primers are described, for example, in references such as Sambrook *et al.* (ed.), *Molecular Cloning: A Laboratory Manual*, 2nd ed., vol. 1-3, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989; Ausubel *et al.* (ed.), Current Protocols in Molecular Biology, Greene Publishing and Wiley-Interscience, New York (with periodic updates), 1987; and Innis *et al.* (eds.), PCR Protocols: A Guide to Methods and Applications, Academic Press: San Diego, 1990. PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge, Mass.). One of skill in the art will appreciate that the specificity of a particular probe or primer increases with the length, but that a probe or primer can range in size from a full-length sequence to sequences as short as five consecutive nucleotides. Thus, for example, a primer of 20 consecutive nucleotides can anneal to a target with a higher specificity than a corresponding primer of only 15 nucleotides. Thus, in order to obtain greater specificity, probes and primers can be selected that comprise, for example, 10, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1250, 1300, 1350, 1400, 1450, 1500, 1550, 1600, 1650, 1700, 1750, 1800, 1850, 1900, 2000, 2050, 2100, 2150, 2200, 2250, 2300, 2350, 2400, 2450, 2500, 2550, 2600, 2650, 2700, 2750, 2800, 2850, 2900, 3000, 3050,3100, 3150, 3200, 3250, 3300, 3350, 3400, 3450, 3500, 3550, 3600, 3650, 3700, 3750, 3800, 3850, 3900, 4000, 4050, 4100, 4150, 4200, 4250, 4300, 4350, 4400, 4450, 4500, 4550, 4600, 4650, 4700, 4750, 4800, 4850, 4900, 5000, 5050, 5100, 5150, 5200, 5250, 5300, 5350, 5400, 5450, or more consecutive nucleotides.

Promoter: An array of nucleic acid control sequences which direct transcription of a nucleic acid. A promoter includes necessary nucleic acid sequences near the start site of transcription, such as, in the case of a polymerase II type promoter, a TATA element. A promoter can include distal enhancer or repressor elements which can be located as much as several thousand base pairs from the start site of transcription.

Purified: The term "purified" as used herein does not require absolute purity; rather, it is intended as a relative term. Thus, for example, a purified polypeptide or nucleic acid preparation can be one in which the subject polypeptide or nucleic acid, respectively, is at a higher concentration than the polypeptide or nucleic acid would be in its natural environment within an organism. For example, a polypeptide preparation can be considered purified if the polypeptide content in the preparation represents at least 50%, 60%, 70%, 80%, 85%, 90%, 92%, 95%, 98%, or 99% of the total protein content of the preparation.

Recombinant: A "recombinant" nucleic acid is one having (1) a sequence that is not naturally occurring in the organism in which it is expressed or (2) a sequence made by an artificial

combination of two otherwise-separated, shorter sequences. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques. "Recombinant" is also used to describe nucleic acid molecules that have been artificially manipulated, but contain the same regulatory sequences and coding regions that are found in the organism from which the nucleic acid was isolated.

Sequence identity: The similarity between amino acid sequences is expressed in terms of the similarity between the sequences, otherwise referred to as sequence identity. Sequence identity is frequently measured in terms of percentage identity (or similarity or homology); the higher the percentage, the more similar the two sequences are. Homologs or variants of a peptide, such as SEQ ID NO: 12, possess a relatively high degree of sequence identity when aligned using standard methods.

Methods of alignment of sequences for comparison are well known in the art. Various programs and alignment algorithms are described in: Smith and Waterman, *Adv. Appl. Math.* 2:482, 1981; Needleman and Wunsch, *J. Mol. Biol.* 48:443-53, 1970; Pearson and Lipman, *Proc. Natl. Acad. Sci. U.S.A.* 85:2444-8, 1988; Higgins and Sharp, *Gene* 73:237-44, 1988; Higgins and Sharp, *CABIOS* 5:151-3, 1989; Corpet et al., *Nucleic Acids Research* 16:10881-90, 1988; and Altschul *et al.*, *Nature Genet.* 6:119-29, 1994.

The NCBI Basic Local Alignment Search Tool (BLAST™) (Altschul *et al.*, *J. Mol. Biol.* 215:403-10, 1990) is available from several sources, including the National Center for Biotechnology Information (NCBI, Bethesda, MD) and on the Internet, for use in connection with the sequence analysis programs blastp, blastn, blastx, tblastn and tblastx.

Variants of a peptide are typically characterized by possession of at least 50% sequence identity counted over the full length alignment with the amino acid sequence using the NCBI Blast 2.0, gapped blastp set to default parameters. For comparisons of amino acid sequences of greater than about 30 amino acids, the Blast 2 sequences function is employed using the default BLOSUM62 matrix set to default parameters, (gap existence cost of 11, and a per residue gap cost of 1). When aligning short peptides (fewer than around 30 amino acids), the alignment is performed using the Blast 2 sequences function, employing the PAM30 matrix set to default parameters (open gap 9, extension gap 1 penalties). Proteins with even greater similarity to the reference sequences will show increasing percentage identities when assessed by this method, such as at least 80%, at least 90%, at least 95%, at least 98%, or even at least 99% sequence identity. When less than the entire sequence is being compared for sequence identity, homologs and variants will typically possess at least 80% sequence identity over short windows of 10-20 amino acids, and may possess sequence identities of at least 85%, at least 90%, at least 95%, or 98% depending on their similarity to the reference sequence. Methods for determining sequence identity over such short windows are described at the website that is maintained by the National Center for Biotechnology Information in Bethesda, Maryland. One of skill in the art will appreciate that these sequence identity ranges are provided for guidance only; it is entirely possible that strongly significant homologs could be obtained that fall outside of the ranges provided.

Similar methods can be used to determine the percent sequence identity of a nucleic acid sequence. In a particular example, a homologous sequence is aligned to a native sequence, and the number of matches is determined by counting the number of positions where an identical nucleotide or amino acid residue is presented in both sequences. The percent sequence identity is determined by dividing the number of matches either by the length of the sequence set forth in the identified sequence (e.g., SEQ ID NO: 11), or by an articulated length (e.g., 100 consecutive nucleotides or amino acid residues from a sequence set forth in an identified sequence), followed by multiplying the resulting value by 100. For example, a nucleic acid sequence that has 1166 matches when aligned with the sequence set forth in SEQ ID NO: 11 is 75.0 percent identical to the sequence set forth in SEQ ID NO: 11 (i.e., 1166÷1554*100=75.0). It is noted that the percent sequence identity value is rounded to the nearest tenth. For example, 75.11, 75.12, 75.13, and 75.14 is rounded down to 75.1, while 75.15, 75.16, 75.17, 75.18, and 75.19 is rounded up to 75.2. It is also noted that the length value will always be an integer. In another example, a target sequence containing a 20-nucleotide region that aligns with 20 consecutive nucleotides from an identified sequence as follows contains a region that shares 75 percent sequence identity to that identified sequence (i.e., 15÷20*100=75).

```
                          1                   20
Target Sequence:          AGGTCGTGTACTGTCAGTCA
                          | || ||| |||| |||| |
Identified Sequence:      ACGTGGTGAACTGCCAGTGA
```

Specific binding agent: An agent that is capable of specifically binding to any of the polypeptide described herein. Examples include, but are not limited to, polyclonal antibodies, monoclonal antibodies (*including* humanized monoclonal antibodies), and fragments of monoclonal antibodies such as Fab, $F(ab')_2$, and Fv fragments as well as any other agent capable of specifically binding to an epitope of such polypeptides.

Antibodies to the polypeptides provided herein (or fragments, variants, or fusions thereof) can be used to purify or identify such polypeptides. The amino acid and nucleic acid sequences provided herein allow for the production of specific antibody-based binding agents that recognize the polypeptides described herein.

Monoclonal or polyclonal antibodies can be produced to the polypeptides, portions of the polypeptides, or variants thereof. Optimally, antibodies raised against one or more epitopes on a polypeptide antigen will specifically detect that polypeptide. That is, antibodies raised against one particular polypeptide would recognize and bind that particular polypeptide, and would not substantially recognize or bind to other polypeptides. The determination that an antibody specifically binds to a particular polypeptide is made by any one of a number of standard immunoassay methods; for instance, Western blotting (See, e.g., Sambrook *et al.* (ed.), *Molecular Cloning: A Laboratory Manual*, 2nd ed., vol. 1-3, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989).

To determine that a given antibody preparation (such as a preparation produced in a mouse against a polypeptide having the amino acid sequence set forth in SEQ ID NO: 12) specifically detects the appropriate polypeptide (e.g., a polypeptide having the amino acid sequence set forth in

SEQ ID NO: 12) by Western blotting, total cellular protein can be extracted from cells and separated by SDS-polyacrylamide gel electrophoresis.

The separated total cellular protein can then be transferred to a membrane (e.g., nitrocellulose), and the antibody preparation incubated with the membrane. After washing the membrane to remove non-specifically bound antibodies, the presence of specifically bound antibodies can be detected using an appropriate secondary antibody (e.g., an anti-mouse antibody) conjugated to a polypeptide such as alkaline phosphatase since application of 5-bromo-4-chloro-3-indolyl phosphate/nitro blue tetrazolium results in the production of a densely blue-colored compound by immuno-localized alkaline phosphatase.

Substantially pure polypeptides suitable for use as an immunogen can be obtained from transfected cells, transformed cells, or wild-type cells. Polypeptide concentrations in the final preparation can be adjusted, for example, by concentration on an Amicon filter device, to the level of a few micrograms per milliliter. In addition, polypeptides ranging in size from full-length polypeptides to polypeptides having as few as nine amino acid residues can be utilized as immunogens. Such polypeptides can be produced in cell culture, can be chemically synthesized using standard methods, or can be obtained by cleaving large polypeptides into smaller polypeptides that can be purified. Polypeptides having as few as nine amino acid residues in length can be immunogenic when presented to an immune system in the context of a Major Histocompatibility Complex (MHC) molecule such as an MHC class I or MHC class II molecule. Accordingly, polypeptides having at least 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 900, 1000, 1050, 1100, 1150, 1200, 1250, 1300, 1350, or more consecutive amino acid residues of any amino acid sequence disclosed herein can be used as immunogens for producing antibodies.

Monoclonal antibodies to any of the polypeptides disclosed herein can be prepared from murine hybridomas according to the classic method of Kohler & Milstein (Nature 256:495-7, 1975) or a derivative method thereof.

Polyclonal antiserum containing antibodies to the heterogeneous epitopes of any polypeptide disclosed herein can be prepared by immunizing suitable animals with the polypeptide (or fragment thereof), which can be unmodified or modified to enhance immunogenicity. An effective immunization protocol for rabbits can be found in Vaitukaitis *et al.* (*J. Clin. Endocrinol. Metab.* 33:988-91,1971).

Antibody fragments can be used in place of whole antibodies and can be readily expressed in prokaryotic host cells. Methods of making and using immunologically effective portions of monoclonal antibodies, also referred to as "antibody fragments," are well known and include those described in Better & Horowitz (Methods Enzymol. 178:476-96,1989), Glockshuber *et al.* (*Biochemistry* 29:1362-7,1990), U.S. Pat. No. 5,648,237 ("Expression of Functional Antibody Fragments"), U.S. Pat. No. 4,946,778 ("Single Polypeptide Chain Binding Molecules"), U.S. Pat. No. 5,455,030 ("Immunotherapy Using Single Chain Polypeptide Binding Molecules"), and references cited therein.

Transformed: A "transformed" cell is a cell into which a nucleic acid molecule has been introduced by, for example, molecular biology techniques. Transformation encompasses all techniques by which a nucleic acid molecule can be introduced into such a cell including, without limitation, transfection with a viral vector, conjugation, transformation with a plasmid vector, and introduction of naked DNA by electroporation, lipofection, and particle gun acceleration.

Variants, fragments or fusion proteins: The disclosed proteins, include variants, fragments, and fusions thereof. DNA sequences (for example SEQ ID NO: 11) which encode for a protein (for example SEQ ID NO: 12), fusion protein, or a fragment or variant of a protein, can be engineered to allow the protein to be expressed in eukaryotic cells, bacteria, insects, and/or plants. To obtain expression, the DNA sequence can be altered and operably linked to other regulatory sequences. The final product, which contains the regulatory sequences and the protein, is referred to as a vector. This vector can be introduced into eukaryotic, bacteria, insect, and/or plant cells. Once inside the cell the vector allows the protein to be produced.

A fusion protein including a protein, such as a tryptophan aminotransferase (or variant, polymorphism, mutant, or fragment thereof), for example SEQ ID NO: 12, linked to other amino acid sequences that do not inhibit the desired activity of the protein, for example the ability to convert tryptophan to indole-3-pyruvate. In one example, the other amino acid sequences are no more than about 10, 12, 15, 20, 25, 30, or 50 amino acids in length.

One of ordinary skill in the art will appreciate that a DNA sequence can be altered in numerous ways without affecting the biological activity of the encoded protein. For example, PCR can be used to produce variations in the DNA sequence which encodes an protein. Such variants can be variants optimized for codon preference in a host cell used to express the protein, or other sequence changes that facilitate expression.

Vector: A nucleic acid molecule as introduced into a cell, thereby producing a transformed cell. A vector may include nucleic acid sequences that permit it to replicate in the cell, such as an origin of replication. A vector may also include one or more selectable marker genes and other genetic elements known in the art.

Overview of Biosynthetic Pathways

As shown in FIGS. 1–3 and 11-13, many biosynthetic pathways can be used to produce monatin or its intermediates such as indole-3-pyruvate or MP. For the conversion of each substrate (glucose, tryptophan, indole-3-lactic acid, indole-3-pyruvate, and MP) to each product (tryptophan, indole-3-pyruvate, MP and monatin) several different polypeptides can be used. Moreover, these reactions can be carried out *in vivo*, *in vitro*, or through a combination of *in vivo* reactions and *in vitro* reactions, such as *in vitro* reactions that include non-enzymatic chemical reactions. Therefore, FIGS. 1-3 and 11-13 are exemplary, and show multiple different pathways that can be used to obtain desired products.

Glucose to Tryptophan

Many organisms can synthesize tryptophan from glucose. The construct(s) containing the gene(s) necessary to produce monatin, MP, and/or indole-3-pyruvate from glucose and/or tryptophan can be cloned into such organisms. It is shown herein that tryptophan can be converted into monatin.

In other examples, an organism is engineered using known polypeptides to produce tryptophan, or overproduce tryptophan. For example, U.S. Patent No. 4,371,614 (herein incorporated by reference) describes an *E. coli* strain transformed with a plasmid containing a wild type tryptophan operon.

Maximum titers of tryptophan disclosed in U.S. Patent No. 4,371,614 are about 230 ppm. Similarly, WO 8701130 (herein incorporated by reference) describes an *E. coli* strain that has been genetically engineered to produce tryptophan and discusses increasing fermentative production of L-tryptophan. Those skilled in the art will recognize that organisms capable of producing tryptophan from glucose are also capable of utilizing other carbon and energy sources that can be converted to glucose or fructose-6-phosphate, with similar results. Exemplary carbon and energy sources include, but are not limited to, sucrose, fructose, starch, cellulose, or glycerol.

Tryptophan to Indole-3-pyruvate

Several polypeptides can be used to convert tryptophan to indole-3-pyruvate. Exemplary polypeptides include members of the enzyme classes (EC) 2.6.1.27, 1.4.1.19, 1.4.99.1, 2.6.1.28, 1.4.3.2, 1.4.3.3, 2.6.1.5, 2.6.1.-, 2.6.1.1, and 2.6.1.21. These classes include polypeptides termed tryptophan aminotransferase (also termed L-phenylalanine-2-oxoglutarate aminotransferase, tryptophan transaminase, 5-hydroxytryptophan-ketoglutaric transaminase, hydroxytryptophan aminotransferase, L-tryptophan aminotransferase, L-tryptophan transaminase, and L-tryptophan:2-oxoglutarate aminotransferase) which converts L-tryptophan and 2-oxoglutarate to indole-3-pyruvate and L-glutamate; D-tryptophan aminotransferase which converts D-tryptophan and a 2-oxo acid to indole-3-pyruvate and an amino acid; tryptophan dehydrogenase (also termed NAD(P)-L-tryptophan dehydrogenase, L-tryptophan dehydrogenase, L-Trp-dehydrogenase, TDH and L-tryptophan:NAD(P) oxidoreductase (deaminating)) which converts L-tryptophan and NAD(P) to indole-3-pyruvate and NH_3 and NAD(P)H; D-amino acid dehydrogenase, which converts D-amino acids and FAD to indole-3-pyruvate and NH_3 and $FADH_2$; tryptophan-phenylpyruvate transaminase (also termed L-tryptophan-α-ketoisocaproate aminotransferase and L-tryptophan:phenylpyruvate aminotransferase) which converts L-tryptophan and phenylpyruvate to indole-3-pyruvate and L-phenylalanine; L-amino acid oxidase (also termed ophio-amino-acid oxidase and L-amino-acid:oxygen oxidoreductase (deaminating)) which converts an L-amino acid and H_2O and O_2 to a 2-oxo acid and NH_3 and H_2O_2; D-amino acid oxidase (also termed ophio-amino-acid oxidase and D-amino-acid:oxygen oxidoreductase (deaminating)) which converts a D-amino acid and H_2O and O_2 to a 2-oxo acid and NH_3 and H_2O_2; and tryptophan oxidase which converts L-tryptophan and H_2O and O_2 to indole-3-pyruvate and NH_3 and H_2O_2. These classes also contain tyrosine (aromatic) aminotransferase, aspartate aminotransferase, D-amino acid (or D-alanine) aminotransferase, and broad (multiple

substrate) aminotransferase which have multiple aminotransferase activities, some of which can convert tryptophan and a 2-oxo acid to indole-3-pyruvate and an amino acid.

Eleven members of the aminotransferase class that have such activity are described below in Example 1, including a novel aminotransferase shown in SEQ ID NOS: 11 and 12. Therefore, this disclosure provides isolated nucleic acid and protein sequences having at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or even at least 99% sequence identity to SEQ ID NOS: 11 and 12. Also encompassed by this disclosure are fragments and fusions of SEQ ID NOS: 11 and 12 that retain aminotransferase activity or encode a protein having aminotransferase activity. Exemplary fragments include, but are not limited to at least 10, 12, 15, 20, 25, 50, 100, 200, 500, or 1000 contiguous nucleotides of SEQ ID NO: 11 or at least 6, 10, 15, 20, 25, 50, 75, 100, 200, 300 or 350 contiguous amino acids of SEQ ID NO: 12. The disclosed sequences (and variants, fragments, and fusions thereof) can be part of a vector. The vector can be used to transform host cells, thereby producing recombinant cells which can produce indole-3-pyruvate from tryptophan, and in some examples can further produce MP and/or monatin.

L-amino acid oxidases (1.4.3.2) are known, and sequences can be isolated from several different sources, such as *Vipera lebetine* (sp P81375), *Ophiophagus hannah* (sp P81383), *Agkistrodon rhodostoma* (spP81382), *Crotalus atrox* (sp P56742), *Burkholderia cepacia*, *Arabidopsis thaliana*, *Caulobacter cresentus*, *Chlamydomonas reinhardtii*, *Mus musculus*, *Pseudomonas syringae*, and *Rhodococcus* str. In addition, tryptophan oxidases are described in the literature and can be isolated, for example, from *Coprinus* sp. SF-1, Chinese cabbage with club root disease, *Arabidopsis thaliana*, and mammalian liver. One member of the L-amino acid oxidase class that can convert tryptophan to indole-3-pyruvate is discussed below in Example 3, as well as alternative sources for molecular cloning. Many D-amino acid oxidase genes are available in databases for molecular cloning.

Tryptophan dehydrogenases are known, and can be isolated, for example, from spinach, *Pisum sativum*, *Prosopis juliflora*, pea, mesquite, wheat, maize, tomato, tobacco, *Chromobacterium violaceum*, and *Lactobacilli*. Many D-amino acid dehydrogenase gene sequences are known.

As shown in FIGS. 11-13, if an amino acid oxidase, such as tryptophan oxidase, is used to convert tryptophan to indole-3-pyruvate, catalase can be added to reduce or even eliminate the presence of hydrogen peroxide.

Indole-3-lactate to Indole-3-pyruvate

The reaction that converts indole-3-lactate to indole-3-pyruvate can be catalyzed by a variety of polypeptides, such as members of the 1.1.1.110, 1.1.1.27, 1.1.1.28, 1.1.2.3, 1.1.1.222, 1.1.1.237, 1.1.3.-, or 1.1.1.111 classes of polypeptides. The 1.1.1.110 class of polypeptides includes indolelactate dehydrogenases (also termed indolelactic acid: NAD^+ oxidoreductase). The 1.1.1.27, 1.1.1.28, and 1.1.2.3 classes include lactate dehydrogenases (also termed lactic acid dehydrogenases, lactate: NAD^+ oxidoreductase). The 1.1.1.222 class contains (R)-4-hydroxyphenyllactate dehydrogenase (also termed D-aromatic lactate dehydrogenase, R-aromatic lactate dehydrogenase,

and R-3-(4-hydroxyphenyl)lactate:NAD(P)$^+$ 2-oxidoreductase) and the 1.1.1.237 class contains 3-(4-hydroxyphenyl)pyruvate) reductase (also termed hydroxyphenylpyruvate reductase and 4-hydroxyphenyllactate: NAD$^+$ oxidoreductase). The 1.1.3.- class contains lactate oxidases, and the 1.1.1.111 class contains (3-imidazol-5-yl) lactate dehydrogenases (also termed (S)-3-(imidazol-5-yl)lactate:NAD(P)$^+$ oxidoreductase). It is likely that several of the polypeptides in these classes allow for the production of indole-3-pyruvate from indole-3-lactic acid. Examples of this conversion are provided in Example 2.

Chemical reactions can also be used to convert indole-3-lactic acid to indole-3-pyruvate. Such chemical reactions include an oxidation step that can be accomplished using several methods, for example: air oxidation using a B2 catalyst (China Chemical Reporter, v 13, n 28, p 18 (1), 2002), dilute permanganate and perchlorate, or hydrogen peroxide in the presence of metal catalysts.

Indole-3-pyruvate to 2-hydroxy 2-(indol-3ylmethyl)-4-keto glutaric acid (MP)

Several known polypeptides can be used to convert indole-3-pyruvate to MP. Exemplary polypeptide classes include 4.1.3.-, 4.1.3.16, 4.1.3.17, and 4.1.2.-. These classes include carbon-carbon synthases/lyases, such as aldolases that catalyze the condensation of two carboxylic acid substrates. Peptide class EC 4.1.3.- are synthases/lyases that form carbon-carbon bonds utilizing oxo-acid substrates (such as indole-3-pyruvate) as the electrophile, while EC 4.1.2.- are synthases/lyases that form carbon-carbon bonds utilizing aldehyde substrates (such as benzaldehyde) as the electrophile.

For example, the polypeptide described in EP 1045-029 (EC 4.1.3.16, 4-hydroxy-2-oxoglutarate glyoxylate-lyase also termed 4-hydroxy-2-oxoglutarate aldolase, 2-oxo-4-hydroxyglutarate aldolase or KHG aldolase) converts glyoxylic acid and pyruvate to 4-hydroxy-2-ketoglutaric acid, and the polypeptide 4-hydroxy-4-methyl-2-oxoglutarate aldolase (EC 4.1.3.17, also termed 4-hydroxy-4-methyl-2-oxoglutarate pyruvate-lyase or ProA aldolase), condenses two keto-acids such as two pyruvates to 4-hydroxy-4-methyl-2-oxoglutarate. Reactions utilizing these lyases are described herein.

FIGS. 1-2 and 11-13 show schematic diagrams of these reactions in which a 3-carbon (C3) molecule is combined with indole-3-pyruvate. Many members of EC 4.1.2.- and 4.1.3.-, particularly PLP-utilizing polypeptides, can utilize C3 molecules that are amino acids such as serine, cysteine, and alanine, or derivatives thereof. Aldol condensations catalyzed by representatives of EC 4.1.2.- and 4.1.3.- require the three carbon molecule of this pathway to be pyruvate or a derivative of pyruvate. However, other compounds can serve as a C3 carbon source and be converted to pyruvate. Alanine can be transaminated by many PLP-utilizing transaminases, including many of those mentioned above, to yield pyruvate. Pyruvate and ammonia can be obtained by beta-elimination reactions (such as those catalyzed by tryptophanase or β-tyrosinase) of L-serine, L-cysteine, and derivatives of serine and cysteine with sufficient leaving groups, such as O-methyl-L-serine, O-benzyl-L-serine, S-methylcysteine, S-benzylcysteine, S-alkyl-L-cysteine, O-acyl-L-serine, and 3-chloro-L-alanine. Aspartate can serve as a source of pyruvate in PLP-mediated beta-lyase reactions

such as those catalyzed by tryptophanase (EC 4.1.99.1) and/or β-tyrosinase (EC 4.1.99.2, also termed tyrosine-phenol lyase). The rate of beta-lyase reactions can be increased by performing site-directed mutagenesis on the (4.1.99.1-2) polypeptides as described by Mozratou *et al.* (*J. Biol. Chem* 274:1320-5, 1999) and in Example 8. These modifications allow the polypeptides to accept dicarboxylic amino acid substrates. Lactate can also serve as a source of pyruvate, and is oxidized to pyruvate by the addition of lactate dehydrogenase and an oxidized cofactor or lactate oxidase and oxygen. Examples of these reactions are described below. For example, as shown in FIG. 2 and FIGS. 11-13, ProA aldolase can be contacted with indole-3-pyruvate when pyruvate is used as the C3 molecule.

The MP can also be generated using chemical reactions, such as the aldol condensations provided in Example 5.

MP to Monatin

Conversion of MP to monatin can be catalyzed by one or more of: tryptophan aminotransferases (2.6.1.27), tryptophan dehydrogenases (1.4.1.19), D-amino acid dehydrogenases (1.4.99.1), glutamate dehydrogenases (1.4.1.2-4), phenylalanine dehydrogenase (EC 1.4.1.20), tryptophan-phenylpyruvate transaminases (2.6.1.28), or more generally members of the aminotransferase family (2.6.1.-) such as aspartate aminotransferase (EC 2.6.1.1); tyrosine (aromatic) aminotransferase (2.6.1.5), D-tryptophan aminotransferase, or D-alanine (2.6.1.21) aminotransferase (FIG. 2). Eleven members of the aminotransferase class are described below (Example 1), including a novel member of the class shown in SEQ ID NOS: 11 and 12, and reactions demonstrating the activity of aminotransferase and dehydrogenase enzymes are provided in Example 7.

This reaction can also be performed using chemical reactions. Amination of the keto acid (MP) is performed by reductive amination using ammonia and sodium cyanoborohydride.

FIGS. 11-13 show additional polypeptides that can be used to convert MP to monatin, as well as providing increased yields of monatin from indole-3-pyruvate or tryptophan. For example, if aspartate is used as the amino donor, aspartate aminotransferase can be used to convert the aspartate to oxaloacetate (FIG. 11). The oxaloacetate is converted to pyruvate and carbon dioxide by a decarboxylase, such as oxaloacetate decarboxylase (FIG. 11). In addition, if lysine is used as the amino donor, lysine epsilon aminotransferase can be used to convert the lysine to allysine (FIG. 12). The allysine is spontaneously converted to 1-piperideine 6-carboxylate (FIG. 12). If a polypeptide capable of catalyzing reductive amination reactions (e.g., glutamate dehydrogenase) is used to convert MP to monatin, a polypeptide that can recycle NAD(P)H and/or produce a volatile product (FIG. 13) can be used, such as formate dehydrogenase.

Additional Considerations in the Design of the Biosynthetic Pathways

Depending on which polypeptides are used to generate indole-3-pyruvate, MP and/or monatin, cofactors, substrates, and/or additional polypeptides can be provided to the production cell to enhance product formation.

Removal of Hydrogen Peroxide

Hydrogen peroxide (H_2O_2) is a product that, if generated, can be toxic to production cells and can damage the polypeptides or intermediates produced. The L-amino acid oxidase described above generates H_2O_2 as a product. Therefore, if L-amino acid oxidase is used, the resulting H_2O_2 can be removed or its levels decreased to decrease potential injury to the cell or product.

Catalases can be used to reduce the level of H_2O_2 in the cell (FIGS. 11-13). The production cell can express a gene or cDNA sequence that encodes a catalase (EC 1.11.1.6), which catalyzes the decomposition of hydrogen peroxide into water and oxygen gas. For example, a catalase can be expressed from a vector transfected into the production cell. Examples of catalases that can be used to include, but are not limited to: tr|Q9HV50 (*Staphylococcus xylosus*), tr|Q9KBE8 (*Bacillus halodurans*), tr|Q9URJ7 (*Candida albicans*), tr|P77948 (*Streptomyces coelicolor*), tr|Q9RBJ5 (*Xanthomonas campestris*) (SwissProt Accession Nos.). Biocatalytic reactors utilizing L-amino acid oxidase, D-amino acid oxidase, or tryptophan oxidase can also contain a catalase polypeptide.

Modulation of PLP (pyridoxal-5'-phosphate) Availability

As shown in FIG. 1, PLP can be utilized in one or more of the biosynthetic steps described herein. The concentration of PLP can be supplemented so that PLP does not become a limitation on the overall efficiency of the reaction.

The biosynthetic pathway for vitamin B_6 (the precursor of PLP) has been thoroughly studied in *E. coli* and some of the proteins have been crystallized (Laber *et al.*, *FEBS Letters*, 449:45-8, 1999). Two of the genes (*epd* or *gapB* and *serC*) are required in other metabolic pathways, while three genes (*pdxA*, *pdxB*, and *pdxJ*) are unique to pyridoxal phosphate biosynthesis. One of the starting materials in the *E. coli* pathway is 1-deoxy-D-xylulose-5-phosphate (DXP). Synthesis of this precursor from common 2 and 3 carbon central metabolites is catalyzed by the polypeptide 1-deoxy-D-xylulose-5-phosphate synthase (DSX). The other precursor is a threonine derivative formed from the 4-carbon sugar, D-erythrose 4-phosphate. The genes required for the conversion to phospho-4-hydroxyl-L threonine (HTP) are *epd*, *pdxB*, and *serC*. The last reaction for the formation of PLP is a complex intramolecular condensation and ring-closure reaction between DXP and HTP, catalyzed by the gene products of *pdxA* and *pdxJ*.

If PLP becomes a limiting nutrient during the fermentation to produce monatin, increased expression of one or more of the pathway genes in a production host cell can be used to increase the yield of monatin. A host organism can contain multiple copies of its native pathway genes or copies of non-native pathway genes can be incorporated into the organism's genome. Additionally, multiple copies of the salvage pathway genes can be cloned into the host organism.

One salvage pathway that is conserved in all organisms recycles the various derivatives of vitamin B_6 to the active PLP form. The polypeptides involved in this pathway are *pdxK* kinase, *pdxH* oxidase, and *pdxY* kinase. Over-expression of one or more of these genes can increase PLP availability.

Vitamin B_6 levels can be elevated by elimination or repression of the metabolic regulation of the native biosynthetic pathway genes in the host organism. PLP represses polypeptides involved in the biosynthesis of the precursor threonine derivative in the bacterium *Flavobacterium sp.* strain 238-7. This bacterial strain, freed of metabolic control, overproduces pyridoxal derivatives and can excrete up to 20 mg/L of PLP. Genetic manipulation of the host organism producing monatin in a similar fashion will allow the increased production PLP without over-expression of the biosynthetic pathway genes.

Ammonium Utilization

Tryptophanase reactions can be driven toward the synthetic direction (production of tryptophan from indole) by making ammonia more available or by removal of water. Reductive amination reactions, such as those catalyzed by glutamate dehydrogenase, can also be driven forward by an excess of ammonium.

Ammonia can be made available as an ammonium carbonate or ammonium phosphate salt in a carbonate or phosphate buffered system. Ammonia can also be provided as ammonium pyruvate or ammonium formate. Alternatively, ammonia can be supplied if the reaction is coupled with a reaction that generates ammonia, such as glutamate dehydrogenase or tryptophan dehydrogenase. Ammonia can be generated by addition of the natural substrates of EC 4.1.99.- (tyrosine or tryptophan), which will be hydrolyzed to phenol or indole, pyruvate and NH_3. This also allows for an increased yield of synthetic product over the normal equilibrium amount by allowing the enzyme to hydrolyze its preferred substrate.

Removal of products and byproducts

The conversion of tryptophan to indole-3-pyruvate via a tryptophan aminotransferase may adversely affect the production rate of indole-3-pyruvate because the reaction produces glutamate and requires the co-substrate 2-oxoglutarate (α-ketoglutarate). Glutamate may cause inhibition of the aminotransferase, and the reaction will consume large amounts of the co-substrate. Moreover, high glutamate concentrations are detrimental to downstream separation processes.

The polypeptide glutamate dehydrogenase (GLDH) converts glutamate to 2-oxoglutarate, thereby recycling the co-substrate in the reaction catalyzed by tryptophan aminotransferase. GLDH also generates reducing equivalents (NADH or NADPH) that can be used to generate energy for the cell (ATP) under aerobic conditions. The utilization of glutamate by GLDH also reduces byproduct formation. Additionally, the reaction generates ammonia, which can serve as a nitrogen source for the cell or as a substrate in a reductive amination for the final step shown in FIG. 1. Therefore, a production cell that over-expresses a *GLDH* polypeptide can be used to increase the yield and reduce the cost of media and/or separation processes.

In the tryptophan to monatin pathway, the amino donor of step three (e.g., glutamate or aspartate) can be converted back to the amino acceptor required for step 1 (e.g., 2-oxo-glutarate or oxaloacetate), if an aminotransferase from the appropriate enzyme classes is used. Utilization of two

separate transaminases for this pathway, in which the substrate of one transaminase does not competitively inhibit the activity of the other transaminase, can increase the efficiency of this pathway.

Many of the reactions in the described pathways are reversible and will, therefore, reach an equilibrium between substrates and products. The yield of the pathway can be increased by continuous removal of the products from the polypeptides. For example, secretion of monatin into the fermentation broth using a permease or other transport protein, or selective crystallization of monatin from a biocatalytic reactor stream with concomitant recycle of substrates will increase the reaction yield.

The removal of byproducts by additional enzymatic reactions or by substitution of amino donor groups is another way to increase the reaction yield. Several examples are discussed in Example 13 and shown in FIGS. 11-13. Ideally a byproduct is produced that is unavailable to react in the reverse direction, either by phase change (evaporation) or by spontaneous conversion to an unreactive endproduct, such as carbon dioxide.

Modulation of the Substrate Pools

The indole pool can be modulated by increasing production of tryptophan precursors and/or altering catabolic pathways involving indole-3-pyruvate and/or tryptophan. For example, the production of indole-3-acetic acid from indole-3-pyruvate can be reduced or eliminated by functionally deleting the gene coding for EC 4.1.1.74 in the host cell. Production of indole from tryptophan can be reduced or eliminated by functionally deleting the gene coding for EC 4.1.99.1 in the host cell. Alternatively, an excess of indole can be utilized as a substrate in an *in vitro or in vivo* process in combination with increased amounts of the gene coding for EC 4.1.99.1 (Kawasaki *et al.*, *J. Ferm. and Bioeng.*, 82:604-6, 1996). Genetic modifications can be made to increase the level of intermediates such as D-erythrose-4-phosphate and chorismate.

Tryptophan production is regulated in most organisms. One mechanism is via feedback inhibition of certain enzymes in the pathway; as tryptophan levels increase, the production rate of tryptophan decreases. Thus, when using a host cell engineered to produce monatin via a tryptophan intermediate, an organism can be used that is not sensitive to tryptophan concentrations. For example, a strain of *Catharanthus roseus* that is resistant to growth inhibition by various tryptophan analogs was selected by repeated exposure to high concentrations of 5-methyltryptophan (Schallenberg and Berlin, *Z Naturforsch* 34:541-5, 1979). The resulting tryptophan synthase activity of the strain was less effected by product inhibition, likely due to mutations in the gene. Similarly, a host cell used for monatin production can be optimized.

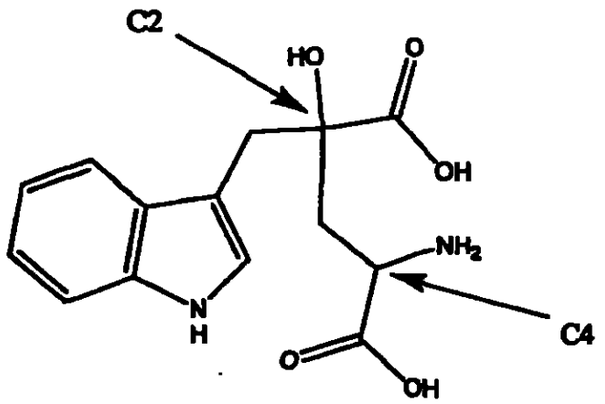
Tryptophan production can be optimized through the use of directed evolution to evolve polypeptides that are less sensitive to product inhibition. For example, screening can be performed on plates containing no tryptophan in the medium, but with high levels of non-metabolizable tryptophan analogs. U.S. Patent Nos. 5,756,345; 4,742,007; and 4,371,614 describe methods used to increase tryptophan productivity in a fermentation organism. The last step of tryptophan biosynthesis

is the addition of serine to indole; therefore the availability of serine can be increased to increase tryptophan production.

The amount of monatin produced by a fermentation organism can be increased by increasing the amount of pyruvate produced by the host organism. Certain yeasts, such as *Trichosporon cutaneum* (Wang *et al., Lett. Appl. Microbiol.* 35:338-42, 2002) and *Torulopsis glabrata* (Li *et al., Appl Microbiol. Biotechnol.* 57:451-9, 2001) overproduce pyruvate and can be used to practice the methods disclosed herein. In addition, genetic modifications can be made to organisms to promote pyruvic acid production, such as those in *E. coli* strain W1485*lip2* (Kawasaki *et al., J. Ferm. and Bioeng.* 82:604-6, 1996).

Controlling Chirality

The taste profile of monatin can be altered by controlling the stereochemistry (chirality) of the molecule. For example, different monatin isomers may be desired in different blends of concentrations for different food systems. Chirality can be controlled via a combination of pH and polypeptides.



Racemization at the C-4 position of monatin (see numbered molecule above) can occur by deprotonation and reprotonation of the alpha carbon, which can occur by a shift in pH or by reaction with the cofactor PLP. In a microorganism, the pH is unlikely to shift enough to cause the racemization, but PLP is abundant. Methods to control the chirality with polypeptides depend upon the biosynthetic route utilized for monatin production.

When monatin is formed using the pathway shown in FIG. 2, the following can be considered. In a biocatalytic reaction, the chirality of carbon-2 is determined by the enzyme that converts indole-3-pyruvate to MP. Multiple enzymes (e.g. from EC 4.1.2.-, 4.1.3.-) can convert indole-3-pyruvate to MP, thus, one can choose the enzyme that forms the desired isomer. Alternatively, the enantiospecificity of the enzyme that converts indole-3-pyruvate to MP can be modified through the use of directed evolution or catalytic antibodies can be engineered to catalyze the desired reaction. Once MP is produced (either enzymatically or by chemical condensation), the amino group can be added stereospecifically using a transaminase, such as those described herein.

Either the R or S configuration of carbon-4 can be generated depending on whether a D- or L-aromatic acid aminotransferase is used. Most aminotransferases are specific for the L-isomer, however D-tryptophan aminotransferases exist in certain plants (Kohiba and Mito, Proceedings of the 8th International Symposium on Vitamin B_6 and Carbonyl Catalysis, Osaka, Japan 1990). Moreover, D-alanine aminotransferases (2.6.1.21), D-methionine-pyruvate aminotransferases (2.6.1.41) and both (R)-3-amino-2-methylpropanoate aminotransferase (2.6.1.61) and (S)-3-amino-2-methylpropanoate aminotransferase (2.6.1.22) have been identified. Certain aminotransferases may only accept the substrate for this reaction with a particular configuration at the C2 carbon. Therefore, even if the conversion to MP is not stereospecific, the stereochemistry of the final product can be controlled through the appropriate selection of a transaminase. Since the reactions are reversible, the unreacted MP (undesired isomer) can be recycled back to its constituents and a racemic mixture of MP can be reformed.

Activation of substrates

Phosphorylated substrates, such as phosphoenolpyruvate (PEP), can be used in the reactions disclosed herein. Phosphorylated substrates can be more energetically favorable and, therefore, can be used to increase the reaction rates and/or yields. In aldol condensations, the addition of a phosphate group stabilizes the enol tautomer of the nucleophilic substrate, making it more reactive. In other reactions, a phosphorylated substrate often provides a better leaving group. Similarly, substrates can be activated by conversion to CoA derivatives or pyrophosphate derivatives.

Example 1

Cloning and Expression of Tryptophan Aminotransferases

This example describes methods that were used to clone tryptophan aminotransferases, which can be used to convert tryptophan to indole-3-pyruvate.

Experimental Overview

Eleven genes encoding aminotransferases were cloned into *E. coli*. These genes were *Bacillus subtilis* D-alanine aminotransferase (*dat*, Genbank Accession No. Y14082.1 bp 28622-29470 and Genbank Accession No. NP_388848.1, nucleic acid sequence and amino acid sequence, respectively), *Sinorhizobium meliloti* (also termed *Rhizobium meliloti*) tyrosine aminotransferase (*tatA*, SEQ ID NOS: 1 and 2, nucleic acid sequence and amino acid sequence, respectively), *Rhodobacter sphaeroides* strain 2.4.1 tyrosine aminotransferase (*tatA* asserted by homology, SEQ ID NOS: 3 and 4, nucleic acid sequence and amino acid sequence, respectively), *R. sphaeroides* 35053 tyrosine aminotransferase (asserted by homology, SEQ ID NOS: 5 and 6, nucleic acid sequence and amino acid sequence, respectively), *Leishmania major* broad substrate aminotransferase (*bsat*, asserted by homology to peptide fragments from *L. mexicana*, SEQ ID NOS: 7 and 8, nucleic acid sequence and amino acid sequence, respectively), *Bacillus subtilis* aromatic aminotransferase (*araT*, asserted by homology, SEQ ID NOS: 9 and 10, nucleic acid sequence and amino acid sequence,

respectively), *Lactobacillus amylovorus* aromatic aminotransferase (*araT* asserted by homology, SEQ ID NOS: 11 and 12, nucleic acid sequence and amino acid sequence, respectively), *R. sphaeroides* 35053 multiple substrate aminotransferase (asserted by homology, SEQ ID NOS: 13 and 14, nucleic acid sequence and amino acid sequence, respectively), *Rhodobacter sphaeroides* strain 2.4.1 multiple substrate aminotransferase (*mat* asserted by homology, Genbank Accession No. AAAE01000093.1, bp 14743-16155 and Genbank Accession No. ZP00005082.1, nucleic acid sequence and amino acid sequence, respectively), *Escherichia coli* aspartate aminotransferase (*aspC*, Genbank Accession No. AE000195.1 bp 2755-1565 and Genbank Accession No. AAC74014.1, nucleic acid sequence and amino acid sequence, respectively), and *E. coli* tyrosine aminotransferase (*tyrB*, SEQ ID NOS: 31 and 32, nucleic acid sequence and amino acid sequence, respectively).

The genes were cloned, expressed, and tested for activity in conversion of tryptophan to indole-3-pyruvate, along with commercially available enzymes. All eleven clones had activity.

Identification of Bacterial Strains that May Contain Polypeptides with the Desired Activity

No genes in the NCBI (National Center for Biotechnology Information) database were designated as tryptophan aminotransferases. However, organisms having this enzymatic activity have been identified. L-tryptophan aminotransferase (TAT) activity has been measured in cell extracts or from purified protein from the following sources: Rhizobacterial isolate from *Festuca octoflora*, pea mitochondria and cytosol, sunflower crown gall cells, *Rhizobium leguminosarum* biovar *trifolii*, *Erwinia herbicola* pv gypsophilae, *Pseudomonas syringae* pv. savastanoi, *Agrobacterium tumefaciens*, *Azospirillum lipferum & brasilense*, *Enterobacter cloacae*, *Enterobacter agglomerans*, *Bradyrhizobium elkanii*, *Candida maltosa*, *Azotobacter vinelandii*, rat brain, rat liver, *Sinorhizobium meliloti*, *Pseudomonas fluorescens* CHA0, *Lactococcus lactis*, *Lactobacillus casei*, *Lactobacillus helveticus*, wheat seedlings, barley, *Phaseolus aureus* (mung bean), *Saccharomyces uvarum* (carlsbergensis), *Leishmania* sp., maize, tomato shoots, pea plants, tobacco, pig, *Clostridium sporogenes*, and *Streptomyces griseus*.

Isolation of Genomic DNA for Cloning

S. meliloti (ATCC number 9930) was grown in TY media at 25°C, pH 7.2. Cells were grown to an optical density at 600 nm (OD_{600}) of 1.85 and a 2% inoculum was used for genomic DNA preparations. The Qiagen genomic tip 20/G kit (Valencia, CA) was used for genomic DNA isolation.

Bacillus subtilis 6051 (ATCC) was grown at 30°C in Bereto Nutrient Broth (Difco; Detroit, MI). The Qiagen genomic tip 20/G protocol was used to isolate the genomic DNA with the following changes: the concentrations of proteinase K and lysozyme were doubled and incubation times were increased 2-3 fold.

Leishmania major ATCC 50122 genomic DNA was supplied by IDI, Inc. (Quebec, Canada) in TE buffer pH 8.0, 17 ng/μL.

Rhodobacter sphaeroides 2.4.1 (provided by Professor Sam Kaplan, University of Texas, Houston), *R. sphaeroides* 35053 (ATCC number), and *L. amyloverus* genomic DNA was prepared by standard phenol extraction. Cells were harvested in late log phase, resuspended in TEN buffer (10 mM Tris-HCl, pH 7.5, 1 mM EDTA, 100 mM NaCl), and lysed by the addition of 0.024 mL sodium lauryl sarcosine per mL cell suspension. After extracting at least three times with an equal volume of phenol saturated with TE buffer (10 mM Tris-HCl, pH 7.5, 1 mM EDTA), the DNA solution was extracted once with 9:1 chloroform:octanol and three times with chloroform. The DNA was precipitated by the addition of 0.1 volume of 3 M sodium acetate, pH 6.8 and 2 volumes ethanol. The precipitate was collected by centrifugation and washed once with 70% ethanol. Finally the DNA was dissolved in 0.10 mL distilled water.

Escherichia coli genomic DNA was isolated from strain DH10B (Invitrogen) and prepared using the Qiagen Genomic-tip™ (500/G) kit. From 30 mL of this strain grown in LB to an OD_{650} of 1.87, 0.3 mg of purified DNA was obtained. The purified DNA was dissolved in Qiagen elution buffer (EB) at a concentration of 0.37 μg/μL.

Polymerase Chain Reaction Protocol

Primers were designed with compatible overhangs for the pET 30 Xa/LIC vector (Novagen, Madison, WI). The pET vector has a 12 base single stranded overhang on the 5' side of the Xa/LIC site and a 15-base single stranded overhang on the 3' side of the Xa/LIC site. The plasmid is designed for ligation independent cloning, with N-terminal His and S-tags and an optional C-terminal His-tag. The Xa protease recognition site (IEGR) sits directly in front of the start codon of the gene of interest, such that the fusion protein tags can be removed.

The following sequences were added to the 5' ends of the organism specific sequences when designing primers: forward primer, 5' GGTATTGAGGGTCGC (SEQ ID NO: 73); reverse primer: 5' AGAGGAGAGTTAGAGCC (SEQ ID NO: 74).

Bacillus subtilis dat primers: N term: 5'-GGTATTGAGGGTCGCATGAAGGTTTTAGTCAATGG-3' and C term: 5'-AGAGGAGAGTTAGAGCCTTATGAAATGCTAGCAGCCT-3' (SEQ ID NOS: 15 and 16).

Sinorhizobium meliloti tmA primers: N term: 5'-GGTATTGAGGGTCGCATGTTCGACGCCCTCGCCCG and C term: 5'-AGAGGAGAGTTAGAGCCTCAGAGACTGGTGAACTTGC (SEQ ID NOS: 17 and 18).

Bacillus subtilis araT primers: N term: 5'-GGTATTGAGGGTCGCATGGAACATTTGCTGAATCC and C term: 5'-AGAGGAGAGTTAGAGCCTTAAACGCCGTTGTTTATCG (SEQ ID NOS: 19 and 20).

Rhodobacter sphaeroides nusa (both 2.4.1 and 35053): N term: 5'-GGTATTGAGGGTCGCATGCGCGAGCCTCTTGCCCT and C term: 5'-AGAGGAGAGTTAGAGCCTCAGCCGGGGAAGCTCCGGG (SEQ ID NOS: 21 and 22).

Leishmania major hsat: N term: 5'-GGTATTGAGGGTCGCATGTCCACGCAGGCGGCCAT and C term: 5'-AGAGGAGAGTTAGAGCCTCACTCACGATTCACATTGC (SEQ ID NOS: 23 and 24).

Lactobacillus amyloverus araT: N term: 5'-GGTATTGAGGGTCGCATGCCAGAATTAGCTAATGA and C term: 5'-AGAGGAGAGTTAGAGCCTTATTCGTCCTCTTGTAAAA (SEQ ID NOS: 25 and 26).

Rhodobacter sphaeroides tatA (both 2.4.1 and 35053 strains): N term: 5'-GGTATTGAGGGTCGCATGCGCTCTACGACGGCTCC and C term: 5'-AGAGGAGAGTTAGAGCCTCAGCCGCGCAGCACCTTGG (SEQ ID NOS: 27 and 28).

Escherichia coli aspC: N term: 5'-GGTATTGAGGGTCGCATGTTTGAGAACATTACCGC-3' and C term: 5'-AGAGGAGAGTTAGAGCCTTACAGCACTGCCACAATCG-3' (SEQ ID NOS: 29 and 30).

Escherichia coli tyrB: N term: 5'-GGTATTGAGGGTCGCGTGTTTCAAAAAGTTGACGC and C term: 5'-AGAGGAGAGTTAGAGCCTTACATCACCGCAGCAAACG-3' (SEQ ID NOS: 33 and 34).

The gene derived from *S. meliloti* (*tatA*) was amplified using the following PCR protocol. In a 50 µL reaction 0.1-0.5 µg template, 1.5 µM of each primer, 0.4 mM each dNTP, 3.5 U Expand High Fidelity Polymerase (Roche, Indianapolis, IN), and 1X Expand™ buffer with Mg were used. The thermocycler program used included a hot start at 96°C for 5 minutes, followed by 29 repetitions of the following steps: 94°C for 30 seconds, 55°C for 2 minutes, and 72°C for 2.5 minutes. After the 29 repetitions the sample was maintained at 72°C for 10 minutes and then stored at 4°C. This PCR protocol produced a product of 1199 bp.

The sequences of the genes derived from *R. sphaeroides (msa* and *tatA)*, *L. amylovorus araT*, and *Bacillus araT* were amplified using the following PCR protocol. In a 50 µL reaction, 0.1-0.5 µg template, 1.5 µM of each primer, 0.4 mM each dNTP, 3.5 U Expand High Fidelity™ Polymerase, and 1X Expand™ buffer with Mg were added. The thermocycler program used included a hot start at 96°C for 5 minutes, followed by 29 repetitions of the following steps: 94°C for 30 seconds, 40-60°C for 1 minute, 45 seconds (gradient thermocycler) and 72°C for 2 minutes, 15 seconds. After the 29 repetitions the sample was maintained at 72°C for 10 minutes and then stored at 4°C.

For each *R. sphaeroides msa* gene, the 42°C and 48°C annealing temperatures produced multiple products, but a distinct band at approximately 1464 bp. For *L. amylovorus araT*, the 42°C, 48°C, and 56°C annealing temperatures yielded single products with intense bands at 1173 bp. For *B. subtilis araT*, the 40°C, 45°C, 50°C, 55°C annealing temperatures generated single intense products (1173 bp), from both genomic DNA and colonies. For *L. major hsat*, the 55°C annealing temperature gave the cleanest product (1239 bp). For *Rhodobacter tatA* genes, the 50-55°C annealing temperatures gave clean products at the correct size (1260 bp). For both *E. coli* genes and the *B. subtilis dat* gene, an annealing temperature of 55-60°C was used, and the annealing time was

shortened to 45 seconds. Clean products of the correct sizes were obtained (approximately 1.3 kb for the *E. coli* genes, 850 bp for the *dat* gene).

Cloning

The PCR products were gel purified from 0.8 or 1% TAE-agarose gels using the Qiagen gel extraction kit (Valencia, CA). The PCR products were quantified by comparison to standards on an agarose gel, and then treated with T4 DNA polymerase following the manufacturer's recommended protocols for Ligation Independent Cloning (Novagen, Madison, WI).

Briefly, approximately 0.2 pmol of purified PCR product was treated with 1 U T4 DNA polymerase in the presence of dGTP for 30 mimtes at 22°C. The polymerase removes successive bases from the 3' ends of the PCR product. When the polymerase encounters a guanine residue, the 5' to 3' polymerase activity of the enzyme counteracts the exonuclease activity to effectively prevent further excision. This creates single stranded overhangs that are compatible with the pET Xa/LIC vector. The polymerase is inactivated by incubating at 75°C for 20 minutes.

The vector and treated insert were annealed as recommended by Novagen. Approximately 0.02 pmol of treated insert and 0.01 pmol vector were incubated for 5 minutes at 22°C, 6.25 mM EDTA (final concentration) was added, and the incubation at 22°C was repeated. The annealing reaction (1 µL) was added to NovaBlue™ singles competent cells (Novagen, Madison, WI), and incubated on ice for 5 minutes. After mixing, the cells were transformed by heat shock for 30 seconds at 42°C. The cells were placed on ice for 2 minutes, and allowed to recover in 250 µL of room temperature SOC for 30 minutes at 37°C with shaking at 225 rpm. Cells were plated on LB plates containing kanamycin (25-50 µg/mL).

Plasmid DNA was purified using the Qiagen spin miniprep kit and screened for the correct inserts by restriction digest with *Xho*I and *Xba*I. The sequences of plasmids that appeared to have the correct insert were verified by dideoxy chain termination DNA sequencing.

SEQ ID NOS: 1-14 and 31-32 show nucleotide and corresponding amino acid sequences of the recombinant aminotransferases, any changes from the Genbank sequences were either silent or generated conservative substitutions in the protein sequence. SEQ ID NOS: 11 and 12 are novel sequences.

Gene Expression and Assays

Plasmid DNA, verified by sequence analysis, was subcloned into *E. coli* expression hosts BLR(DE3) or BL21(DE3) (Novagen, Madison, WI). The cultures were grown and the plasmids were isolated using Qiagen miniprep kit, and analyzed by restriction digest to confirm identity.

Induction was initially performed with *L. amylovorus araT*, *B. subtilis araT*, and *S. meliloti tatA* in both BLR(DE3) and BL21(DE3) cells. A time course study was performed with cultures grown in LB containing kanamycin (30 mg/L) to an OD_{600} of 0.5-0.8 and induced with 1 mM IPTG (isopropyl thiogalacatoside) and sampled at 0, 1, 2, and 4 hours post induction. Cells from 2.0 mL were resuspended in 0.10 mL 120 mM Tris-HCl, pH 6.8 containing 10% sodium dodecyl sulfate,

10% 2-mercaptoethanol, and 20% glycerol, heated at 95°C for 10 min, and cooled, and diluted with 0.10 mL H_2O. Aliquots of these total cellular protein samples were analyzed by SDS-PAGE using a 4-15% gradient gel. There were no significant differences in the amount of protein expressed between the 2 hour and 4 hour induction, nor between the BLR(DE3) and BL21(DE3) cells.

Cell extracts were also prepared from the 4 hour samples by suspending cell pellets from 2 mL of culture in 0.25 mL Novagen BugBuster™ reagent containing 0.25 µL benzonase nuclease, incubating at room temperature for 20 minutes with gentle shaking, and centrifuging at 16,000 x g to remove cell debris. The supernatants (cell extracts) were loaded onto 4-15% gradient gels for analysis of the cellular soluble proteins.

The three clones, (*L. amylovorus araT* (SEQ ID NOS: 11 and 12), *B. subtilis araT* (SEQ ID NOS: 9 and 10), and *S. meliloti tatA* (SEQ ID NOS: 1 and 2) showed soluble protein that corresponded to the correct size (approximately 45 kDa). The *B. subtilis araT* gene product was over-expressed at the highest level and/or was more soluble than the other two gene products.

In subsequent expression methods, plasmid DNA from positive clones was subcloned into BL21(DE3) due to the better growth characteristics of this host. Induction was repeated using 1 mM IPTG with cultures grown in LB containing kanamycin at 50 mg/L, inducing when the OD_{600} reached approximately 0.8. Cells were harvested after 4 hours of growth at 37°C, centrifuged at 3000 rpm for 10 minutes (4°C), washed with TEGGP buffer (50 mM Tris-HCl (pH 7.0), 0.5 mM EDTA, 100 mg/L glutathione, 5% glycerol, with Roche complete protease inhibitor cocktail), and flash frozen in -80°C ethanol.

Samples were resuspended in 5 mL/g wet cell weight of BugBuster™ (Novagen) reagent containing 5 µL/mL protease inhibitor cocktail set #3 (Calbiochem-Novabiochem Corp., San Diego, CA) and 1 µL/mL benzonase nuclease. Samples were incubated at room temperature for 20 minutes on an orbital shaker. Insoluble cell debris was removed by centrifugation at 16,000 X g for 20 minutes at 4°C.

Cell extracts were analyzed by SDS-PAGE, and assayed for tryptophan aminotransferase activity by following production of indole-pyruvic acid using the following protocol. One mL reactions were carried out in 50 mM sodium tetraborate (pH 8.5), 0.5 mM EDTA, 0.5 mM sodium arsenate, 50 µM pyridoxal phosphate, 5 mM α-ketoglutarate, and 5 mM L-tryptophan. The reactions were initiated by the addition of cell free extracts or purified enzyme and were incubated 30 minutes at 30°C. 20% TCA (200 µL) was added to stop the reaction, and the precipitated protein was removed by centrifugation. The absorbance at 327 nm was measured and compared to a standard curve of freshly prepared indole-3-pyruvate in the assay buffer. Control reactions without the substrate tryptophan or using cell-free extracts from clones transformed with pET30a alone were also performed.

Due to background from the native *E. coli* aminotransferases in cell extracts, recombinant proteins were purified using His-Bind cartridges following manufacturer's protocols (Novagen, Madison, WI). The eluent fractions were desalted on PD-10 (Amersham Biosciences, Piscataway,

NJ) columns and eluted in 50 mM Tris, pH 7.0. Purified proteins were analyzed by SDS-PAGE and assayed for aminotransferase activity.

Results from the 37°C induction with 1 mM IPTG (4 hours) demonstrate that *L. major bsat*, *S. meliloti tatA*, *E. coli aspC*, and both *R. sphaeroides tatA* clones have significant levels of tryptophan aminotransferase activity. The *araT* protein from *B. subtilis* was over-expressed and soluble, but showed little enzymatic activity. The *L. amylovorus araT* gene product appeared to be soluble in the cell extract, but purification using a His-Bind cartridge resulted in only small amounts of protein with the correct molecular weight. The *msa* gene products were insoluble and further expression experiments were done at 24°C to minimize inclusion body formation. Several concentrations of IPTG between 10 μM and 1 mM were used to maximize the amount of soluble protein.

Table 1 lists the specific activities measured in micrograms of indole-3-pyruvate (I3P) formed per milligram protein per minute. In some cases, very small amounts of recombinant protein showed high levels of activity above the effective linear range of the assay. In these cases a '>' precedes the specific activity number.

Table 1: Specific Activities of Clones in Cell Extracts (CE) and Purified (P) and Commercial Enzymes

Enzyme	Specific Activity (μg I3P/mg protein/min)	Note
L. major bsat CE	>49.3	
L. major bsat P	>4280	
S. meliloti tatA CE	>28.6	
S. meliloti tatA P	>931	
2.4.1 *tatA* CE	>41.2	
2.4.1 *tatA* P	1086	
35053 *tatA* CE	>62.3	
35053 *tatA* P	>486	
L. amylovorus araT CE	1.26	
L. amylovorus araT P	0	little protein after His-Bind cartridge
B. subtilis araT CE	0	undetectable
B. subtilis araT P	1.5-4.5	
2.4.1 *msa* CE	2.05	very little soluble protein
2.4.1 *msa* P	0	no protein after His-Bind cartridge
35053 *msa* CE	3.97	very little soluble protein
35053 *msa* P	0	no protein after His-Bind cartridge
E. coli aspC (P)	800	
E. coli tyrB (P)	1	not very soluble
B. subtilis D-aminotransf.(P)	2.7	using D-tryptophan as substrate
broad range transaminase	22	Sigma cat # T 7684
Porcine type II-A	1.5	Sigma G7005
Porcine type I	1	Sigma G2751

An alignment comparing all of the recombinant proteins cloned illustrates that there are not many highly conserved areas between the *araT*, *tatA*, *bsat*, and *msa sequences*. An alignment of

highest activity recombinant proteins: *Rhodobacter tatA* gene product homologs, *L. major* broad substrate aminotransferase, and the *Sinorhizobium meliloti* tyrosine aminotransferase showed several conserved regions, however they are only approximately 30-43% identical at the protein level. The availability of the broad range, D-specific (D-alanine) aminotransferase can be useful in the production of other stereoisomers of monatin.

Example 2

Conversion of Indole-3-lactate to Indole-3-pyruvate

As shown in FIGS. 1 and 3, indole-3-lactic acid can be used to produce indole-3-pyruvate. Conversion between lactic acid and pyruvate is a reversible reaction, as is conversion between indole-3-pyruvate and indole-3-lactate. The oxidation of indole-lactate was typically followed due to the high amount of background at 340 nm from indole-3-pyruvate.

The standard assay mixture contained 100 mM potassium phosphate, pH 8.0, 0.3 mM NAD^+, 7 units of lactate dehydrogenase (LDH) (Sigma-L2395, St. Louis, MO), and 2 mM substrate in 0.1 mL. The assay was performed in duplicate in a UV-transparent microtiter plate, using a Molecular Devices SpectraMax Plus platereader. Polypeptide and buffer were mixed and pipetted into wells containing the indole-3-lactic acid and NAD^+ and the absorbance at 340 nm of each well was read at intervals of 9 seconds after brief mixing. The reaction was held at 25°C for 5 minutes. The increase in absorbance at 340 nm follows the production of NADH from NAD^+. Separate negative controls were performed without NAD^+ and without substrate. D-LDH from *Leuconostoc mesenteroides* (Sigma catalog number L2395) appeared to exhibit more activity with the indole-derivative substrates than did *L-LDH from Bacillus stearothermophilus* (Sigma catalog number L5275).

Similar methods were utilized with D-lactic acid and NAD+ or NADH and pyruvate, the natural substrates of D-LDH polypeptides. The V_{max} for the reduction of pyruvate was 100-1000 fold higher than the V_{max} for the oxidation of lactate. The V_{max} for the oxidation reaction of indole-3-lactic with *D-LDH* was approximately one-fifth of that with lactic acid. The presence of indole-3-pyruvate was also measured by following the change in absorbance at 327 (the enol-borate derivative) using 50 mM sodium borate buffer containing 0.5 mM EDTA and 0.5 mM sodium arsenate. Small, but repeatable, absorbance changes were observed, as compared to the negative controls for both L and D-LDH polypeptides.

Additionally, broad specificity lactate dehydrogenases (enzymes with activity associated with EC 1.1.1.27, EC 1.1.1.28, and/or EC 1.1.2.3), can be cloned and used to make indole-3-pyruvate from indole-3-lactic acid. Sources of broad specificity dehydrogenases include *E. coli, Neisseria gonorrhoeae*, and *Lactobacillus plantarum*.

Alternatively, indole-3-pyruvate can be produced by contacting indole-3-lactate with cellular extracts from *Clostridium sporogenes* which contain an indolelactate dehydrogenase (EC 1.1.1.110); or *Trypanosoma cruzi epimastigotes* cellular extracts which contain *p*-hydroxyphenylactate dehydrogenase (EC 1.1.1.222) known to have activity on indole-3-pyruvate; or *Pseudomonas*

acidovorans or *E. coli* cellular extracts, which contain an imidazol-5-yl lactate dehydrogenase (EC 1.1.1.111); or *Coleus blumei*, which contains a hydroxyphenylpyruvate reductase (EC 1.1.1.237); or *Candida maltosa* which contains a D-aromatic lactate dehydrogenase (EC 1.1.1.222). References describing such activities include, Nowicki *et al.* (*FEMS Microbiol Lett* 71:119-24, 1992), Jean and DeMoss (*Canadian J. Microbiol.* 14 1968, Coote and Hassall (*Biochem. J.* 111: 237-9, 1969), Cortese *et al.* (*C.R. Seances Soc. Biol. Fil.* 162 390-5, 1968), Petersen and Alfermann (*Z. Naturforsch. C: Biosci.* 43 501-4, 1988), and Bhatnagar *et al.* (*J. Gen Microbiol* 135:353-60, 1989). In addition, a lactate oxidase such as the one from *Pseudomonas sp.* (Gu *et al.* J. Mol. Catalysis B: Enzymatic: 18:299-305, 2002), can be utilized for oxidation of indole-3-lactic to indole-3-pyruvate.

Example 3

Conversion of L-tryptophan to Indole-3-pyruvate utilizing L-amino acid oxidase

This example describes methods used to convert tryptophan to indole-3-pyruvate via an oxidase (EC 1.4.3.2), as an alternative to using a tryptophan aminotransferase as described in Example 1. L-amino acid oxidase was purified from *Crotalus durissus* (Sigma, St. Louis, MO, catalog number A-2805). The accession numbers of L-amino acid oxidases for molecular cloning include: CAD21325.1, AAL14831, NP_490275, BAB78253, A38314, CAB71136, JE0266, T08202, S48644, CAC00499, P56742, P81383, O93364, P81382, P81375, S62692, P23623, AAD45200, AAC32267, CAA88452, AP003600, and Z48565.

Reactions were performed in microcentrifuge tubes in a total volume of 1 mL, incubated for 10 minutes while shaking at 37°C. The reaction mix contained 5 mM L-tryptophan, 100 mM sodium phosphate buffer pH 6.6, 0.5 mM sodium arsenate, 0.5 mM EDTA, 25 mM sodium tetraborate, 0.016 mg catalase (83 U, Sigma C-3515), 0.008 mg FAD (Sigma), and 0.005-0.125 Units of L-amino acid oxidase. Negative controls contained all components except tryptophan, and blanks contained all components except the oxidase. Catalase was used to remove the hydrogen peroxide formed during the oxidative deamination. The sodium tetraborate and arsenate were used to stabilize the enol-borate form of indole-3-pyruvate, which shows a maximum absorbance at 327 nm. Indole-3-pyruvate standards were prepared at concentrations of 0.1-1 mM in the reaction mix.

The purchased L-amino acid oxidase had a specific activity of 540 μg indole-3-pyruvate formed per minute per mg protein. This is the same order of magnitude as the specific activity of tryptophan aminotransferase enzymes.

Example 4

Converting Indole-3-pyruvate to 2-hydroxy 2-(indol-3-ylmethyl)-4-keto glutaric acid with an Aldolase

This example describes methods that can be used to convert indole-3-pyruvate to the 2-hydroxy 2-(indol-3-ylmethyl)-4-keto glutaric acid monatin precursor (MP) using an aldolase (lyase) (FIG. 2). Aldol condensations are reactions that form carbon-carbon bonds between the β-carbon of an aldehyde or ketone and the carbonyl carbon of another aldehyde or ketone. A carbanion is formed

on the carbon adjacent to the carbonyl group of one substrate, and serves as a nucleophile attacking the carbonyl carbon of the second substrate (the electrophilic carbon). Most commonly, the electrophilic substrate is an aldehyde, so most aldolases fall into EC 4.1.2.- category. Quite often, the nucleophilic substrate is pyruvate. It is less common for aldolases to catalyze the condensation between two keto-acids or two aldehydes.

However, aldolases that catalyze the condensation of two carboxylic acids have been identified. For example, EP 1045-029 describes the production of L-4-hydroxy-2-ketoglutaric acid from glyoxylic acid and pyruvate using a *Pseudomonas* culture (EC 4.1.3.16). In addition, 4-hydroxy-4-methyl-2-oxoglutarate aldolase (4-hydroxy-4-methyl-2-oxoglutarate pyruvate lyase, EC 4.1.3.17) can catalyze the condensation of two keto acids. Therefore, similar aldolase polypeptides were used to catalyze the condensation of indole-3-pyruvate with pyruvate.

Cloning

4-Hydroxy-4-methyl-2-oxoglutarate pyruvate lyases (ProA aldolase, EC 4.1.3.17) and 4-hydroxy-2-oxoglutarate glyoxylate-lyase (KHG aldolase, EC 4.1.3.16) catalyze reactions very similar to the aldolase reaction of FIG. 2. Primers were designed with compatible overhangs for the pET30 Xa/LIC vector (Novagen, Madison, WI). The design of these primers is described above in Example 1.

The following primers were designed for pET30 Xa/LIC cloning:

1. *Pseudomonas stramineae proA* gene (Genbank Accession No.: 12964663 Version: 12964663) and *Comamonas testosteroni proA* gene (SEQ ID NOS: 65-66, nucleic acid sequence and amino acid sequence, respectively) forward 5'-GGTATTGAGGGTCGCATGTACGAACTGGGAGTTGT-3' and reverse 5'-AGAGGAGAGTTAGAGCCTTAGTCAATATATTTCAGGC-3' (SEQ ID NOS: 55 and 56).

2. *Sinorhizobium meliloti* 1021 SMc00502 gene (homologous to *proA*, Genbank Accession Nos.: 15074579 and CAC46344, nucleic acid sequence and amino acid sequence, respectively) forward 5'-GGTATTGAGGGTCGCATGAGCGTGGTTCACCGGAA-3' and reverse 5'-AGAGGAGAGTTAGAGCCTCAATCGATATATTTCAGTC-3' (SEQ ID NOS: 61 and 62).

3. *Sphingomonas sp.* LB126 fldZ gene (Genbank Accession No.: 7573247 Version: 7573247, codes for a putative acyl transferase) forward 5'-GGTATTGAGGGTCGCATGTCCGGCATCGTTGTCCA-3' and reverse 5'-AGAGGAGAGTTAGAGCCTCAGACATATTTCAGTCCCA-3' (SEQ ID NOS: 57 and 58).

4. *Arthrobacter keyseri* pcmB gene (Genbank Accession No.: AF331043 Version: AF331043.1, codes for an oxalocitramalate aldolase) forward 5'-GGTATTGAGGGTCGCATGCGACTGAACAACCTCGG-3' and reverse 5'-AGAGGAGAGTTAGAGCCTCAGTTCTCCACGTATTCCA-3' (SEQ ID NOS: 59 and 60).

5. *Yersinia pestis strain* CO92 YPO0082 gene (Genbank Accession No.: 15978115 Version: 15978115, codes for a possible transferase) forward 5'-

GGTATTGAGGGTCGCATGAGCCTGGTTAATATGAA-3' and reverse 5'-AGAGGAGAGTTAGAGCCTTATGACTTTAACGCGTTGA-3' (SEQ ID NOS: 63 and 64).

6. *Bacillus subtilis kbg* gene (Genbank Accession Nos. Z99115.1 GI:2634478, 126711-127301and CAB14127.1, nucleic acid sequence and amino acid sequence, respectively) forward 5'-GGTATTGAGGGTCGCATGGAGTCCAAAGTCGTTGA-3' and reverse 5'-AGAGGAGAGTTAGAGCCTTACACTTGGAAAACAGCCT-3' (SEQ ID NOS: 35 and 36).

7. *E. coli kbg* gene (Genbank Accession Nos. AE000279.1 1331-1972 and AAC74920.1, nucleic acid and amino acid sequence, respectively) forward 5'-GGTATTGAGGGTCGCATGAAAAACTGGAAAACAAG-3' and reverse 5'-AGAGGAGAGTTAGAGCCTTACAGCTTAGCGCCTTCTA-3' (SEQ ID NOS: 37 and 38).

8. *S. meliloti kbg gene* (Genbank Accession Nos. AL591792.1 GI:15075850, 65353-64673 and CAC47463.1, nucleic acid and amino acid sequence, respectively) forward 5'-GGTATTGAGGGTCGCATGCGAGGGGCATTATTCAA-3' and reverse 5'-AGAGGAGAGTTAGAGCCTCAGCCCTTGAGCGCGAAG-3' (SEQ ID NOS: 39 and 40).

Genomic DNA from the organisms described in 1-2 and 6-8, above, was purified using the Qiagen genomic-tip protocol. Using similar techniques the genomic DNA from organisms described in 3-5 can be purified.

Pseudomonas straminea (ATCC 33636) was grown at 30°C in Nutrient Broth and hydroxybenzoate medium. *Comamonas testosteroni* (ATCC 49249) was grown at 26°C in Nutrient Broth and hydroxybenzoate medium. *Sphingomonas* sp. LB126 (Flemish Institute for Technological Research,VITO, B-2400 Mol, Belgium) is grown according to the method described by Wattiau *et al.* (*Research in Microbiol.* 152:861-72, 2001). *Arthrobacter keyseri* (Gulf Ecology Division, National Health and Environmental Effects Research Laboratory, U.S. Environmental Protection Agency, Gulf Breeze, FL 32561, USA) is grown according to the protocol described by Eaton (*J. Bacteriol.* 183:3689-3703, 2001). *Sinorhizobium meliloti* 1021 (ATCC 51124) was grown at 26°C in ATCC TY medium and hydroxybenzoate medium. *Yersinia pestis* strain CO92 (ATCC) is grown at 26°C in ATCC medium 739 Horse blood agar. *Bacillus subtilis* 6051 (ATCC) was grown at 30°C in Bacto Nutrient Broth (Difco; Detroit, MI). *E. coli* genomic DNA was isolated from strain DH10B (Invitrogen) as described in Example 1.

The PCR, cloning, and screening protocols described in Example 1 were used to clone the *C. testosteroni* and the *S. meliloti pvdA* sequences, as well as the *E. coli*, *B. subtilis*, and *S. meliloti kbg* sequences. The same methods can be used to clone the other sequences described above.

Positive clones were sequenced using dideoxy chain termination sequencing (Seqwright, Houston, TX) with S-tag and T7 terminator primers (Novagen), and internal primers from Integrated DNA Technologies, Inc. (Coralville, IA).

Expression and Activity Assays

Plasmid DNA (verified by sequence analysis) was subcloned into expression host BL21(DE3) (Novagen). The cultures were grown in LB medium with 50 mg/L kanamycin, the

plasmids isolated using a Qiagen spin plasmid miniprep kit and subsequently analyzed by restriction digest to confirm identity. Induction experiments were done with the BL21(DE3) constructs grown in LB medium containing 50 mg/L kanamycin at 37°C. Protein expression was induced using 0.1 mM IPTG after the OD_{600} reached approximately 0.6. The cells were grown for 4 hours at 30°C and harvested by centrifugation. The cells were then lysed using Bugbuster™ reagent (Novagen) and the His-tag recombinant proteins were purified using His-Bind cartridges as described above (Example 1). Purified proteins were desalted on PD-10 disposable columns and eluted in 50 mM Tris-HCl buffer, pH 7.3 with 2 mM $MgCl_2$.

The proteins were analyzed by SDS-PAGE on 4-15% gradient gels to detect soluble protein levels at the predicted MW of the recombinant fusion protein.

The proteins were assayed for activity using indole-3-pyruvate and sodium pyruvate as substrates. The assay mixture contained 100 mM Tris-HCl (pH 7-pH 8.9), 0-8 mM $MgCl_2$, 3 mM potassium phosphate (pH 8), and 6 mM of each substrate in 1 mL. The reaction was started by adding varying amounts of polypeptide (for example from 10 to 100 µg), and was incubated at 25°C-37°C for 30 minutes, filtered, and then frozen at -80 °C.

Activity Results with *proA* gene products

Both the *C. testosteroni proA* and *S. meliloti SMc00502* gene constructs had high levels of expression when induced with IPTG. The recombinant proteins were highly soluble, as determined by SDS-PAGE analysis of total protein and cellular extract samples. The *C. testosteroni* gene product was purified to > 95% purity. Because the yield of the *S. meliloti* gene product was very low after affinity purification using a His-Bind cartridge, cellular extract was used for the enzymatic assays.

Both recombinant aldolases catalyzed the formation of MP from indole-3-pyruvate and pyruvate. The presence of both divalent magnesium and potassium phosphate were required for enzymatic activity. No product was apparent when indole-3-pyruvate, pyruvate, or potassium phosphate was absent. A small amount of the product was also formed in the absence of enzyme (typically one order of magnitude less than when enzyme was present).

The product peak eluted from the reverse phase C18 column slightly later than the indole-3-pyruvate standard, the mass spectrum of this peak showed a collisionally-induced parent ion ([M + H]+) of 292.1, the parent ion expected for the product MP. The major daughter fragments present in the mass spectrum included those with m/z =158 (1*H*-indole-3-carbaldehyde carbonium ion), 168 (3-buta-1,3-dienyl-1*H*-indole carbonium ion), 274 (292 - H_2O), 256 (292 - 2 H_2O), 238 (292 - 3 H2O), 228 (292 - CH4O3), and 204 (loss of pyruvate). The product also exhibited a UV spectrum characteristic of other indole-containing compounds such as tryptophan, with the λ_{max} of 279-280 and a small shoulder at approximately 290 nm.

The amount of MP produced by the *C. testosteroni* aldolase increased with an increase in reaction temperature from room temperature to 37°C, amount of substrate, and amount of magnesium. The synthetic activity of the enzyme decreased with increasing pH, the maximum

product observed was at pH 7. Based on tryptophan standards, the amount of MP produced under a standard assay using 20 µg of purified protein was approximately *10-40 µg per one mL* reaction.

Due to the high degree of homology of the *S. meliloti* and *C. testosteroni* ProA aldolase coding sequences with the other genes described above, it is expected that all of the recombinant gene products can catalyze this reaction. Moreover, it is expected that aldolases that have threonine (T) at positions 59 and 87, arginine (R) at 119, aspartate (D) at 120, and histidine (H) at 31 and 71, (based on the numbering system of *C. testosteroni*) will have similar activity.

Activity Results with *khg* gene products

Both the *B. subtilis* and *E. coli khg* gene constructs had high levels of expression of protein when induced with IPTG, while the *S. meliloti khg* had a lower level of expression. The recombinant proteins were highly soluble, as judged by SDS-PAGE analysis of total proteins and cellular extracts. The *B. subtilis* and *E. coli khg* gene products were purified to > 95% purity; the yield of the *S. meliloti* gene product was not as high after affinity purification using a His-Bind cartridge.

There is no evidence that magnesium and phosphate are required for activity for this enzyme. However, the literature reports performing the assays in sodium phosphate buffer, and the enzyme reportedly is bifunctional and has activity on phosphorylated substrates such as 2-keto-3-deoxy-6-phosphogluconate (KDPG). The enzymatic assays were performed as described above, and in some instances the phosphate was omitted. The results indicate that the recombinant KHG aldolases produced MP, but were not as active as the ProA aldolases. In some cases the level of MP produced by KHG was almost identical to the amount produced by magnesium and phosphate alone. Phosphate did not appear to increase the KHG activities. The *Bacillus* enzyme had the highest activity, approximately 20-25% higher activity than the magnesium and phosphate alone, as determined by SRM (see Example 10). The *Sinorhizobium* enzyme had the least amount of activity, which may be associated with folding and solubility problems noted in the expression. All three enzymes have the active site glutamate (position 43 in *B. subtilis* numbering system) as well as the lysine required for Shiff base formation with pyruvate (position 130); however, the *B. subtilis* enzyme contains a threonine in position 47, an active site residue, rather than arginine. The *B. subtilis* KHG is smaller and appears to be in a cluster distinct from the *S. meliloti* and *E. coli* enzymes, with other enzymes having the active site threonine. The differences in the active site may be the reason for the increased activity of the *B. subtilis* enzyme.

Improvement of Aldolase Activity

Catalytic antibodies can be as efficient as natural aldolases, accept a broad range of substrates, and can be used to catalyze the reaction shown in FIG. 2.

Aldolases can also be improved by directed evolution, for example as previously described for a KDPG aldolase (highly homologous to KHG described above) evolved by DNA shuffling and error-prone PCR to remove the requirement for phosphate and to invert the enantioselectivity. The KDPG aldolase polypeptides are useful in biochemical reactions since they are highly specific for the donor substrate (herein, pyruvate), but are relatively flexible with respect to the acceptor substrate

(i.e. indole-3-pyruvate) (Koeller & Wong, *Nature* 409:232-9, 2001). KHG aldolase has activity for condensation of pyruvate with a number of carboxylic acids. Mammalian versions of the KHG aldolase are thought to have broader specificity than bacterial versions, including higher activity on 4-hydroxy 4-methyl 2-oxoglutarate and acceptance of both stereoisomers of 4-hydroxy-2-ketoglutarate. Bacterial sources appear to have a 10-fold preference for the R isomer. There are nearly 100 KHG homologs available in genomic databases, and activity has been demonstrated in *Pseudomonas, Paracoccus, Providencia, Sinorhizobium, Morganella, E. coli,* and mammalian tissues. These enzymes can be used as a starting point for tailoring the enantiospecificity that is desired for monatin production.

Aldolases that utilize pyruvate and another substrate that is either a keto acid and/or has a bulky hydrophobic group like indole can be "evolved" to tailor the polypeptide's specificity, speed, and selectivity. In addition to KHG and ProA aldolases demonstrated herein, examples of these enzymes include, but are not limited to: KDPG aldolase and related polypeptides (KDPH); transcarboxybenzalpyruvate hydratase-aldolase from *Nocardioides* sp; 4-(2-carboxyphenyl)-2-oxobut-3-enoate aldolase (2'-carboxybenzalpyruvate aldolase) which condenses pyruvate and 2-carboxybenzaldehyde (an aromatic ring-containing substrate); trans-O-hydroxybenzylidenepyruvate hydratase-aldolase from *Pseudomonas putida* and *Sphingomonas aromaticivorans*, which also utilizes pyruvate and an aromatic-containing aldehyde as substrates; 3-hydroxyaspartate aldolase (erythro-3-hydroxy-L-aspartate glyoxylate lyase), which uses 2-oxo acids as the substrates and is thought to be in the organism *Micrococcus dentrificans*; benzoin aldolase (benzaldehyde lyase), which utilizes substrates containing benzyl groups; dihydroneopterin aldolase; L-threo-3-phenylserine benzaldehyde-lyase (phenylserine aldolase) which condenses glycine with benzaldehyde; 4-hydroxy-2-oxovalerate aldolase; 1,2-dihydroxybenzylpyruvate aldolase; and 2-hydroxybenzalpyruvate aldolase.

A polypeptide having the desired activity can be selected by screening clones of interest using the following methods. Tryptophan auxotrophs are transformed with vectors carrying the clones of interest on an expression cassette and are grown on a medium containing small amounts of monatin or MP. Since aminotransferases and aldolase reactions are reversible, the cells are able to produce tryptophan from a racemic mixture of monatin. Similarly, organisms (both recombinant and wildtype) can be screened by ability to utilize MP or monatin as a carbon and energy source. One source of target aldolases is expression libraries of various *Pseudomonas* and rhizobacterial strains. Pseudomonads have many unusual catabolic pathways for degradation of aromatic molecules and they also contain many aldolases; whereas the rhizobacteria contain aldolases, are known to grow in the plant rhizosphere, and have many of the genes described for construction of a biosynthetic pathway for monatin.

Example 5

Chemical Synthesis of the Monatin Precursor

Example 4 described a method of using an aldolase to convert indole-3-pyruvate to the 2-hydroxy 2-(indol-3-ylmethyl)-4-keto glutaric acid monatin precursor (MP). This example describes an alternative method of chemically synthesizing MP.

The MP is formed by using a typical aldol-type condensation (FIG. 4). Briefly, a typical aldol-type reaction involves the generation of a carbanion of the pyruvate ester using a strong base, such as LDA (lithium diisopropylamide), lithium hexamethyldisilazane or butyl lithium. The carbanion that is generated reacts with the indole-pyruvate to form the coupled product.

Protecting groups that can be used for protecting the indole nitrogen include, but are not limited to: t-butyloxycarbonyl (Boc), and benzyloxycarbonyl (Cbz). Blocking groups for carboxylic acids include, but are not limited to, alkyl esters (for example, methyl, ethyl, benzyl esters). When such protecting groups are used, it is not possible to control the stereochemistry of the product that is formed. However, if R2 and/or R3 are chiral protecting groups (FIG. 4), such as (S)-2-butanol, menthol, or a chiral amine, this can favor the formation of one MP enantiomer over the other.

Example 6

Conversion of Tryptophan or Indole-3-Pyruvate to Monatin

An *in vitro* process utilizing two enzymes, an aminotransferase and an aldolase, produced monatin from tryptophan and pyruvate. In the first step alpha-ketoglutarate was the acceptor of the amino group from tryptophan in a transamination reaction generating indole-3-pyruvate and glutamate. An aldolase catalyzed the second reaction in which pyruvate was reacted with indole-3-pyruvate, in the presence of Mg^{2+} and phosphate, generating the alpha-keto derivative of monatin (MP), 2-hydroxy-2-(indol-3-ylmethyl)-4-ketoglutaric acid. Transfer of the amino group from the glutamate formed in the first reaction produced the desired product, monatin. Purification and characterization of the product established that the isomer formed was S,S-monatin. Alternative substrates, enzymes, and conditions are described as well as improvements that were made to this process.

Enzymes

The aldolase, 4-hydroxy-4-methyl-2-oxoglutarate pyruvate lyase (ProA aldolase, *proA* gene) (EC 4.1.3.17) from *Comamonas testosteroni* was cloned, expressed and purified as described in Example 4. The 4-hydroxy-2-oxoglutarate glyoxylate lyases (KHG aldolases) (EC 4.1.3.16) from *B. subtilis*, *E. coli*, and *S. meliloti* were cloned, expressed and purified as described in Example 4.

The aminotransferases used in conjunction with the aldolases to produce monatin were L-aspartate aminotransferase encoded by the *E. coli aspC* gene, the tyrosine aminotransferase encoded by the *E. coli tyrB* gene, the *S. meliloti* TatA enzyme, the broad substrate aminotransferase encoded by the *L. major hsat* gene, or the glutamic-oxaloacetic transaminase from pig heart (Type IIa). The

cloning, expression and purification of the non-mammalian proteins are described in Example 1. Glutamic-oxaloacetic transaminase from pig heart (type IIa) was obtained from Sigma (# G7005).

Method using ProA aldolase and L-aspartate aminotransferase

The reaction mixture contained 50 mM ammonium acetate, pH 8.0, 4 mM $MgCl_2$, 3 mM potassium phosphate, 0.05 mM pyridoxal phosphate, 100 mM ammonium pyruvate, 50 mM tryptophan, 10 mM-alpha-ketoglutarate, 160 mg of recombinant *C. testosteroni* ProA aldolase (unpurified cell extract, ~30% aldolase), 233 mg of recombinant *E. coli* L-aspartate aminotransferase (unpurified cell extract, ~40% aminotransferase) in one liter. All components except the enzymes were mixed together and incubated at 30°C until the tryptophan dissolved. The enzymes were then added and the reaction solution was incubated at 30°C with gentle shaking (100 rpm) for 3.5 hours. At 0.5 and 1 hour after the addition of the enzymes aliquots of solid tryptophan (50 mmoles each) were added to the reaction. All of the added tryptophan did not dissolve, but the concentration was maintained at 50 mM or higher. After 3.5 hours, the solid tryptophan was filtered off. Analysis of the reaction mixture by LC/MS using a defined amount of tryptophan as a standard showed that the concentration of tryptophan in the solution was 60.5 mM and the concentration of monatin was 5.81 mM (1.05 g).

The following methods were used to purify the final product. Ninety percent of the clear solution was applied to a column of BioRad AG50W-X8 resin (225 mL; binding capacity of 1.7 meq/mL). The column was washed with water, collecting 300 mL fractions, until the absorbance at 280 nm was <5% of the first flow through fraction. The column was then eluted with 1 M ammonium acetate, pH 8.4, collecting 4 300-mL fractions. All 4 fractions contained monatin and were evaporated to 105 mL using a roto-evaporator with a tepid water bath. A precipitate formed as the volume reduced and was filtered off over the course of the evaporation process.

Analysis of the column fractions by LC/MS showed that 99% of the tryptophan and monatin bound to the column. The precipitate that formed during the evaporation process contained >97% tryptophan and <2% of monatin. The ratio of tryptophan to product in the supernatant was approximately 2:1.

The supernatant (7 ml) was applied to a 100 mL Fast Flow DEAE Sepharose (Amersham Biosciences) column previously converted to the acetate form by washing with 0.5 L 1 M NaOH, 0.2 L water, 1.0 L of 1.0 M ammonium acetate, pH 8.4, and 0.5 L water. The supernatant was loaded at <2 mL/min and the column was washed with water at 3-4 mL/min until the absorbance at 280 nm was ~0. Monatin was eluted with 100 mM ammonium acetate, pH 8.4, collecting 4 100-mL fractions.

Analysis of the fractions showed that the ratio of tryptophan to monatin in the flow through fractions was 85:15 and the ratio in the eluent fractions was 7:93. Assuming the extinction coefficient at 280 nm of monatin is the same as tryptophan, the eluent fractions contained 0.146

mmole of product. Extrapolation to the total 1 L reaction would produce ~2.4 mmoles (~710 mg) of monatin, for a recovery of 68%.

The eluent fractions from the DEAE Sepharose column were evaporated to <20 mL. An aliquot of the product was further purified by application to a C_8 preparative reversed-phase column using the same chromatographic conditions as those described in Example 10 for the analytical-scale monatin characterization. Waters FractionLynx™ software was employed to trigger automated fraction collection of monatin based on detection of the $m/z = 293$ ion. The fraction from the C_8 column with the corresponding protonated molecular ion for monatin was collected, evaporated to dryness, and then dissolved in a small volume of water. This fraction was used for characterization of the product.

The resulting product was characterized using the following methods.

UV/Visible Spectroscopy. UV/visible spectroscopic measurements of monatin produced enzymatically were carried out using a Cary 100 Bio UV/visible spectrophotometer. The purified product, dissolved in water, showed an absorption maximum of 280 nm with a shoulder at 288 nm, characteristics typical of indole containing compounds.

LC/MS Analysis. Analyses of mixtures for monatin derived from the *in vitro* biochemical reactions were carried out as described in Example 10. A typical LC/MS analysis of monatin in an *in vitro* enzymatic synthetic mixture is illustrated in FIG. 5. The lower panel of FIG. 5 illustrates a selected ion chromatogram for the protonated molecular ion of monatin at $m/z = 293$. This identification of monatin in the mixture was corroborated by the mass spectrum illustrated in FIG. 6. Analysis of the purified product by LC/MS showed a single peak with a molecular ion of 293 and absorbance at 280 nm. The mass spectrum was identical to that shown in FIG. 6.

MS/MS Analysis. LC/MS/MS daughter ion experiments, as described in Example 10, were also performed on monatin. A daughter ion mass spectrum of monatin is illustrated in FIG. 7. Tentative structural assignments of all fragment ions labeled in FIG. 7 were made. These include fragment ions of $m/z = 275$ ($293 - H_2O$), 257 ($293-(2 \times H_2O)$), 230 (275-COOH), 212 (257-COOH), 168 (3-buta-1,3-dienyl-1*H*-indole carbonium ion), 158 (1*H*-indole-3-carbaldehyde carbonium ion), 144 (3-ethyl-1*H*-indole carbonium ion), 130 (3-methylene-1*H*-indole carbonium ion), and 118 (indole carbonium ion). Many of these are the same as those obtained for MP (Example 4), as expected if derived from the indole portion of the molecule. Some are 1 mass unit higher than those seen for MP, due to the presence of an amino group instead of a ketone.

High Resolution MS analysis. FIG. 8 illustrates the mass spectrum obtained for purified monatin employing an Applied Biosystems-Perkin Elmer Q-Star hybrid quadrupole/time-of-flight mass spectrometer. The measured mass for protonated monatin using tryptophan as an internal mass calibration standard was 293.1144. The calculated mass of protonated monatin, based on the elemental composition $C_{14}H_{17}N_2O_5$ is 293.1137. This is a mass measurement error of less than 2 parts per million (ppm), providing conclusive evidence of the elemental composition of monatin produced enzymatically.

NMR Spectroscopy. The NMR experiments were performed on a Varian Inova 500 MHz instrument. The sample of monatin (~3 mg) was dissolved in 0.5 ml of D_2O. Initially, the solvent (D_2O) was used as the internal reference at 4.78 ppm. Since the peak for water was large, the ^1H-NMR was run with suppression of the peak for water. Subsequently, due to the broadness of the water peak, the C-2 proton of monatin was used as the reference peak, and set at the published value of 7.192 ppm.

For ^{13}C-NMR, an initial run of several hundred scans indicated that the sample was too dilute to obtain an adequate ^{13}C spectrum in the allotted time. Therefore, a heteronuclear multiple quantum coherence (HMQC) experiment was performed, which enabled the correlation of the hydrogens and the carbons to which they were attached, and also providing information on the chemical shifts of the carbons.

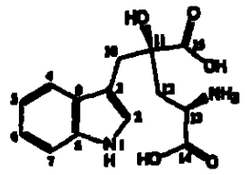
A summary of the ^1H and HMQC data is shown in Tables 2 and 3. By comparison to published values, the NMR data indicated that the enzymatically produced monatin was either (S,S), (R,R), or a mixture of both.

Chiral LC/MS Analysis. To establish that the monatin produced *in vitro* was one isomer, and not a mixture of the (R,R) and (S,S) enantiomers, chiral LC/MS analyses were carried out using the instrumentation described in Example 10.

Chiral LC separations were made using an Chirobiotic T (Advanced Separations Technology) chiral chromatography column at room temperature. Separation and detection, based on published protocols from the vendor, were optimized for the R- (D) and S- (L) isomers of tryptophan. The LC mobile phase consisted of A) water containing 0.05% (v/v) trifluoroacetic acid; B) Methanol containing 0.05% (v/v) trifluoroacetic acid. The elution was isocratic at 70% A and 30% B. The flow rate was 1.0 mL/min, and PDA absorbance was monitored from 200 nm to 400 nm. The instrumental parameters used for chiral LC/MS analysis of tryptophan and monatin are identical to those described in Example 10 for LC/MS analysis. Collection of mass spectra for the region *m/z* 150–400 was utilized. Selected ion chromatograms for protonated molecular ions ($[M + H]^+ = 205$ for both R- and S-tryptophan and $[M + H]^+ = 293$ for monatin) allowed direct identification of these analytes in the mixtures.

The chromatograms of R- and S-tryptophan and monatin, separated by chiral chromatography and monitored by MS, are shown in FIG. 9. The single peak in the chromatogram of monatin indicates that the compound is one isomer, with a retention time almost identical to S-tryptophan.

Table 2: ^{1}H NMR data



	Cargill		Vleggaar et al.[1]		Takeuhi et al.[2]	
Atom	δ_H	J(HH) Hz	δ_H	J(HH) Hz	δ_H	J(HH) Hz
2	7.192 (1H, s)		7.192 (s)		7.18 (s)	
4	7.671 (d)	7.99	7.686 (d)	7.9	7.67 (d)	8.0
5	7.104 (dd)	7.99	7.102 (dd)	8.0, 8.0	7.11 (dd)	7.5, 7.5
6	7.178 (dd)	*	7.176 (dd)	8.0, 8.0	7.17 (dd)	7.5, 7.5
7	7.439 (d)	7.99	7.439 (d)	8.1	7.43 (d)	8.0
10a	3.242(d)	14.5	3.243 (d)	14.3	3.24 (d)	14.5
10b	3.033 (d)	14.5	3.051 (d)	14.3	3.05 (d)	14.5
12	2.626 (dd) 2.015 (dd)	15.5, 1.5 15.0, 12.0	2.651 (dd) 2.006 (dd)	15.3, 1.7 15.3, 11.7	2.62 (dd) 2.01 (dd)	15.5, 1.8 15.5, 12.0
13	3.571 (dd)	10.75*, 1.5	3.168 (dd)	11.6, 1.8	3.57 (dd)	12.0, 1.8

[1] Vleggaar *et al.* (*J.C.S. Perkin Trans.* 1:3095-8, 1992).

[2] Takeuhi and Shusuke (JP2002060382, 2002-02-26).

Table 3: ^{13}C NMR data (from HMQC spectrum)

	Cargill	Vleggaar et al.[1]
Atom	δ_C	δ_C
2	126.1	126.03
3	*	110.31
4	120.4	120.46
5	120.2	120.25
6	122.8	122.74
7	112.8	112.79
8	*	137.06
9	*	129.23
10a	36.4	36.53
12	39.5	39.31
13	54.9	54.89
14	*	175.30
15	*	181.18

[1] Vleggaar *et al.* (*J.C.S. Perkin Trans.* 1:3095-8, 1992).

Polarimetry. The optical rotation was measured on a Rudolph Autopol III polarimeter. The monatin was prepared as a 14.6 mg/mL solution in water. The expected specific rotation ($[\alpha]_D^{20}$) for S,S monatin (salt form) is -49.6 for a 1 g/mL solution in water (Vleggaar *et al*). The observed $[\alpha]_D^{20}$ was -28.1 for the purified, enzymatically produced monatin indicating that it was the S, S isomer.

Improvements

The reaction conditions, including reagent and enzyme concentrations, were optimized and yields of 10 mg/mL were produced using the following reagent mix:
50 mM ammonium acetate pH 8.3, 2 mM $MgCl_2$, 200 mM pyruvate (sodium or ammonium salt), 5 mM alpha-ketoglutarate (sodium salt), 0.05 mM pyridoxal phosphate,
deaerated water to achieve a final volume of 1 mL after the addition of the enzymes,
3 mM potassium phosphate, 50 µg/mL of recombinant ProA aldolase (cell extract; total protein concentration of 167 µg/mL), 1000 µg/mL of L-aspartate aminotransferase encoded by the *E. coli aspC* gene (cell extract; total protein concentration of 2500 µg/mL), and solid tryptophan to afford a concentration of > 60 mM (saturated; some undissolved throughout the reaction). The mixture was incubated at 30°C for 4 hours with gentle stirring or mixing.

Substitutions

The concentration of alpha-ketoglutarate can be reduced to 1 mM and supplemented with 9 mM aspartate with an equivalent yield of monatin. Alternative amino acid acceptors can be utilized in the first step, such as oxaloacetate.

When recombinant *L. major* broad substrate aminotransferase was used in place of the *E. coli* L-aspartate aminotransferase, similar yields of monatin were achieved. However, a second unidentified product (3-10% of the major product) with a molecular mass of 292 was also detected by LC-MS analysis. Monatin concentrations of 0.1-0.5 mg/mL were produced when the *E. coli tyrB* encoded enzyme, the *S. meliloti tat A* encoded enzyme or the glutamic-oxaloacetic transaminase from pig heart (type IIa) was added as the aminotransferase. When starting the reaction from indole-3-pyruvate, a reductive amination can be done for the last step with glutamate dehydrogenase and NADH (as in Example 7).

The KHG aldolases from *B. subtilis*, *E. coli*, and *S. meliloti* were also used with the *E. coli* L-aspartate aminotransferase to produce monatin enzymatically. The following reaction conditions were used: 50 mM NH_4-OAc pH 8.3, 2 mM $MgCl_2$, 200 mM pyruvate, 5 mM glutamate, 0.05 mM pyridoxal phosphate, deaerated water to achieve a final volume of 0.5 mL after the addition of the enzymes, 3 mM potassium phosphate, 20 µg/mL of recombinant *B. subtilis* KHG aldolase (purified), ca. 400 µg/mL of *E. coli* L-aspartate aminotransferase (AspC) unpurified from cell extract, and 12 mM indole-3-pyruvate. The reactions were incubated at 30°C for 30 minutes with shaking. The amount of monatin produced using the *B. subtilis* enzyme was 80 ng/mL, and increased with increasing amounts of aldolase. If indole-3-pyruvate and glutamate were replaced by saturating amounts of tryptophan and 5 mM alpha-ketoglutarate, the production of monatin was increased to 360 ng/mL. Reactions were repeated with 30 µg/mL of each of the three KHG enzymes in 50 mM Tris pH 8.3, with saturating amounts of tryptophan, and were allowed to proceed for an hour in order to increase detection. The *Bacillus* enzyme had the highest activity as in Example 4, producing approximately 4000 ng/mL monatin. The *E. coli* KHG produced 3000 ng/mL monatin, and the *S. meliloti* enzyme produced 2300 ng/mL.

Example 7

Interconversion between MP and Monatin

The amination of MP to form monatin can be catalyzed by aminotransferases such as those identified in Examples 1 and 6, or by dehydrogenases that require a reducing cofactor such as NADH or NADPH. These reactions are reversible and can be measured in either direction. The directionality, when using a dehydrogenase enzyme, can be largely controlled by the concentration of ammonium salts.

Dehydrogenase activity. The oxidative deamination of monatin was monitored by following the increase in absorbance at 340 nm as NAD(P)+ was converted to the more chromophoric NAD(P)H. Monatin was enzymatically produced and purified as described in Example 6.

A typical assay mixture contained 50 mM Tris-HCl, pH 8.0 to 8.9, 0.33 mM NAD^+ or $NADP^+$, 2 to 22 units of glutamate dehydrogenase (Sigma), and 10-15 mM substrate in 0.2 mL. The assay was performed in duplicate in a UV-transparent microtiter plate, on a Molecular Devices SpectraMax Plus platereader. A mix of the enzyme, buffer, and $NAD(P)^+$ were pipetted into wells containing the substrate and the increase in absorbance at 340 nm was monitored at 10 second intervals after brief mixing. The reaction was incubated at 25°C for 10 minutes. Negative controls were carried out without the addition of substrate, and glutamate was utilized as a positive control. The type III glutamate dehydrogenase from bovine liver (Sigma # G-7882) catalyzed the conversion of the monatin to the monatin precursor at a rate of conversion approximately one-hundredth the rate of the conversion of glutamate to alpha-ketoglutarate.

Transamination activity. Monatin aminotransferase assays were conducted with the aspartate aminotransferase (AspC) from *E. coli*, the tyrosine aminotransferase (TyrB) from *E. coli*, the broad substrate aminotransferase (BSAT) from *L. major*, and the two commercially available porcine glutamate-oxaloacetate aminotransferases described in Example 1. Both oxaloacetate and alpha-ketoglutarate were tested as the amino acceptor. The assay mixture contained (in 0.5 mL) 50 mM Tris-HCl, pH 8.0, 0.05 mM PLP, 5 mM amino acceptor, 5 mM monatin, and 25 µg of aminotransferase. The assays were incubated at 30°C for 30 minutes, and the reactions were stopped by addition of 0.5 mL isopropyl alcohol. The loss of monatin was monitored by LC/MS (Example 10). The highest amount of activity was noted with *L. major* BSAT with oxaloacetate as the amino acceptor, followed by the same enzyme with alpha-ketoglutarate as the amino acceptor. The relative activity with oxaloacetate was: BSAT > AspC > porcine type IIa > porcine type I ≈ TyrB. The relative activity with alpha-ketoglutarate was: BSAT > AspC > porcine type I > porcine type IIa > TyrB.

Example 8

Production of Monatin from Tryptophan and C3 Sources Other than Pyruvate

As described above in Example 6, indole-3-pyruvate or tryptophan can be converted to monatin using pyruvate as the C3 molecule. However, in some circumstances, pyruvate may not be a desirable raw material. For example, pyruvate may be more expensive than other C3 carbon sources, or may have adverse effects on fermentations if added to the medium. Alanine can be transaminated by many PLP-enzymes to produce pyruvate.

Tryptophanase-like enzymes perform beta-elimination reactions at faster rates than other PLP enzymes such as aminotransferases. Enzymes from this class (4.1.99.-) can produce ammonia and pyruvate from amino acids such as L-serine, L-cysteine, and derivatives of serine and cysteine with good leaving groups such as O-methyl-L-serine, O-benzyl-L-serine, S-methylcysteine, S-benzylcysteine, S-alkyl-L-cysteine, O-acyl-L-serine, 3-chloro-L-alanine.

Processes to produce monatin using EC 4.1.99.- polypeptides can be improved by mutating the β-tyrosinase (TPL) or tryptophanase according to the method of Mouratou *et al.* (*J. Biol. Chem* 274:1320-5, 1999). Mouratou *et al.* describe the ability to covert the β-tyrosinase into a dicarboxylic amino acid β-lyase, which has not been reported to occur in nature. The change in specificity was accomplished by converting valine (*V*) 283 to arginine (*R*) and arginine (*R*) 100 to threonine (*T*). These amino acid changes allow for the lyase to accept a dicarboxylic amino acid for the hydrolytic deamination reaction (such as aspartate). Aspartate, therefore, can also be used as a source of pyruvate for subsequent aldol condensation reactions.

Additionally, cells or enzymatic reactors can be supplied with lactate and an enzyme that converts lactate to pyruvate. Examples of enzymes capable of catalyzing this reaction include lactate dehydrogenase and lactate oxidase.

Isolation of Genomic DNA

Tryptophanase polypeptides have previously been reported in, for example, Mouratou *et al.* (*JBC* 274:1320-5, 1999). To isolate genes that encode tryptophanase polypeptides, genomic DNA from *E. coli* DH10B was used as a template for PCR as described in Example 1.

The gene for tyrosine-phenol lyase was isolated from *C. freundii* (ATCC catalog number 8090, Designation ATCC 13316; NCTC 9750) and grown on Nutrient agar (Difco 0001) and nutrient broth (Difco 0003) at 37°C to an OD of 2.0. The genomic DNA was purified using a Qiagen Genomic-tip ™ 100/G kit.

PCR Amplification of Coding Sequences

Primers were designed with compatible overhangs for the pET 30 Xa/LIC vector (Novagen, Madison, WI) as described above in Example 1.

E. coli tna (SEQ ID NO: 41). N-terminal primer for pET30 Xa/LIC cloning: 5'-GGT ATT GAG GGT CGC ATG GAA AAC TTT AAA CAT CT-3' (SEQ ID NO: 43). C-terminal primer for

pET30 Xa/LIC cloning: 5'-AGA GGA GAG TTA GAG CCT TAA ACT TCT TTA AGT TTT G-3' (SEQ ID NO: 44).

C. freundii tpl (SEQ ID NO: 42). N-terminal primer for pET30 Xa/LIC cloning: 5'-GGT ATT GAG GGT CGC ATGAATTATCCGGCAGAACC-3' (SEQ ID NO: 45). C-terminal primer for pET 30 Xa/LIC cloning: 5'-AGA GGA GAG TTA GAG CCTTAGATGTAATCAAAGCGTG-3' (SEQ ID NO: 46).

The Eppendorf Mastercycler™ Gradient 5331 Thermal Cycler was used for all PCR reactions. In 50 µL was added 0.5 µg template (genomic DNA), 1.0 µM of each primer, 0.4 mM each dNTP, 3.5 U Expand High Fidelity Polymerase (Roche), 1X Expand buffer with Mg, and 5% DMSO (final concentration). The thermocycler PCR program used was as follows: 96°C hot start (5 minutes), 94°C - 30 seconds, 40-60°C - 1 minute 45 seconds, 72°C - 2 minutes 15 seconds; 30 repetitions. The final polymerization step was for 7 minutes, and the samples were then stored at 4°C.

Cloning

Cloning and positive clone identification procedures detailed above in Example 1 were used to identify the appropriate clones.

Gene Expression and Activity Assays

Plasmid DNA (verified by sequence analysis) was subcloned into the expression host BL21(DE3) (Novagen). The cultures were grown in LB medium with 30 mg/L kanamycin, the plasmids were isolated using a Qiagen miniprep kit, and analyzed by restriction digest to confirm identity.

Induction experiments were done with the BL21(DE3) expression host, the constructs were grown in LB medium containing 50 mg/L kanamycin at 37°C. Protein expression was induced using 0.1 mM IPTG after the OD_{600} of the culture reached approximately 0.6. The cells were grown for 4 hours at 30°C and harvested by centrifugation. The cells were then lysed in 5 mL/g wet cell weight BugBuster™ (Novagen) reagent containing 5 µL/mL protease inhibitor cocktail set #III (Calbiochem) and 1 µL/mL benzonase nuclease (Novagen), and the His-tagged recombinant proteins were purified using the His-Bind cartridges as described above in Example 1. Purified proteins were desalted on a PD-10 (G25 Sephadex, Amersham Biosciences) column and eluted in 100 mM Tris-Cl buffer, pH 8.0. The proteins were analyzed by SDS-PAGE on 4-15% gradient gels to check for soluble protein levels at the predicted MW of the recombinant fusion protein.

Mutagenesis

Some members of polypeptide class 4.1.99.- (tryptophanase and β-tyrosinase) will perform the beta-lyase reaction with aspartate or similar amino acids without any modification. However, some members of the class may need to be mutagenized to allow for the use of the substrates and/or

the creation of the product. Moreover, in some cases polypeptides that can perform the conversion may be further optimized by mutagenesis.

Site directed mutagenesis was performed based on 3D structure analysis of PLP-binding polypeptides. Two examples for changing the *substrate specificity* of the polypeptides are shown below.

Mutagenesis of Tryptophanase Example 1

The mutagenesis protocol provided below introduced two point mutations in the amino acid sequence. The first point mutation changed arginine (R) at position 103 to threonine (T) and the second point mutation changed valine (V) at position 299 to arginine (R) (numbering system for *E. coli* mature protein). Mutagenesis experiments were performed by ATG Laboratories (Eden Prairie, MN). Mutations were introduced sequentially by PCR of gene fragments and reassembly of the fragments was accomplished by PCR as well. Primers for converting arginine (R)103 to threonine (T): 5'-CCAGGGCACCGGCGCAGAGCAAATCTATATT-3' (SEQ ID NO: 47) and 5'-TGCGCCGGTGCCCTGGTGAGTCGGAATGGT-3' (SEQ ID NO: 48).

Primers for converting valine (V)299 to arginine (R): 5'-TCCTGCACGCGGCAAAGGGTTCTGCACTCGGT-3' (SEQ ID NO: 49) and 5'-CTTTGCCGCGTGCAGGAAGGCTTCCCGACA-3' (SEQ ID NO: 50).

Mutants were screened by restriction digest with Xba I/HindIII and SphI, and verified by sequencing.

Mutagenesis of Tyrosine Phenol Lyase (β-tyrosinase) Example 2

Two point mutations were made to the tyrosine phenol lyase amino acid sequence. These mutations converted arginine (R) at position 100 to threonine (T) and valine (V) at position 283 to arginine (R) (in *C. freundii* mature protein sequence).

Primers for the R100T conversion were: 5'-AGGGGACCGGCGCAGAAAACCTGTTATCG-3' (SEQ ID NO: 51) and 5'-AGGGGACCGGCGCAGAAAACCTGTTATCG-3' (SEQ ID NO: 52). Primers for the V283R conversion were: 5'-GTTAGTCCGCGTCTACGAAGGGATGCCAT-3' (SEQ ID NO: 53) and 5'-GTAGACGCGGACTAACTCTTTGGCAGAAG-3' (SEQ ID NO: 54).

The methods described above were used, and the clones were screened by KpnI/SacI digestion, and BstX I digestion. The sequences were verified by dideoxy chain termination sequencing. Recombinant protein was produced as described above for the wildtype enzymes.

The reaction mixture consisted of 50 mM Tris-Cl pH 8.3, 2 mM $MgCl_2$, 200 mM C3 carbon source, 5 mM alpha-ketoglutarate, sodium salt, 0.05 mM pyridoxal phosphate, deaerated water to achieve a final volume of 0.5 mL after the addition of the enzymes, 3 mM potassium phosphate pH 7.5, 25 µg of crude recombinant *C. testosteroni* ProA aldolase as prepared as in Example 4, 500 µg of crude L-aspartate aminotransferase (AspC) as prepared in Example 1, and solid tryptophan to afford a

concentration of > 60 mM (saturated; some undissolved throughout the reaction). The reaction mix was incubated at 30°C for 30 minutes with mixing. Serine, alanine, and aspartate were supplied as 3-carbon sources. Assays were performed with and without secondary PLP enzymes (purified) capable of performing beta-elimination and beta-lyase reactions (tryptophanase (TNA), double mutant tryptophanase, β-tyrosinase (TPL)). The results are shown in Table 4:

Table 4: Production of monatin utilizing alternative C3-carbon sources

C3-carbon source	Additional PLP enzyme	Relative Activity
none	none	0%
pyruvate	none	100%
serine	none	3%
serine	11 µg wildtype TNA (1 U)	5.1%
serine	80 µg double mutant TNA	4.6%
alanine	none	32%
alanine	11 µg wildtype TNA	41.7%
alanine	80 µg mutant TNA	43.9%
aspartate	110 µg wildtype TNA (10 U)	7.7%
aspartate	5 U wildtype TPL (crude)	5.1%
aspartate	80 µg mutant TNA	3.3%

The monatin produced from alanine and serine as 3-carbon sources was verified by *LC/MS/MS* daughter scan analysis, and was identical to the characterized monatin produced in Example 6. Alanine was the best alternative tested, and was transaminated by the AspC enzyme. The amount of monatin produced was increased by addition of the tryptophanase, which is capable of transamination as a secondary activity. The amount of monatin produced with serine as a carbon source nearly doubled with the addition of the tryptophanase enzymes, even though only one-fifth of the amount of tryptophanase was added in comparison to the aminotransferase. AspC is capable of some amount of beta-elimination activity alone. The results with aspartate indicate that the tryptophanase activity on aspartate does not increase with the same site-directed mutations as previously suggested for β-tyrosinase. It is expected that the mutant β-tyrosinase will have higher activity for production of monatin.

Example 9

Chemical Synthesis of Monatin

The addition of alanine to indole-3-pyruvic acid produces monatin, and this reaction can be performed synthetically with a Grignard or organolithium reagent.

For example, to 3-chloro- or 3-bromo-alanine which has been appropriately blocked at the carboxyl and amino groups, is added magnesium under anhydrous conditions. Indole-3-pyruvate (appropriately blocked) is then added to form the coupled product followed by removal of the protecting groups to form monatin. Protecting groups that are particularly useful include THP (tetrahydropyranyl ether) which is easily attached and removed.

Example 10

Detection of Monatin and MP

This example describes methods used to detect the presence of monatin, or its precursor 2-hydroxy 2-(indol-3-ylmethyl)-4-keto glutaric acid.

LC/MS Analysis

Analyses of mixtures for the alpha-keto acid form of monatin (monatin precursor, MP) and monatin derived from *in vitro* or *in vivo* biochemical reactions were performed using a Waters/Micromass liquid chromatography-tandem mass spectrometry (LC/MS/MS) instrument including a Waters 2690 liquid chromatograph with a Waters 996 Photo-Diode Array (PDA) absorbance monitor placed in series between the chromatograph and a Micromass Quattro Ultima triple quadrupole mass spectrometer. LC separations were made using a Supelco Discovery C_{18} reversed-phase chromatography column, 2.1mm x 150 mm, or an Xterra MS C_8 reversed-phase chromatography column, 2.1mm x 250 mm, at room temperature. The LC mobile phase consisted of A) water containing 0.05% (v/v) trifluoroacetic acid and B) methanol containing 0.05% (v/v) trifluoroacetic acid.

The gradient elution was linear from 5% B to 35% B, 0-9 min, linear from 35% B to 90% B, 9-16 min, isocratic at 90% B, 16-20 min, linear from 90% B to 5% B, 20-22 min, with a 10 min re-equilibration period between runs. The flow rate was 0.25 mL/min, and PDA absorbance was monitored from 200 nm to 400 nm. All parameters of the ESI-MS were optimized and selected based on generation of protonated molecular ions ($[M + H]^+$) of the analytes of interest, and production of characteristic fragment ions.

The following instrumental parameters were used for LC/MS analysis of monatin: Capillary: 3.5 kV; Cone: 40 V; Hex 1: 20 V; Aperture: 0 V; Hex 2: 0 V; Source temperature: 100°C; Desolvation temperature: 350°C; Desolvation gas: 500 L/h; Cone gas: 50 L/h; Low mass resolution (Q1): 15.0; High mass resolution (Q1): 15.0; Ion energy: 0.2; Entrance: 50V; Collision Energy: 2; Exit: 50V; Low mass resolution (Q2): 15; High mass resolution (Q2): 15; Ion energy (Q2): 3.5; Multiplier: 650. Uncertainties for reported mass/charge ratios (m/z) and molecular masses are ± 0.01%. Initial detection of the alpha-keto acid form of monatin (MP) and monatin in the mixtures was accomplished by LC/MS monitoring with collection of mass spectra for the region m/z 150–400. Selected ion chromatograms for protonated molecular ions ($[M + H]^+ = 292$ for MP, $[M + H]^+ = 293$ for monatin) allowed direct identification of these analytes in the mixtures.

MS/MS Analysis

LC/MS/MS daughter ion experiments were performed on monatin as follows. A daughter ion analysis involves transmission of the parent ion (e.g., $m/z = 293$ for monatin) of interest from the first mass analyzer (Q1) into the collision cell of the mass spectrometer, where argon is introduced and chemically dissociates the parent into fragment (daughter) ions. These fragment ions are then

detected with the second mass analyzer (Q2), and can be used to corroborate the structural assignment of the parent.

The following instrumental parameters were used for LC/MS/MS analysis of monatin: Capillary: 3.5 kV; Cone: 40 V; Hex 1: 20 V; Aperture: 0 V; Hex 2: 0 V; Source temperature: 100 °C; Desolvation temperature: 350 °C; Desolvation gas: 500 L/h; Cone gas: 50 L/h; Low mass resolution (Q1): 13.0; High mass resolution (Q1): 13.0; Ion energy: 0.2; Entrance: -5 V; Collision Energy: 14; Exit: 1V; Low mass resolution (Q2): 15; High mass resolution (Q2): 15; Ion energy (Q2): 3.5; Multiplier: 650.

High-Throughput Determination of Monatin

High-throughput analyses (< 5 min/sample) of mixtures for monatin derived from *in vitro* or *in vivo* reactions were carried out using instrumentation described above, and the same parameters as described for LC/MS/MS. LC separations were made using Waters Xterra MS C_8 (2.1mm x 50 mm) chromatography at room temperature with isocratic elution in 15% aqueous MeOH, 0.25% acetic acid at a flow rate of 0.3 mL/min. Detection of monatin in the mixtures was accomplished using selected reaction monitoring (SRM)-tandem mass spectrometry. This involved monitoring specific collisionally-induced parent ion ($[M + H]^+ = 293.1$) to daughter ion (e.g., the fragment ion at m/z = 168.1, tentatively assigned as a 3-buta-1,3-dienyl-1*H*-indole carbonium ion) transitions to maximize sensitivity, selectivity, and throughput for the detection of monatin. PDA absorbance data were collected in parallel for further verification of monatin identity.

Example 11

Production of Monatin in Bacteria

This example describes methods used to produce monatin in *E. coli* cells. One skilled in the art will understand that similar methods can be used to produce monatin in other bacterial cells. In addition, vectors containing other genes in the monatin synthesis pathway (FIG. 2) can be used.

Trp-1 + glucose medium, a minimal medium that has been used for increased production of tryptophan in *E. coli* cells (Zeman *et al. Folia Microbiol.* 35:200-4, 1990), was prepared as follows. To 700 mL nanopure water the following reagents were added: 2 g $(NH_4)_2SO_4$, 13.6 g KH_2PO_4, 0.2 g $MgSO_4 \cdot 7H_2O$, 0.01 g $CaCl_2 \cdot 2H_2O$, and 0.5 mg $FeSO_4 \cdot 7H_2O$. The pH was adjusted to 7.0, the volume was increased to 850 mL, and the medium was autoclaved. A 50% glucose solution was prepared separately, and sterile-filtered. Forty mL was added to the base medium (850 mL) for a 1 L final volume.

A 10 g/L L-tryptophan solution was prepared in 0.1 M sodium phosphate pH 7, and sterile-filtered. One-tenth volume was typically added to cultures as specified below. A 10% sodium pyruvate solution was also prepared and sterile-filtered. A 10 mL aliquot was typically used per liter of culture. Stocks of ampicillin (100 mg/mL), kanamycin (25 mg/mL) and IPTG (840 mM) were prepared, sterile-filtered, and stored at –20°C before use. Tween 20 (polyoxyethylene 20-Sorbitan

monolaurate) was utilized at a 0.2% (vol/vol) final concentration. Ampicillin was used at non-lethal concentrations, typically 1-10 μg/mL final concentration.

Fresh plates of *E. coli* BL21(DE3)::*C. tertosteroni proA*/pET 30 Xa/LIC (described in Example 4) were prepared on LB medium containing 50 μg/mL kanamycin. Overnight cultures (5 mL) were inoculated from a single colony and grown at 30°C in LB medium with kanamycin. Typically a 1 to 50 inoculum was used for induction in trp-1 + glucose medium. Fresh antibiotic was added to a final concentration of 50 mg/L. Shake flasks were grown at 37°C prior to induction.

Cells were sampled every hour until an OD_{600} of 0.35-0.8 was obtained. Cells were then induced with 0.1 mM IPTG, and the temperature reduced to 34 °C. Samples (1 ml) were collected prior to induction (zero time point) and centrifuged at 5000 x g. The supernatant was frozen at –20°C for LC/MS analysis. Four hours post-induction, another 1 mL sample was collected, and centrifuged to separate the broth from the cell pellet. Tryptophan, sodium pyruvate, ampicillin, and Tween were added as described above.

The cells were grown for 48 hours post-induction, and another 1 mL sample was taken and prepared as above. At 48 hours, another aliquot of tryptophan and pyruvate were added. The entire culture volume was centrifuged after approximately 70 hours of growth (post-induction), for 20 minutes at 4°C and 3500 rpm. The supernatant was decanted and both the broth and the cells were frozen at –80°C. The broth fractions were filtered and analyzed by LC/MS. The heights and areas of the $[M+H]^+ = 293$ peaks were monitored as described in Example 10. The background level of the medium was subtracted. The data was also normalized for cell growth by plotting the height of the $[M+H]^+ = 293$ peak divided by the optical density of the culture at 600 nm.

Higher levels of monatin were produced when pyruvate, ampicillin, and Tween were added 4 hours post induction rather than at induction. Other additives such as PLP, additional phosphate, or additional $MgCl_2$ did not increase the production of monatin. Higher titers of monatin were obtained when tryptophan was utilized instead of indole-3-pyruvate, and when the tryptophan was added post-induction rather than at inoculation, or at induction. Prior to induction, and 4 hours post-induction (at time of substrate addition), there was typically no detectable level of monatin in the fermentation broth or cellular extracts. Negative controls were done utilizing cells with pET30a vector only, as well as cultures where tryptophan and pyruvate were not added. A parent MS scan demonstrated that the compound with $(m+1)/z = 293$ was not derived from larger molecules, and daughter scans (performed as in Example 10) were similar to monatin made *in vitro*.

The effect of Tween was studied by utilizing 0, 0.2% (vol/vol), and 0.6% final concentrations of Tween-20. The highest amount of monatin produced by shake flasks was at 0.2% Tween. The ampicillin concentration was varied between 0 and 10 μg/mL. The amount of monatin in the cellular broth increased rapidly (2.5 X) between 0 and 1 μg/mL, and increased 1.3 X when the ampicillin concentration was increased from 1 to 10 μg/mL.

A time course experiment showing typical results is shown in FIG. 10. The amount of monatin secreted into the cell broth increased, even when the values are normalized for cell growth.

By using the molar extinction coefficient of tryptophan, the amount of monatin in the broth was estimated to be less than 10 μg/mL. The same experiment was repeated with the cells containing vector without *proA* insert. Many of the numbers were negative, indicating the peak height at m/z=293 was less in these cultures than in the medium alone (FIG. 10). The numbers were consistently lower when tryptophan and pyruvate were absent, demonstrating that monatin production is a result of an enzymatic reaction catalyzed by the aldolase enzyme.

The *in vivo* production of monatin in bacterial cells was repeated in 800 mL shake flask experiments and in fermentors. A 250 mL sample of monatin (in cell-free broth) was purified by anion exchange chromatography and preparative reverse-phase liquid chromatography. This sample was evaporated, and submitted for high resolution mass analysis (described in Example 6). The high resolution MS indicated that the metabolite being produced is monatin.

In vitro assays indicate that aminotransferase needs to be present at higher levels than aldolase (see Example 6), therefore the aspartate aminotransferase from *E. coli* was overexpressed in combination with the aldolase gene to increase the amount of monatin produced. Primers were designed to introduce *C. tartosteroni proA* into an operon with *aspC*/pET30 Xa/LIC, as follows: 5' primer: ACTCGGATCCGAAGGAGATATACATATGTACGAACTGGGACT (SEQ ID NO: 67) and 3' primer: CGGCTGTCGACCGTTAGTCAATATATTTCAGGC (SEQ ID NO: 68). The 5' primer contains a BamHI site, the 3' primer contains a SalI site for cloning. PCR was performed as described in Example 4, and gel purified. The *aspC*/pET30 Xa/LIC construct was digested with BamHI and SalI, as was the PCR product. The digests were purified using a Qiagen spin column. The *proA* PCR product was ligated to the vector using the Roche Rapid DNA Ligation kit (Indianapolis, IN) according to manufacturer's instructions. Chemical transformations were done using Novablues Singles (Novagen) as described in Example 1. Colonies were grown up in LB medium containing 50 mg/L kanamycin and plasmid DNA was purified using the Qiagen spin miniprep kit. Clones were screened by restriction digest analysis and sequence was confirmed by Seqwright (Houston, TX). Constructs were subcloned into BLR(DE3), BLR(DE3)pLysS, BL21(DE3) and BL21(DE3)pLysS (Novagen). The *proA*/pET30 Xa/LIC construct was also transformed into BL21(DE3)pLysS.

Initial comparisons of BLR(DE3) shake flask samples under the standard conditions described above demonstrated that the addition of the second gene (*aspC*) improved the amount of monatin produced by seven-fold. To hasten growth, BL21(DE3)-derived host strains were used. The *proA* clones and the two gene operon clones were induced in Trp-1 medium as above, the pLysS hosts had chloramphenicol (34 mg/L) added to the medium as well. Shake flask experiments were performed with and without the addition of 0.2% Tween-20 and 1 mg/L ampicillin. The amount of monatin in the broth was calculated using *in vitro* produced purified monatin as a standard. SRM analyses were performed as described in Example 10. Cells were sampled at zero, 4 hours, 24 hours, 48 hours, 72 hours, and 96 hours of growth.

The results are shown in Table 5 for the maximum amounts produced in the culture broths. In most instances, the two gene construct gave higher values than the *proA* construct alone. The

pLysS strains, which should have leakier cell envelopes, had higher levels of monatin secreted, even though these strains typically grow at a slower rate. The additions of Tween and ampicillin were beneficial.

Table 5: Amount of Monatin Produced by *E. coli* Bacteria

Construct	Host	Tween + Amp	µg/mL monatin	time
proA	BL21(DE3)	-	0.41	72 hr
proA	BL21(DE3)	+	1.58	48 hr
proA	BL21(DE3)pLysS	-	1.04	48 hr
proA	BL21(DE3)pLysS	+	1.60	48 hr
aspC:proA	BL21(DE3)	-	0.09	48 hr
aspC:proA	BL21(DE3)	+	0.58	48 hr
aspC:proA	BL21(DE3)pLysS	-	1.39	48 hr
aspC:proA	BL21(DE3)pLysS	+	6.68	48 hr

Example 12

Production of Monatin in Yeast

This example describes methods used to produce monatin in eukaryotic cells. One skilled in the art will understand that similar methods can be used to produce monatin in any cell of interest. In addition, other genes can be used (e.g., those listed in FIG. 2) in addition to, or alternatively to those described in this example.

The pESC Yeast Epitope Tagging Vector System (Stratagene, La Jolla, CA) was used to clone and express the *E. coli aspC* and *C. testosteroni proA* genes into *Saccharomyces cerevisiae*. The pESC vectors contain both the GAL1 and the GAL10 promoters on opposite strands, with two distinct multiple cloning sites, allowing for expression of two genes at the same time. The pESC-His vector also contains the *His3* gene for complementation of histidine auxotrophy in the host (YPH500). The GAL1 and GAL10 promoters are repressed by glucose and induced by galactose; a Kozak sequence is utilized for optimal expression in yeast. The pESC plasmids are shuttle vectors, allowing the initial construct to be made in *E. coli* (with the *bla* gene for selection); however, no bacterial ribosome binding sites are present in the multiple cloning sites.

The following primers were designed for cloning into pESC-His (restriction sites are underlined, Kozak sequence is in bold): *aspC* (*BamHI/SalI*), GAL1: 5'-CGCGGATCCATAATGGTTGAGAACATTACCG-3' (SEQ ID NO: 69) and 5'-ACGCGTCGACTTACAGCACTGCCACAATCG-3' (SEQ ID NO: 70). *proA* (*EcoRI/NotI*), GAL10: 5'-CCGGAATTCATAATGGTCGAACTGGGAGTTGT-3' (SEQ ID NO: 71) and 5'-GAATGCGGCCGCTTAGTCAATATATTTCAGGCC-3' (SEQ ID NO: 72).

The second codon for both mature proteins was changed from an aromatic amino acid to valine due to the introduction of the Kozak sequence. The genes of interest were amplified using pET30 Xa/LIC miniprep DNA from the clones described in Examples 1 and Example 4 as template. PCR was performed using an Eppendorf Master cycler gradient thermocycler and the following protocol for a 50 µL reaction: 1.0 µL template, 1.0 µM of each primer, 0.4 mM each dNTP, 3.5 U

Expand High Fidelity Polymerase (Roche, Indianapolis, IN), and 1X Expand™ buffer with Mg. The thermocycler program used consisted of a hot start at 94°C for 5 minutes, followed by 29 repetitions of the following steps: 94°C for 30 seconds, 50°C for 1 minute 45 seconds, and 72°C for 2 minutes 15 seconds. After the 29 repetitions the sample was maintained at 72°C for 10 minutes and then stored at 4°C. The PCR products were purified by separation on a 1% TAE-agarose gel followed by recovery using a QIAquick Gel Extraction Kit (Qiagen, Valencia, CA).

The pESC-His vector DNA (2.7 µg) was digested with *Bam*HI/*Sal*I and gel-purified as above. The *aspC* PCR product was digested with *Bam*HI/*Sal*I and purified with a QIAquick PCR Purification Column. Ligations were performed with the Roche Rapid DNA Ligation Kit following the manufacturer's protocols. Desalted ligations were electroporated into 40 µl Electromax DH10B competent cells (Invitrogen) in a 0.2 cm Biorad disposable cuvette using a Biorad Gene Pulser II with pulse controller plus, according to the manufacturer's instructions. After 1 hour of recovery in 1 mL of SOC medium, the transformants were plated on LB medium containing 100 µg/mL ampicillin. Plasmid DNA preparations for clones were done using QIAprep Spin Miniprep Kits. Plasmid DNA was screened by restriction digest, and sequenced (Seqwright) for verification using primers designed for the vector.

The *aspC*/pESC-His clone was digested with *Eco*RI and *Not*I, as was the *proA* PCR product. DNA was purified as above, and ligated as above. The two gene construct was transformed into DH10B cells and screened by restriction digest and DNA sequencing.

The construct was transformed into *S. cerevisiae* strain YPH500 using the S.c. EasyComp™ Transformation Kit (Invitrogen). Transformation reactions were plated on SC-His minimal medium (Invitrogen pYES2 manual) containing 2% glucose. Individual yeast colonies were screened for the presence of the *proA* and *aspC* genes by colony PCR using the PCR primers above. Pelleted cells (2 µl) were suspended in 20 µL of Y-Lysis Buffer (Zymo Research) containing 1 µl of zymolase and heated at 37°C for 10 minutes. Four µL of this suspension was then used in a 50 µL PCR reaction using the PCR reaction mixture and program described above.

Five mL cultures were grown overnight on SC-His + glucose at 30°C and 225 rpm. The cells were gradually adjusted to growth on raffinose in order to minimize the lag period prior to induction with galactose. After approximately 12 hours of growth, absorbance measurements at 600 nm were taken, and an appropriate volume of cells was spun down and resuspended to give an OD of 0.4 in the fresh SC-His medium. The following carbon sources were used sequentially: 1% raffinose + 1 % glucose, 0.5% glucose + 1.5% raffinose, 2% raffinose, and finally 1% raffinose + 2% galactose for induction.

After approximately 16 hours of growth in induction medium, the 50 mL cultures were divided into duplicate 25 mL cultures, and the following were added to only one of the duplicates: (final concentrations) 1 g/L L-tryptophan, 5 mM sodium phosphate pH 7.1, 1 g/L sodium pyruvate, 1 mM $MgCl_2$. Samples of broths and cell pellets from the non-induction medium, and from the 16 hour cultures prior to addition of substrates for the monatin pathway, were saved as negative controls. In

addition, constructs containing only a functional *aspC* gene (and a truncated *proA* gene) were utilized as another negative control. The cells were allowed to grow for a total of 69 hours post-induction. Occasionally the yeast cells were induced at a lower OD, and only grown for 4 hours prior to addition of tryptophan and pyruvate. However, these monatin substrates appear to inhibit growth and the addition at higher OD was more effective.

The cell pellets from the cultures were lysed with 5 mL of YeastBuster™ + 50 µl THP (Novagen) per gram (wet weight) of cells following manufacturer's protocols, with the addition of protease inhibitors and benzonase nuclease as described in previous examples. The culture broth and cell extracts were filtered and analyzed by SRM as described in Example 10. Using this method, no monatin was detected in the broth samples, indicating that the cells could not secrete monatin under these conditions. The proton motive force may be insufficient under these conditions or the general amino acid transporters may be saturated with tryptophan. Protein expression was not at a level that allowed for detection of changes using SDS-PAGE.

Monatin was detectable (approximately 60 µg/mL) transiently in cell extracts of the culture with two functional genes, when tryptophan and pyruvate were added to the medium. Monatin was not detected in any of the negative control cell extracts. *In vitro* assays for monatin were performed in duplicate with 4.4 mg/mL of total protein (about double what is typically used for *E. coli* cell extracts) using the optimized assay described in Example 6. Other assays were performed with the addition of either 32 µg/mL *C. testosteroni* ProA aldolase or 400 µg/mL AspC aminotransferase, to determine which enzyme was limiting in the cell extract. Negative controls were performed with no addition of enzyme, or the addition of only AspC aminotransferase (the aldol condensation can occur to some extent without enzyme). Positive controls were performed with partially pure enzymes (30-40%), using 16 µg/mL aldolase and 400 µg/mL aminotransferase.

In vitro results were analyzed by SRM. The analysis of cell extracts showed that tryptophan was effectively transported into the cells when it was added to the medium post-induction, resulting in tryptophan levels two orders of magnitude higher than those in which no additional tryptophan was added. The results for *in vitro* monatin analysis are shown in Table 6 (numbers indicate ng/mL).

Table 6: Monatin production with yeast cell extracts.

	aspC construct	+ aldolase	+ AspC	two-gene construct	+ aldolase	+ AspC
repressed (glucose medium)	0	888.3	173.5	0	465.2	829
24 hr induced	0	2832.8	642.4	0	1375.6	9146.6
69 hr induced	0	4937.3	340.3	71.9	1652.8	23693.5
69 hr + subs.	0	556.9	659.1	21.9	755.6	16688.2
+ control (purified enzymes)	21853			21853		
-control (no enzymes)	0		254.3	0		254.3

Positive results were obtained with the full two-gene construct cell extracts with and without substrate added to the growth medium. These results, in comparison to the positive controls, indicate that the enzymes were expressed at levels of close to 1% of the total protein in yeast. The amount of monatin produced when the cell extract of the *aspC* construct (with truncated *proA*) was assayed with aldolase was significantly greater than when cell extracts were assayed alone, and indicates that the recombinant AspC aminotransferase comprises approximately 1-2% of the yeast total protein. The cell extracts of uninduced cultures had a small amount of activity when assayed with aldolase due to the presence of native aminotransferases in the cells. When assayed with AspC aminotransferase, the activity of the extracts from uninduced cells increased to the amount of monatin produced by the negative control *with AspC* (ca. 200 ng/ml). In contrast, the activity observed when assaying the two gene construct cell extract increases more when aminotransferase is supplemented than when aldolase is added. Since both genes should be expressed at the same level, this indicates that the amount of monatin produced is maximized when the level of aminotransferase is higher than that of aldolase, in agreement with results shown in Example 6.

The addition of pyruvate and tryptophan not only inhibits cellular growth, but apparently inhibits protein expression as well. The addition of the pESC-Trp plasmid can be used to correct for tryptophan auxotrophy of the YPH500 host cells, to provide a means of supplying tryptophan with fewer effects on growth, expression, and secretion.

Example 13

Improvement of Enzymatic Processes using Coupled Reactions

In theory, if no side reactions or degradation of substrates or intermediates occurs, the maximum amount of product formed from the enzymatic reaction illustrated in FIG. 1 is directly proportional to the equilibrium constants of each reaction, and the concentrations of tryptophan and pyruvate. Tryptophan is not a highly soluble substrate, and concentrations of pyruvate greater than 200 mM appear to have a negative effect on the yield (see Example 6).

Ideally, the concentration of monatin is maximized with respect to substrates, in order to decrease the cost of separation. Physical separations can be performed such that the monatin is removed from the reaction mixture, preventing the reverse reactions from occurring. The raw materials and catalysts can then be regenerated. Due to the similarity of monatin in size, charge, and hydrophobicity to several of the reagents and intermediates, physical separations will be difficult unless there is a high amount of affinity for monatin (such as an affinity chromatography technique). However, the monatin reactions can be coupled to other reactions such that the equilibrium of the system is shifted toward monatin production. The following are examples of processes for improving the yield of monatin obtained from tryptophan or indole-3-pyruvate.

Coupled reactions using oxaloacetate decarboxylase (EC 4.1.1.3)

FIG. 11 is an illustration of the reaction. Tryptophan oxidase and catalase are utilized to drive the reaction in the direction of indole-3-pyruvate production. Catalase is used in excess such

that hydrogen peroxide is not available to react in the reverse direction or to damage the enzymes or intermediates. Oxygen is regenerated during the catalase reaction. Alternatively, indole-3-pyruvate can be used as the substrate.

Aspartate is used as the amino donor for the amination of MP, and an aspartate aminotransferase is utilized. Ideally, an aminotransferase that has a low specificity for the tryptophan/indole-3-pyruvate reaction in comparison to the MP to monatin reaction is used so that the aspartate is not utilized to reaminate the indole-3-pyruvate. Oxaloacetate decarboxylase (from *Pseudomonas sp.*) can be added to convert the oxaloacetate to pyruvate and carbon dioxide. Since CO_2 is volatile, it is not available for reaction with the enzymes, decreasing or even preventing the reverse reactions. The pyruvate produced in this step can also be utilized in the aldol condensation reaction. Other decarboxylase enzymes can be used, and homologs are known to exist in *Actinobacillus actinomycetemcomitans, Aquifex aeolicus, Archaeoglobus fulgidus, Azotobacter vinelandii, Bacteroides fragilis,* several *Bordetella* species, *Campylobacter jejuni, Chlorobium tepidum, Chloroflexus aurantiacus, Enterococcus faecalis, Fusobacterium nucleatum, Klebsiella pneumoniae, Legionella pneumophila, Magnetococcus MC-1, Mannheimia haemolytica, Methylobacillus flagellatus KT, Pasteurella multocida* Pm70, *Petrotoga miotherma, Porphyromonas gingivalis,* several *Pseudomonas* species, several *Pyrococcus* species, *Rhodococcus,* several *Salmonella* species, several *Streptococcus* species, *Thermochromatium tepidum, Thermotoga maritima, Treponema pallidum,* and several *Vibrio* species.

Tryptophan aminotransferase assays were performed with the aspartate aminotransferase (AspC) from *E. coli,* the tyrosine aminotransferase (TyrB) from *E. coli,* the broad substrate aminotransferase (BSAT) from *L. major,* and the two commercially available porcine glutamate-oxaloacetate aminotransferases as described in Example 1. Both oxaloacetate and alpha-ketoglutarate were tested as the amino acceptor. The ratio of activity using monatin (Example 7) versus activity using tryptophan was compared, to determine which enzyme had the highest specificity for the monatin aminotransferase reaction. These results indicated that the enzyme with the highest specificity for the monatin reaction verses the tryptophan reaction is the Porcine type II-A glutamate-oxaloacetate aminotransferase, GOAT (Sigma G7005). This specificity was independent of which amino acceptor was utilized. Therefore, this enzyme was used in the coupled reactions with oxaloacetate decarboxylase.

A typical reaction starting from indole-3-pyruvate included (final concentrations) 50 mM Tris-Cl pH 7.3, 6 mM indole-3-pyruvate, 6 mM sodium pyruvate, 6 mM aspartate, 0.05 mM PLP, 3 mM potassium phosphate, 3 mM $MgCl_2$, 25 μg/mL aminotransferase, 50 μg/mL *C. testosteroni* ProA aldolase, and 3 Units/mL of decarboxylase (Sigma O4878). The reactions were allowed to proceed for 1 hour at 26°C. In some cases, the decarboxylase was omitted or the aspartate was substituted with alpha-ketoglutarate (as negative controls). The aminotransferase enzymes described above were also tested in place of the GOAT to confirm earlier specificity experiments. Samples were filtered and analyzed by LC/MS as described in Example 10. The results demonstrate that the GOAT enzyme produced the highest amount of monatin per mg of protein, with the least amount of

tryptophan produced as a byproduct. In addition, there was a 2-3 fold benefit from having the decarboxylase enzyme added. The *E. coli* AspC enzyme also produced large amounts of monatin in comparison to the other aminotransferases.

Monatin production was increased by: 1) periodically adding 2 mM additions of indole-pyruvate, pyruvate, and aspartate (every half hour to hour), 2) performing the reactions in an anaerobic environment or with degassed buffers, 3) allowing the reactions to proceed overnight, and 4) using freshly prepared decarboxylase that has not been freeze-thawed multiple times. The decarboxylase was inhibited by concentrations of pyruvate greater than 12 mM. At concentrations of indole-3-pyruvate higher than 4 mM, side reactions with indole-3-pyruvate were hastened. The amount of indole-3-pyruvate used in the reaction could be increased if the amount of aldolase was also increased. High levels of phosphate (50 mM) and aspartate (50 mM) were found to be inhibitory to the decarboxylase enzyme. The amount of decarboxylase enzyme added could be reduced to 0.5 U/mL with no decrease in monatin production in a one hour reaction. The amount of monatin produced increased when the temperature was increased from 26°C to 30°C and from 30°C to 37°C; however, at 37°C the side reactions of indole-3-pyruvate were also hastened. The amount of monatin produced increased with increasing pH from 7 to 7.3, and was relatively stable from pH 7.3-8.3.

A typical reaction starting with tryptophan included (final concentrations) 50 mM Tris-Cl pH 7.3, 20 mM tryptophan, 6 mM aspartate, 6 mM sodium pyruvate, 0.05 mM PLP, 3 mM potassium phosphate, 3 mM $MgCl_2$, 25 μg/mL aminotransferase, 50 μg/mL *C. testosteroni* ProA aldolase, 4 Units/mL of decarboxylase, 5-200 mU/mL L-amino acid oxidase (Sigma A-2805), 168 U/mL catalase (Sigma C-3515), and 0.008 mg FAD. Reactions were carried out for 30 minutes at 30°C. Improvement was observed with the addition of decarboxylase. The greatest amount of monatin was produced when 50 mU/mL of oxidase was used. Improvements were similar to those observed when indole-3-pyruvate was used as the substrate. In addition, the amount of monatin produced increased when 1) the tryptophan level was low (i.e., below the K_m of the aminotransferase enzyme and therefore unable to compete with MP in the active site), and 2) the ratio of oxidase to aldolase and aminotransferase was maintained at a level such that indole-3-pyruvate could not accumulate.

Whether starting with either indole-3-pyruvate or tryptophan, the amount of monatin produced in assays with incubation times of 1-2 hours increased when 2-4 times the amounts of all the enzymes were used while maintaining the same enzyme ratio. Using either substrate, concentrations of approximately 1 mg/mL of monatin were achieved. The amount of tryptophan produced if starting from indole-pyruvate was typically less than 20% of the amount of product, which shows the benefit of utilizing coupled reactions. With further optimization and control of the concentrations of intermediates and side reactions, the productivity and yield can be improved greatly.

Coupled reactions using lysine epsilon aminotransferase (EC 2.6.1.36)

Lysine epsilon aminotransferase (L-Lysine 6-transaminase) is found in several organisms, including *Rhodococcus, Mycobacterium, Streptomyces, Nocardia, Flavobacterium, Candida utilis,*

and *Streptomyces*. It is utilized by organisms as the first step in the production of some beta-lactam antibiotics (Rius and Demain, *J. Microbiol. Biotech.*, 7:95-100, 1997). This enzyme converts lysine to L-2-aminoadipate 6-semialdehyde (allysine), by a PLP-mediated transamination of the C-6 of lysine, utilizing alpha-ketoglutarate as the amino acceptor. Allysine is unstable and spontaneously undergoes an intramolecular dehydration to form 1-piperideine 6-carboxylate, a cyclic molecule. This effectively inhibits any reverse reaction from occuring. The reaction scheme is depicted in FIG. 12. An alternative enzyme, lysine-pyruvate 6-transaminase (EC 2.6.1.71), can also be used.

A typical reaction contained in 1 mL: 50 mM Tris-HCl pH 7.3, 20 mM indole-3-pyruvate, 0.05 mM PLP, 6 mM potassium phosphate pH 8, 2-50 mM sodium pyruvate, 1.5 mM $MgCl_2$, 50 mM lysine, 100 µg aminotransferase (lysine epsilon aminotransferase LAT-101, BioCatalytics Pasadena, CA), and 200 µg *C. testosteroni* ProA aldolase. The amount of monatin produced increased with increasing concentrations of pyruvate. The maximum amount using these reaction conditions (at 50 mM pyruvate) was 10-fold less than what was observed with coupled reactions using oxaloacetate decarboxylase (approximately 0.1 mg/mL).

A peak with $[M+H]^+ = 293$ eluted at the expected time for monatin and the mass spectrum contained several of the same fragments observed with other enzymatic processes. A second peak with the correct mass to charge ratio (293) eluted slightly earlier than what is typically observed for the S,S monatin produced in Example 6, and may indicate the presence of another isomer of monatin. Very little tryptophan was produced by this enzyme. However, there is likely some activity on pyruvate (producing alanine as a byproduct). Also, the enzyme is known to be unstable. Improvements can be made by performing directed evolution experiments to increase stability, reduce the activity with pyruvate, and increase the activity with MP. These reactions can also be coupled to L-amino acid oxidase/catalase as described above.

Other coupled reactions

Another coupling reaction that can improve monatin yield from tryptophan or indole-pyruvate is shown in FIG. 13. Formate dehydrogenase (EC 1.2.1.2 or 1.2.1.43) is a common enzyme. Some formate dehydrogenases require NADH while others can utilize NADPH. Glutamate dehydrogenase catalyzed the interconversion between the monatin precursor and monatin in previous examples, using ammonium based buffers. The presence of ammonium formate and formate dehydrogenase is an efficient system for regeneration of cofactors, and the production of carbon dioxide is an efficient way to decrease the rate of the reverse reactions (Bommarius *et al.*, *Biocatalysis* 10:37, 1994 and Galkin *et al. Appl. Environ. Microbiol.* 63:4651-6, 1997). In addition, large amounts of ammonium formate can be dissolved in the reaction buffer. The yield of monatin produced by glutamate dehydrogenase reactions (or similar reductive aminations) can be improved by the addition of formate dehydrogenase and ammonium formate.

Other processes can be used to drive the equilibrium toward monatin production. For instance, if aminopropane is utilized as the amino acid donor in the conversion of MP to monatin with

an omega-amino acid aminotransferase (EC 2.6.1.18) such as those described by in US patents 5,360,724 and 5,300,437, one of the resulting products would be acetone, a more volatile product than the substrate, aminopropane. The temperature can be raised periodically for short periods to flash off the acetone, thereby alleviating equilibrium. Acetone has a boiling point of 47°C, a temperature not likely to degrade the intermediates if used for short periods of time. Most aminotransferases that have activity on alpha-ketoglutarate also have activity on the monatin precursor. Similarly, if a glyoxylate/aromatic acid aminotransferase (EC 2.6.1.60) is used with glycine as the amino donor, glyoxylate is produced which is relatively unstable and has a highly reduced boiling point in comparison to glycine.

EXAMPLE 14

Recombinant Expression

With publicly available enzyme cDNA and amino acid sequences, and the enzymes and sequences disclosed herein, such as SEQ ID NOS: 11 and 12, as well as variants, polymorphisms, mutants, fragments and fusions thereof, the expression and purification of any protein, such as an enzyme, by standard laboratory techniques is enabled. One skilled in the art will understand that enzymes and fragments thereof can be produced recombinantly in any cell or organism of interest, and purified prior to use, for example prior to production of SEQ ID NO: 12 and derivatives thereof.

Methods for producing recombinant proteins are well known in the art. Therefore, the scope of this disclosure includes recombinant expression of any protein or fragment thereof, such as an enzyme. For example, see U.S. Patent No: 5,342,764 to Johnson *et al.*; U.S. Patent No: 5,846,819 to Pausch *et al.*; U.S. Patent No: 5,876,969 to Fleer *et al.* and Sambrook *et al.* (*Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor, New York, 1989, Ch. 17).

Briefly, partial, full-length, or variant cDNA sequences, which encode for a protein or peptide, can be ligated into an expression vector, such as a bacterial or eukaryotic expression vector. Proteins and/or peptides can be produced by placing a promoter upstream of the cDNA sequence. Examples of promoters include, but are not limited to *lac*, *trp*, *tac*, *trc*, major operator and promoter regions of phage lambda, the control region of fd coat protein, the early and late promoters of SV40, promoters derived from polyoma, adenovirus, retrovirus, baculovirus and simian virus, the promoter for 3-phosphoglycerate kinase, the promoters of yeast acid phosphatase, the promoter of the yeast alpha-mating factors and combinations thereof.

Vectors suitable for the production of intact native proteins include pKC30 (Shimatake and Rosenberg, 1981, *Nature* 292:128), pKK177-3 (Amann and Brosius, 1985, *Gene* 40:183) and pET-3 (Studier and Moffatt, 1986, *J. Mol. Biol.* 189:113). A DNA sequence can be transferred to other cloning vehicles, such as other plasmids, bacteriophages, cosmids, animal viruses and yeast artificial chromosomes (YACs) (Burke *et al.*, 1987, *Science* 236:806-12). These vectors can be introduced into a variety of hosts including somatic cells, and simple or complex organisms, such as bacteria, fungi (Timberlake and Marshall, 1989, *Science* 244:1313-7), invertebrates, plants (Gasser and Fraley,

1989, *Science* 244:1293), and mammals (Pursel *et al.*, 1989, *Science* 244:1281-8), which are rendered transgenic by the introduction of the heterologous cDNA.

For expression in mammalian cells, a cDNA sequence can be ligated to heterologous promoters, such as the simian virus SV40, promoter in the pSV2 vector (Mulligan and Berg, 1981, *Proc. Natl. Acad. Sci. USA* 78:2072-6), and introduced into cells, such as monkey COS-1 cells (Gluzman, 1981, *Cell* 23:175-82), to achieve transient or long-term expression. The stable integration of the chimeric gene construct may be maintained in mammalian cells by biochemical selection, such as neomycin (Southern and Berg, 1982, *J. Mol. Appl. Genet.* 1:327-41) and mycophoenolic acid (Mulligan and Berg, 1981, *Proc. Natl. Acad. Sci. USA* 78:2072-6).

The transfer of DNA into eukaryotic, such as human or other mammalian cells, is a conventional technique. The vectors are introduced into the recipient cells as pure DNA (transfection) by, for example, precipitation with calcium phosphate (Graham and vander Eb, 1973, *Virology* 52:466) strontium phosphate (Brash *et al.*, 1987, *Mol. Cell Biol.* 7:2013), electroporation (Neumann *et al.*, 1982, *EMBO J.* 1:841), lipofection (Felgner *et al.*, 1987, *Proc. Natl. Acad. Sci USA* 84:7413), DEAE dextran (McCuthan *et al.*, 1968, *J. Natl. Cancer Inst.* 41:351), microinjection (Mueller *et al.*, 1978, *Cell* 15:579), protoplast fusion (Schafner, 1980, *Proc. Natl. Acad. Sci. USA* 77:2163-7), or pellet guns (Klein *et al.*, 1987, *Nature* 327:70). Alternatively, the cDNA can be introduced by infection with virus vectors, for example retroviruses (Bernstein *et al.*, 1985, *Gen. Engrg.* 7:235) such as adenoviruses (Ahmad *et al.*, 1986, *J. Virol.* 57:267) or Herpes (Spaete *et al.*, 1982, *Cell* 30:295).

In view of the many possible embodiments to which the principles of our disclosure may be applied, it should be recognized that the illustrated embodiments are only particular examples of the disclosure and should not be taken as a limitation on the scope of the disclosure. Rather, the scope of the disclosure is in accord with the following claims. We therefore claim as our invention all that comes within the scope and spirit of these claims.

We claim:

1. A method for producing monatin comprising converting glucose to monatin.

2. A method of producing monatin, comprising converting tryptophan and/or indole-3-lactic acid to monatin, wherein the method comprises:

contacting tryptophan or indole-3-lactic acid with a first polypeptide, wherein the first polypeptide converts the tryptophan or indole-3-lactic acid to indole-3-pyruvate;

contacting the indole-3-pyruvate with a second polypeptide and a C3 carbon source, wherein the second polypeptide converts the indole-3-pyruvate and the C3 carbon source to 2-hydroxy 2-(indol-3-ylmethyl)-4-keto glutaric acid; and

contacting the 2-hydroxy 2-(indol-3ylmethyl)-4-keto glutaric acid with a third polypeptide, wherein the third polypeptide converts the 2-hydroxy 2-(indol-3ylmethyl)-4-keto glutaric acid to monatin.

3. The method of claim 2, wherein the first polypeptide is selected from the group consisting of tryptophan aminotransferase (EC 2.6.1.27), tryptophan dehydrogenase (EC 1.4.1.19), tyrosine (aromatic) aminotransferase (EC 2.6.1.5), tryptophan-phenylpyruvate transaminase (EC 2.6.1.28), multiple substrate aminotransferase (EC 2.6.1.-), aspartate aminotransferase (EC 2.6.1.1), tryptophan oxidase, L-amino acid oxidase (EC 1.4.3.2), D-amino acid dehydrogenase (EC 1.4.99.1), D-amino acid oxidase (EC 1.4.3.3), D-tryptophan aminotransferase, D-alanine aminotransferase (EC 2.6.1.21), glutamate dehydrogenase (EC 1.4.1.2-4), phenylalanine dehydrogenase (EC 1.4.1.20), and a combination thereof, and wherein the first polypeptide converts tryptophan to indole-3-pyruvate.

4. The method of claim 2, wherein the first polypeptide is selected from the group consisting of indolelactate dehydrogenase (EC 1.1.1.110), R-4-hydroxyphenyllactate dehydrogenase (EC 1.1.1.222), 3-(4)-hydroxyphenylpyruvate reductase (EC 1.1.1.237), lactate dehydrogenase (EC 1.1.1.27, 1.1.1.28, 1.1.2.3), (3-imidazol-5-yl) lactate dehydrogenase (EC 1.1.1.111), and lactate oxidase (EC 1.1.3.-), or a combination thereof, and wherein the first polypeptide converts indole-3-lactic acid to indole-3-pyruvate.

5. The method of claim 2 or 3, wherein the first and/or third polypeptide comprises:

an amino acid sequence shown in SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 32, Genbank Accession No: NP_388848.1, Genbank Accession No: ZP00005082.1, or Genbank Accession No: AAC74014.1;

an amino acid sequence comprising at least 90% sequence identity with the sequence shown in SEQ ID NO:2, 4, 6, 8, 10, 12, 14, 32, Genbank Accession No: NP_388848.1, Genbank Accession No: ZP00005082.1, or Genbank Accession No: AAC74014.1; or

amino acid sequences that differ from SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 32, Genbank Accession No: NP_388848.1, Genbank Accession No: ZP00005082.1, or Genbank Accession No: AAC74014.1 by less than 50 conservative amino acid substitutions, wherein the amino acid sequence converts tryptophan to indole-3-pyruvate.

6. The method of claim 2, wherein the second polypeptide is an EC 4.1.2.- or EC 4.1.3.- enzyme.

7. The method of claim 6, wherein the EC 4.1.3.- enzyme is 4-hydroxy-2-oxoglutarate glyoxylate-lyase (EC 4.1.3.16), 4-hydroxy-4-methyl-2-oxoglutarate pyruvate-lyase (EC 4.1.3.17) or a combination thereof.

8. The method of claim 2, wherein the third polypeptide is selected from the group consisting of tyrosine (aromatic) aminotransferase (EC 2.6.1.5), tryptophan dehydrogenase (EC 1.4.1.19), tryptophan-phenylpyruvate transaminase (EC 2.6.1.28), tryptophan aminotransferase (EC 2.6.1.27), aspartate aminotransferase (2.6.1.1), D-alanine aminotransferase (EC 2.6.1.21), D-tryptophan aminotranaferase, glutamate dehydrogenase (EC 1.4.1.2-4), phenylalanine dehydrogenase (EC 1.4.1.20), multiple substrate aminotransferase (EC 2.6.1.-), D-amino acid dehydrogenase (EC 1.4.99.1), and a combination thereof.

9. The method of claim 2, wherein the first polypeptide comprises an amino acid oxidase (EC 1.4.3.-) and a catalase (EC 1.11.1.6), and wherein the first polypeptide converts tryptophan to indole-3-pyruvate.

10. The method of claim 9, wherein the amino acid oxidase comprises a tryptophan oxidase.

11. The method of claim 2, wherein the C3 carbon source is selected from the group consisting of pyruvate, phosphoenolpyruvate, alanine, serine, cysteine, or a combination thereof.

12. The method of claim 11, wherein the pyruvate is produced by alanine and a polypeptide capable of transamination of alanine; serine and a polypeptide capable of beta-elimination of serine; cysteine and a polypeptide capable of beta-elimination of cysteine; aspartate and a polypeptide capable of beta-lyase reactions with a dicarboxylic amino acid; lactate and lactate oxidase (EC 1.1.3.-) or lactate dehydrogenase (EC 1.1.1.27, 1.1.1.28, 1.1.2.3); or a combination thereof.

13. The method of claim 2, wherein the method further comprises reducing an amount of hydrogen peroxide produced when tryptophan is converted to indole-3-pyruvate, by contacting the hydrogen peroxide with a catalase.

14. The method of claim 2, wherein the second polypeptide comprises an amino acid sequence selected from the group consisting of:

SEQ ID NO: 66, Genbank Accession No: CAC46344, Genbank Accession No: CAB14127.1, Genbank Accession No: AAC74920.1, or Genbank Accession No: CAC47463.1;

an amino acid sequence comprising at least 90% sequence identity to SEQ ID NO: 66, Genbank Accession No: CAC46344, Genbank Accession No: CAB14127.1, Genbank Accession No: AAC74920.1, or Genbank Accession No: CAC47463.1; or

amino acid sequences that differ from SEQ ID NO: 66, Genbank Accession No: CAC46344, Genbank Accession No: CAB14127.1, Genbank Accession No: AAC74920.1, or Genbank Accession No: CAC47463.1 by less than 50 conservative amino acid substitutions, wherein the amino acid sequence converts indole-3-pyruvate and pyruvate to MP.

15. The method of claim 8, wherein the third polypeptide comprises an aspartate aminotransferase, wherein the 2-hydroxy 2-(indol-3-ylmethyl)-4-keto glutaric acid is contacted with aspartate to produce oxaloacetate and monatin.

16. The method of claim 15, further comprising contacting oxaloacetate with a decarboxylase.

17. The method of claim 2, wherein the third polypeptide comprises a lysine epsilon aminotransferase, and wherein the 2-hydroxy 2-(indol-3-ylmethyl)-4-keto glutaric acid is further contacted with lysine.

18. The method of claim 2, wherein the third polypeptide comprises an enzyme having reductive amination activity and a polypeptide that is capable of recycling NAD(P)H.

19. The method of claim 18, wherein the enzyme having reductive amination activity is selected from the group consisting of glutamate dehydrogenase, phenylalanine dehydrogenase, tryptophan dehydrogenase, or a combination thereof.

20. The method of claim 18, wherein the polypeptide capable of recycling NAD(P)H further produces a volatile product.

21. The method of claim 20, wherein the polypeptide that is capable of recycling NAD(P)H and producing a volatile product comprises formate dehydrogenase, and wherein the volatile product produced is carbon dioxide.

22. The method of claim 2, wherein the method comprises converting indole-3-lactic acid to monatin.

23. The method of claim 2, where at least 10 g/L of monatin is produced.

24. A method of producing monatin comprising converting indole-3-pyruvate to monatin.

25. A method of producing monatin comprising converting tryptophan to monatin.

26. A cell capable of producing monatin, wherein the cell comprises at least one exogenous nucleic acid sequence encoding at least one polypeptide, wherein the at least one polypeptide converts a first intermediate in a pathway to make monatin into a second intermediate or into monatin.

27. The cell of claim 26, wherein the at least one polypeptide is selected from the group consisting of: tryptophan aminotransferase (EC 2.6.1.27), tryptophan dehydrogenase (EC 1.4.1.19), tyrosine (aromatic) aminotransferase (EC 2.6.1.5), tryptophan-phenylpyruvate transaminase (EC 2.6.1.28), multiple substrate aminotransferase (EC 2.6.1.-), tryptophan oxidase, L-amino acid oxidase (EC 1.4.3.2), aspartate aminotransferase (EC 2.6.1.1), D-amino acid dehydrogenase (EC 1.4.99.1), D-amino acid oxidase (EC 1.4.3.3), D-alanine aminotransferase (EC 2.6.1.21), D-tryptophan aminotransferase, indolelactate dehydrogenase (EC 1.1.1.110), R-4-hydroxyphenyllactate dehydrogenase (EC 1.1.1.222), 3-(4)-hydroxyphenylpyruvate reductase (EC 1.1.1.237), lactate dehydrogenase (EC 1.1.1.27, 1.1.1.28, 1.1.2.3), (3-imidazol-5-yl) lactate dehydrogenase (EC 1.1.1.111), lactate oxidase (EC 1.1.3.-), synthase/lyase (4.1.2.-), synthase/lyase (4.1.3.-), phenylalanine dehydrogenase (EC 1.4.1.20), glutamate dehydrogenase (EC 1.4.1.2, 1.4.1.3, 1.4.1.4), and a combination thereof.

28. The cell of claim 27, wherein the at least one polypeptide is selected from the group consisting of: tryptophan aminotransferase (EC 2.6.1.27), tryptophan dehydrogenase (EC 1.4.1.19), tyrosine (aromatic) aminotransferase (EC 2.6.1.5), tryptophan-phenylpyruvate transaminase (EC 2.6.1.28), multiple substrate aminotransferase (EC 2.6.1.-), aspartate aminotransferase (EC 2.6.1.1), tryptophan oxidase, L-amino acid oxidase (EC 1.4.3.2), D-amino acid dehydrogenase (EC 1.4.99.1), D-amino acid oxidase (EC 1.4.3.3), D-alanine aminotransferase (EC 2.6.1.21), D-tryptophan aminotransferase, synthase/lyase (EC 4.1.3.-), synthase/lyase (4.1.2.-), phenylalanine dehydrogenase (EC 1.4.1.20), glutamate dehydrogenase (EC 1.4.1.2, 1.4.1.3, 1.4.1.4) and a combination thereof.

29. The cell of claim 27, wherein the at least one polypeptide comprises an activity selected from the group consisting of: indolelactate dehydrogenase (EC 1.1.1.110), R-4-hydroxyphenyllactate dehydrogenase (EC 1.1.1.222), 3-(4)-hydroxyphenylpyruvate reductase (EC 1.1.1.237), lactate dehydrogenase (EC 1.1.1.27, 1.1.1.28, 1.1.2.3), (3-imidazol-5-yl) lactate dehydrogenase (EC 1.1.1.111), lactate oxidase (EC 1.1.3.-), synthase/lyase (4.1.2.-), synthase/lyase (4.1.3.-), tryptophan dehydrogenase (EC 1.4.1.19), tryptophan-phenylpyruvate transaminase (EC 2.6.1.28), tryptophan aminotransferase (EC 2.6.1.27), tyrosine (aromatic) aminotransferase (EC 2.6.1.5), multiple substrate aminotransferase (EC 2.6.1.-), aspartate aminotransferase (EC 2.6.1.1), glutamate dehydrogenase (EC

1.4.1.2, 1.4.1.3, 1.4.1.4), phenylalanine dehydrogenase (EC 1.4.1.20), D-tryptophan aminotransferase, D-amino acid dehydrogenase (EC 1.4.99.1), *D*-alanine aminotransferase (EC 2.6.1.21) and a combination thereof.

30. The cell of claim 27, wherein the synthase/lyase (EC 4.1.3.-) comprises 4-hydroxy-2-oxoglutarate aldolase (EC 4.1.3.16) or 4-hydroxy-4-methyl-2-oxoglutarate aldolase (EC 4.1.3.17).

31. The recombinant cell of claim 26, wherein the cell is a bacterial cell.

32. The recombinant cell of claim 26, wherein the cell is a yeast cell.

33. The cell of claim 27, wherein the cell further comprises one or more polypeptides that are encoded on at least one exogenous nucleic acid sequence, wherein the polypeptides comprise activities selected from the group consisting of: catalase, oxaloacetate decarboxylase, lysine epsilon aminotransferase, and formate dehydrogenase.

34. A polypeptide comprising an amino acid sequence selected from the group consisting of Genbank Accession No: P00913), Genbank Accession No: NP_290344, Genbank Accession No: CAA34096, Genbank Accession No: NP_246359, Genbank Accession No: AAB96579, Genbank Accession No: NP_232561, Genbank Accession No: Q59342, Genbank Accession No: P28796Genbank Accession No: 1AX4_A, Genbank Accession No: BAA34638, Genbank Accession No: P31014, Genbank Accession No: P31015, Genbank Accession No: NP_280989, Genbank Accession No: AAC33284, Genbank Accession No: NP_147838, Genbank Accession No: AAF63441, or Genbank Accession No: AAA24679, where the amino acid residues at position 103 and 299 with respect to Genbank Accession No: NP_418164 have been changed to broaden a specificity of the polypeptide to accept dicarboxylic amino acid substrates.

35. The polypeptide of claim 34, where the change in the amino acid sequence is selected from the group consisting of: arginine 103 changed to threonine, arginine 103 changed to threonine and valine 299 to arginine; and valine 299 to arginine.

36. A polypeptide comprising an amino acid sequence selected from the group consisting of Genbank Accession No: 2TPL_A, Genbank Accession No: P31012, Genbank Accession No: P31013, A49493, Genbank Accession No: P31011, Genbank Accession No: AAB24234, Genbank Accession No: 1TPL_A, , Genbank Accession No: 1TPL_B, Genbank Accession No: NP_245748, Genbank Accession No: AAB41499, Genbank Accession No: Q08897, Genbank Accession No: T45297, and Genbank Accession No: AAA71928, where the amino acid residues at position 100 and 283 with respect to Genbank Accession No: X66978 have been changed to broaden a specificity of the polypeptide to accept dicarboxylic acid amino acid substrates.

37. The polypeptide of claim 36, where the change in the amino acid sequence is selected from the group consisting of: arginine 100 changed to threonine, arginine 100 changed to threonine and valine 283 to arginine; and valine 283 to arginine.

38. The polypeptide of claim 34 or 36, wherein the substrate is aspartate.

39. A method of making monatin comprising reacting at least one alanine derivative blocked at carboxyl and amino groups, and at least one indole-3-pyruvate derivative blocked at a carboxyl group.

40. The method of making monatin according to claim 39, wherein the alanine derivative is 3-chloro-alanine or 3-bromo-alanine, and wherein the indole-3-pyruvate is reacted with the alanine derivative in a Grignard or organolithium reaction.

41. A method of making monatin comprising converting a substrate selected from the group consisting of tryptophan, indole-3-lactic acid, indole-3-pyruvate, and 2-hydroxy 2-(indol-3-ylmethyl)-4-keto glutaric acid, to one or more intermediates, wherein monatin is made.

42. The method of claim 41, wherein at least one of the intermediate products is converted via a chemical reaction.

43. The method of claim 41, wherein at least one of the intermediate products is converted via a polypeptide.

44. An isolated nucleic acid comprising at least 90% sequence identity to SEQ ID NO: 11.

45. The isolated nucleic acid of claim 44, wherein the nucleic acid comprises SEQ ID NO: 11.

46. The isolated nucleic acid of claim 44, wherein the nucleic acid comprises at least 20 consecutive nucleotides of SEQ ID NO: 11.

47. A vector comprising the isolated nucleic acid of claim 44.

48. A recombinant cell comprising the vector of claim 47.

49. An isolated protein encoded by the isolated nucleic acid of claim 44.

50. The isolated protein of claim 49, wherein the protein comprises a sequence that has at least 90% sequence identity to SEQ ID NO: 12.

51. The isolated protein of claim 50, wherein the protein comprises SEQ ID NO: 12.

52. A method for producing monatin comprising converting 2-hydroxy 2-(indol-3-ylmethyl)-4-keto glutaric acid to monatin.

53. The method of claim 52, wherein the method further comprises detection of the 2-hydroxy 2-(indol-3-ylmethyl)-4-keto glutaric acid by mass spectrometry.

54. A composition comprising 2-hydroxy 2-(indol-3-ylmethyl)-4-keto glutaric acid.

55. The method of claim 2, wherein the tryptophan is produced from indole using a polypeptide from EC 4.1.99.1.

56. A method for producing monatin comprising reacting tryptophan with one or more polypeptides.

57. The method of claim 56, wherein the one or more polypeptides is selected from the group consisting of : tryptophan aminotransferase (EC 2.6.1.27), tryptophan dehydrogenase (EC 1.4.1.19), tyrosine (aromatic) aminotransferase (EC 2.6.1.5), tryptophan-phenylpyruvate transaminase (EC 2.6.1.28), multiple substrate aminotransferase (EC 2.6.1.-), aspartate aminotransferase (EC 2.6.1.1), tryptophan oxidase, L-amino acid oxidase (EC 1.4.3.2), D-amino acid dehydrogenase (EC 1.4.99.1), D-amino acid oxidase (EC 1.4.3.3), D-alanine aminotransferase (EC 2.6.1.21), D-tryptophan aminotransferase, synthase/lyase (EC 4.1.3.-), synthase/lyase (4.1.2.-), phenylalanine dehydrogenase (EC 1.4.1.20), glutamate dehydrogenase (EC 1.4.1.2, 1.4.1.3, 1.4.1.4) and a combination thereof.

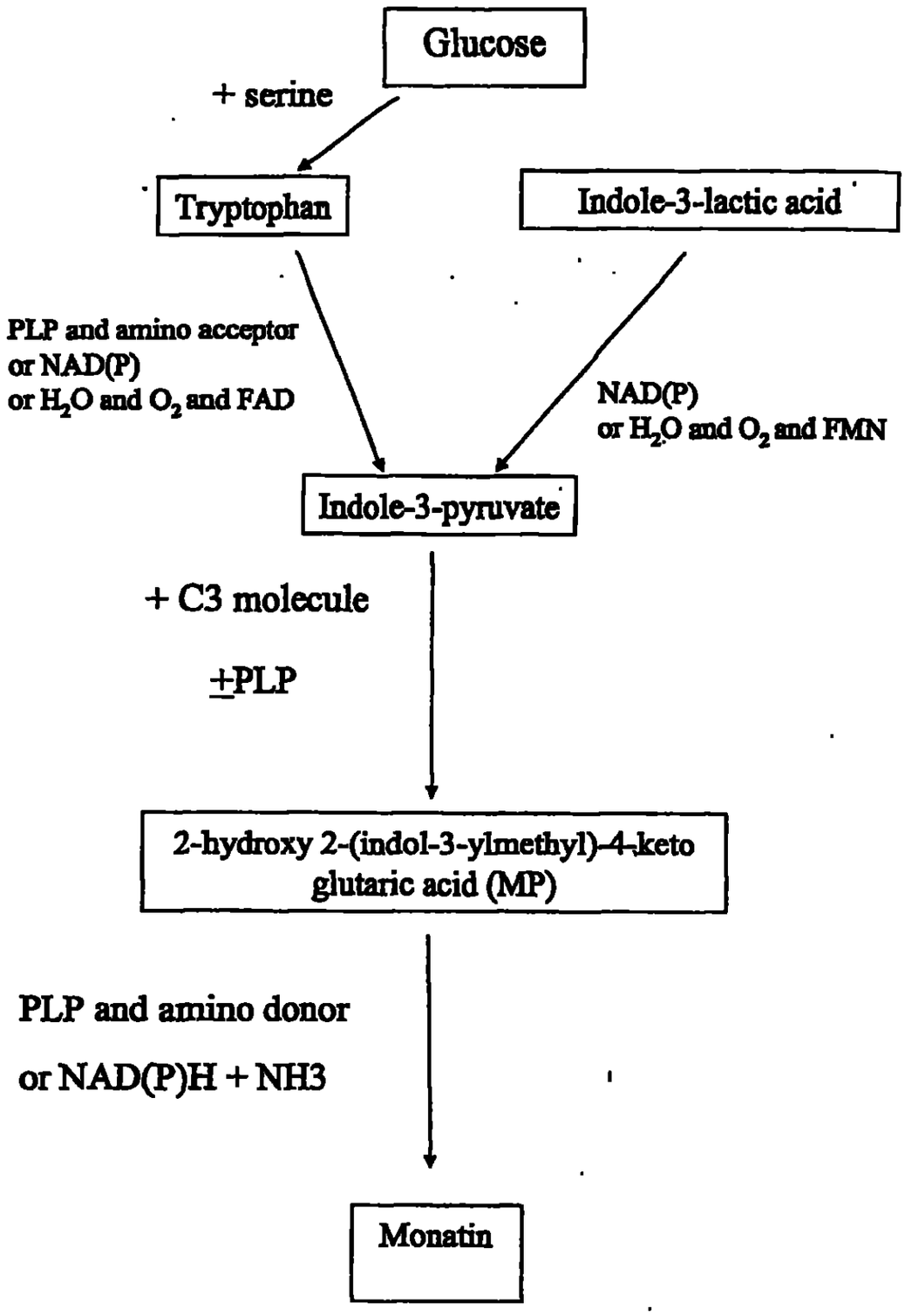

Glucose
+ serine
Tryptophan
Indole-3-lactic acid
PLP and amino acceptor
or NAD(P)
or H_2O and O_2 and FAD
NAD(P)
or H_2O and O_2 and FMN
Indole-3-pyruvate
+ C3 molecule
±PLP
2-hydroxy 2-(indol-3-ylmethyl)-4-keto glutaric acid (MP)
PLP and amino donor
or NAD(P)H + NH3
Monatin

FIG. 1

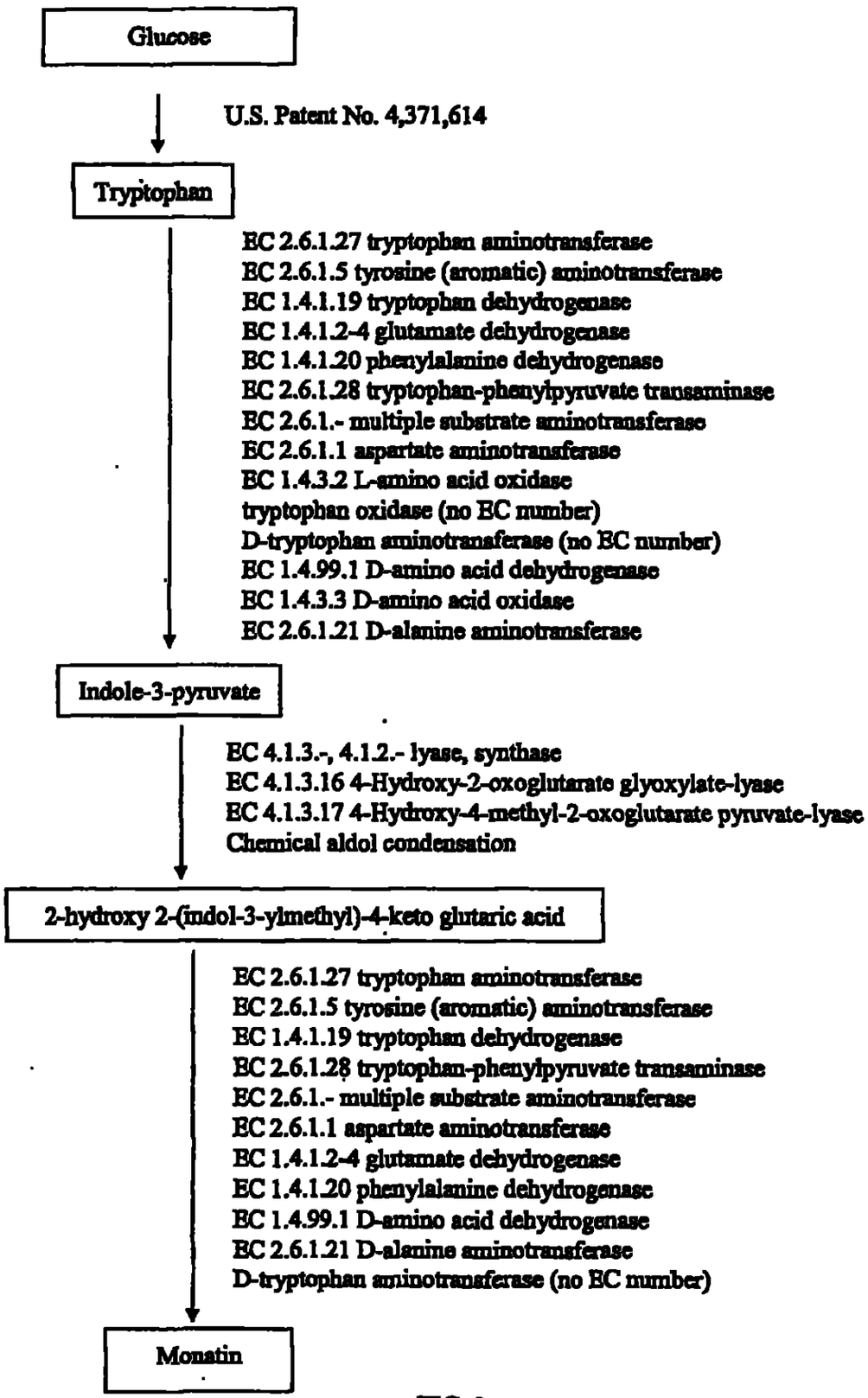

Glucose
U.S. Patent No. 4,371,614
Tryptophan
EC 2.6.1.27 tryptophan aminotransferase
EC 2.6.1.5 tyrosine (aromatic) aminotransferase
EC 1.4.1.19 tryptophan dehydrogenase
EC 1.4.1.2-4 glutamate dehydrogenase
EC 1.4.1.20 phenylalanine dehydrogenase
EC 2.6.1.28 tryptophan-phenylpyruvate transaminase
EC 2.6.1.- multiple substrate aminotransferase
EC 2.6.1.1 aspartate aminotransferase
EC 1.4.3.2 L-amino acid oxidase
tryptophan oxidase (no EC number)
D-tryptophan aminotransferase (no EC number)
EC 1.4.99.1 D-amino acid dehydrogenase
EC 1.4.3.3 D-amino acid oxidase
EC 2.6.1.21 D-alanine aminotransferase
Indole-3-pyruvate
EC 4.1.3.-, 4.1.2.- lyase, synthase
EC 4.1.3.16 4-Hydroxy-2-oxoglutarate glyoxylate-lyase
EC 4.1.3.17 4-Hydroxy-4-methyl-2-oxoglutarate pyruvate-lyase
Chemical aldol condensation
2-hydroxy 2-(indol-3-ylmethyl)-4-keto glutaric acid
EC 2.6.1.27 tryptophan aminotransferase
EC 2.6.1.5 tyrosine (aromatic) aminotransferase
EC 1.4.1.19 tryptophan dehydrogenase
EC 2.6.1.28 tryptophan-phenylpyruvate transaminase
EC 2.6.1.- multiple substrate aminotransferase
EC 2.6.1.1 aspartate aminotransferase
EC 1.4.1.2-4 glutamate dehydrogenase
EC 1.4.1.20 phenylalanine dehydrogenase
EC 1.4.99.1 D-amino acid dehydrogenase
EC 2.6.1.21 D-alanine aminotransferase
D-tryptophan aminotransferase (no EC number)
Monatin

FIG. 2

FIG. 3

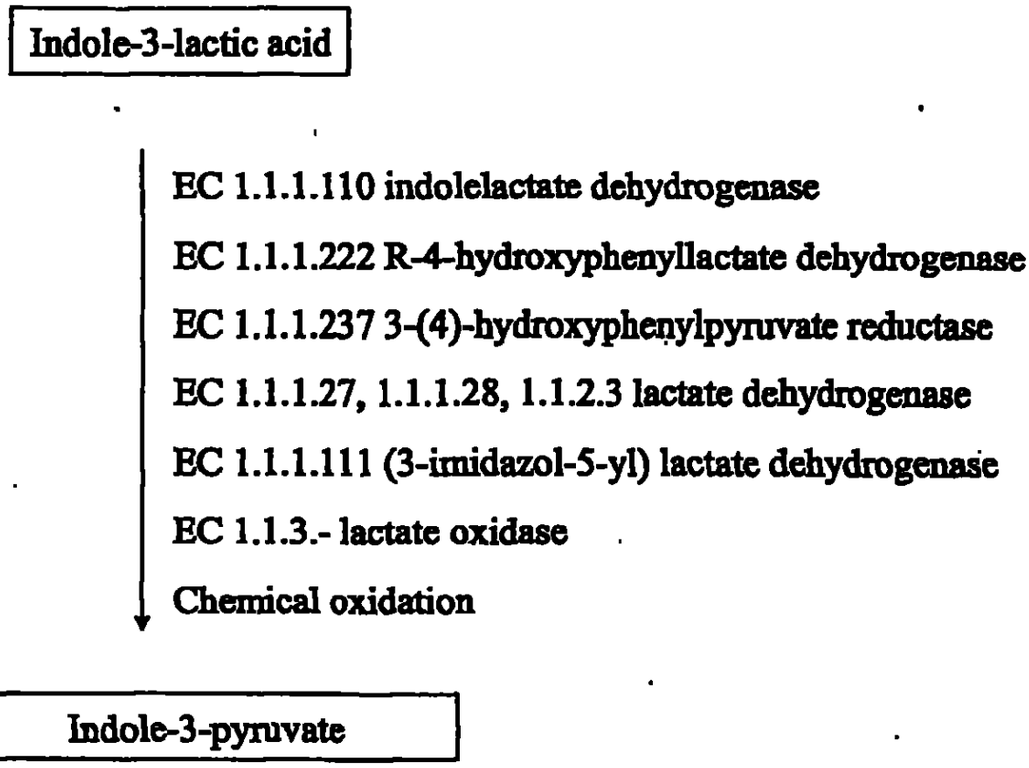
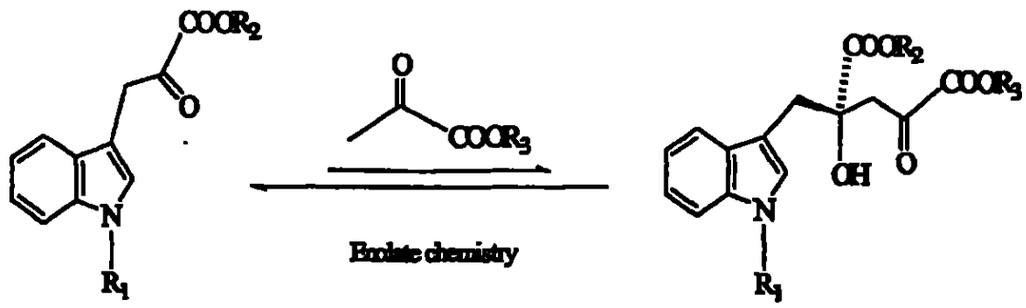


FIG. 4



R_1 = Boc, Cbz, etc.
R_2 and R_3 = Alkyl, Aryl, etc.

FIG. 5A

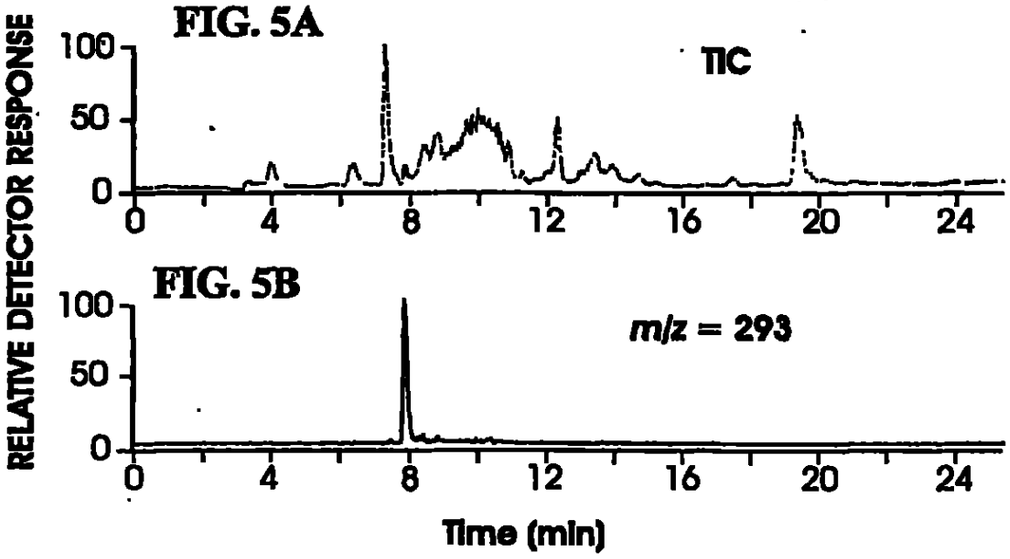


FIG. 5B

FIG. 6

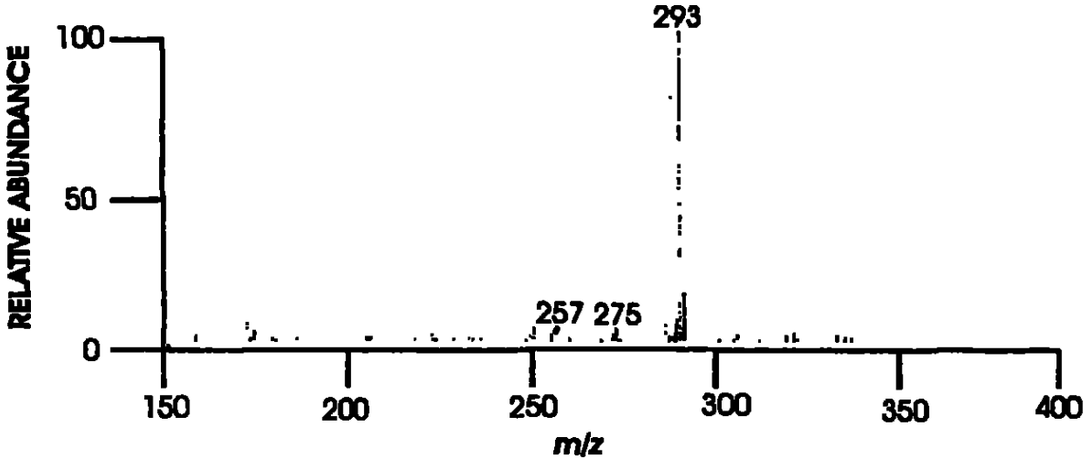


FIG. 7A

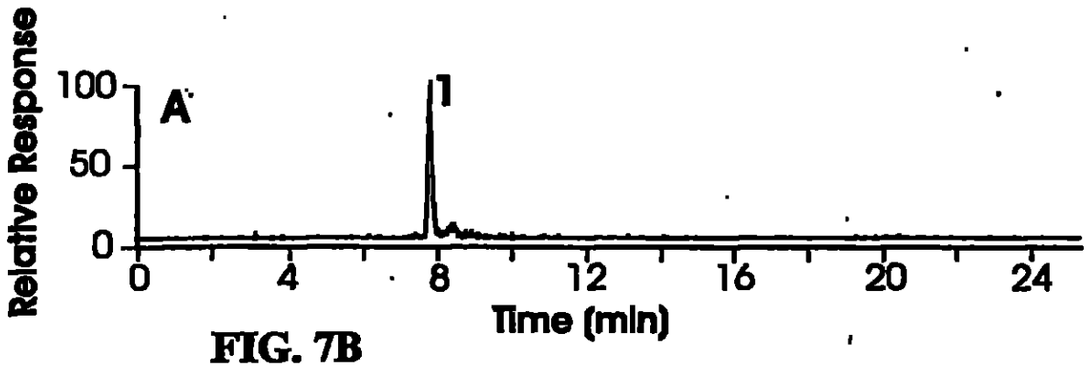


FIG. 7B

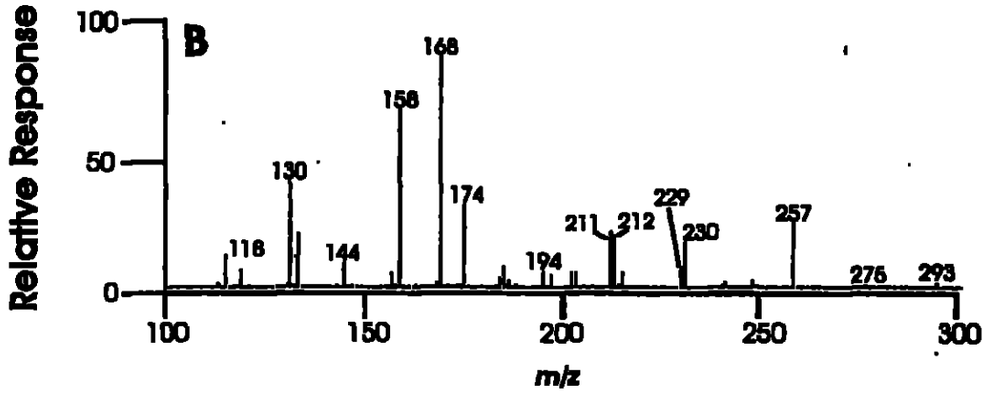
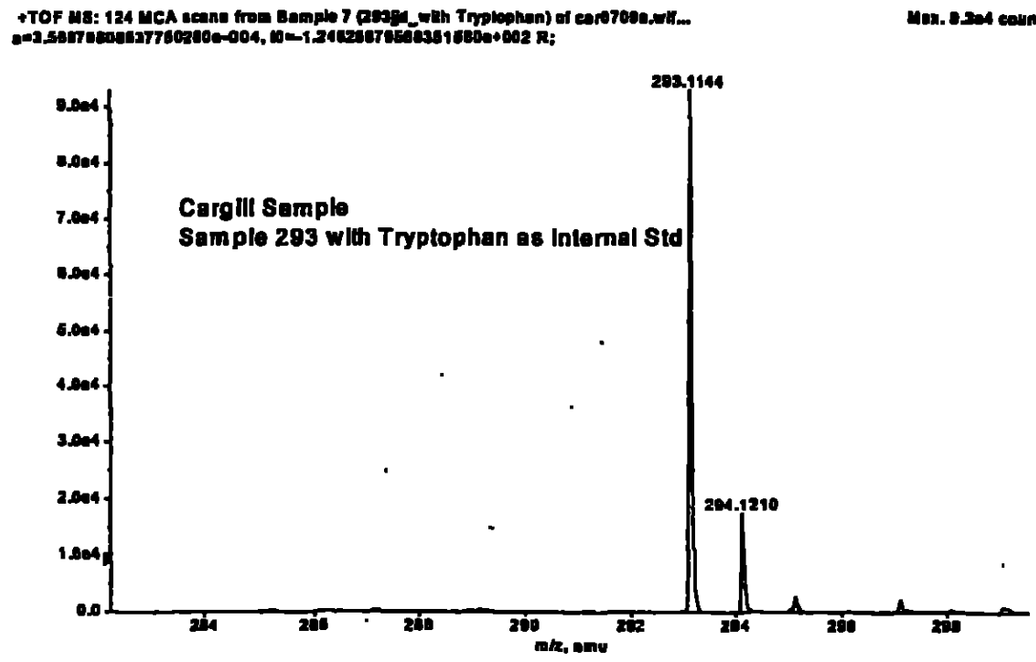


FIG. 8



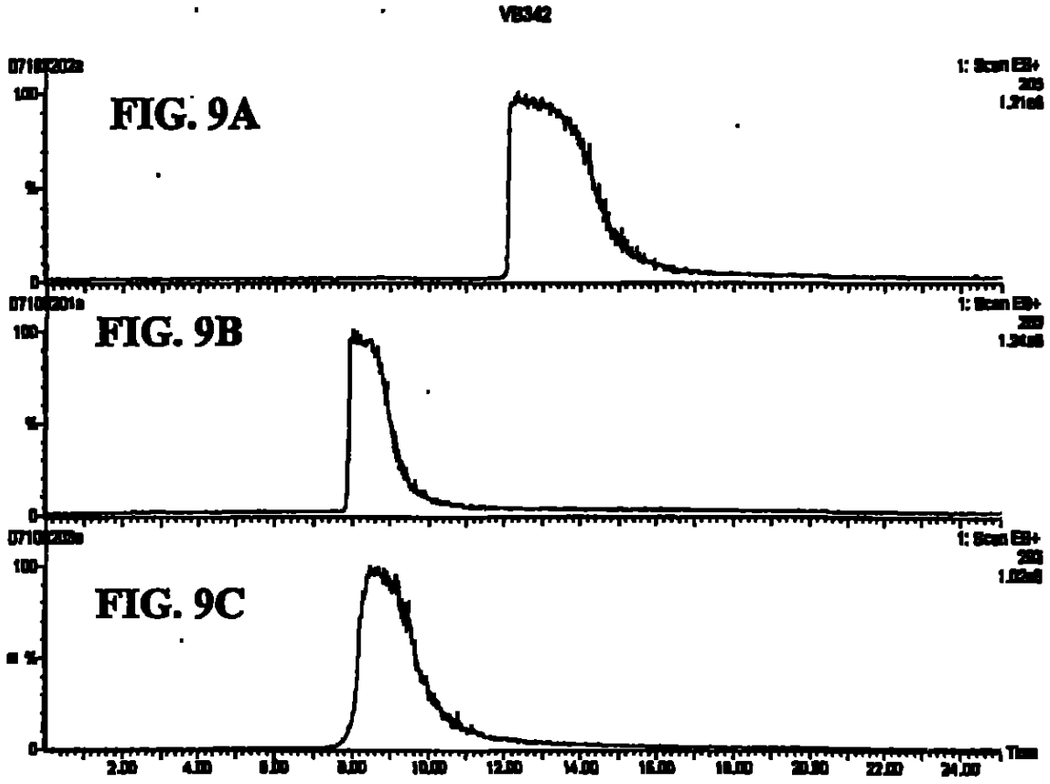


FIG. 9A
FIG. 9B
FIG. 9C

FIG. 10

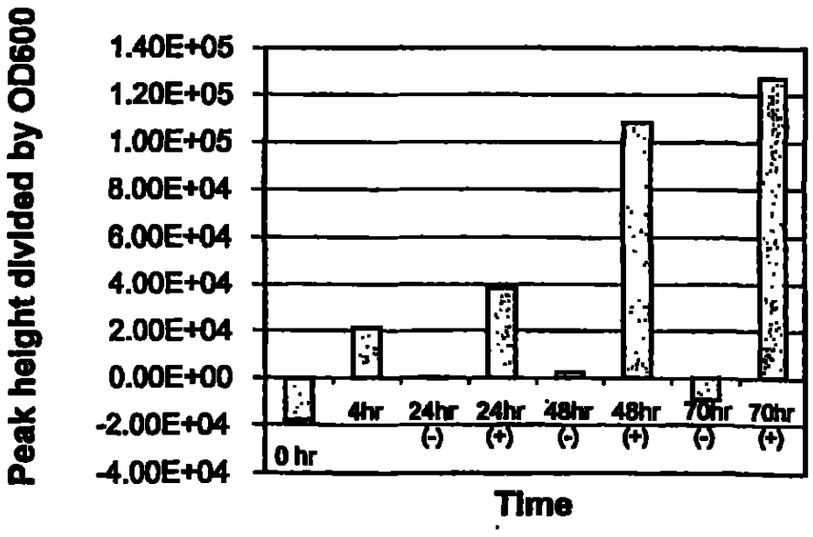

Peak height divided by OD600
1.40E+05
1.20E+05
1.00E+05
8.00E+04
6.00E+04
4.00E+04
2.00E+04
0.00E+00
-2.00E+04
-4.00E+04
0 hr
4hr
24hr (-)
24hr (+)
48hr (-)
48hr (+)
70hr (-)
70hr (+)
Time

FIG. 11

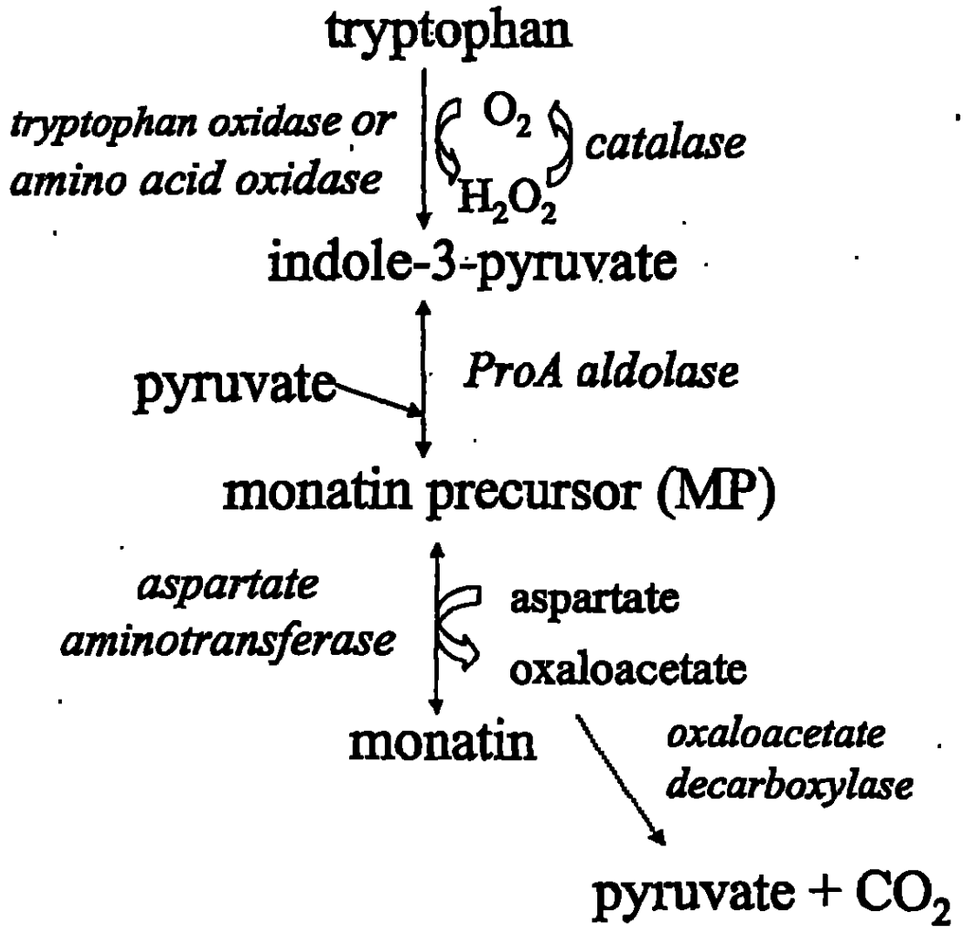

tryptophan
tryptophan oxidase or
amino acid oxidase
O2
H2O2
catalase
indole-3-pyruvate
pyruvate
ProA aldolase
monatin precursor (MP)
aspartate
aminotransferase
aspartate
oxaloacetate
monatin
oxaloacetate
decarboxylase
pyruvate + CO2

FIG. 12

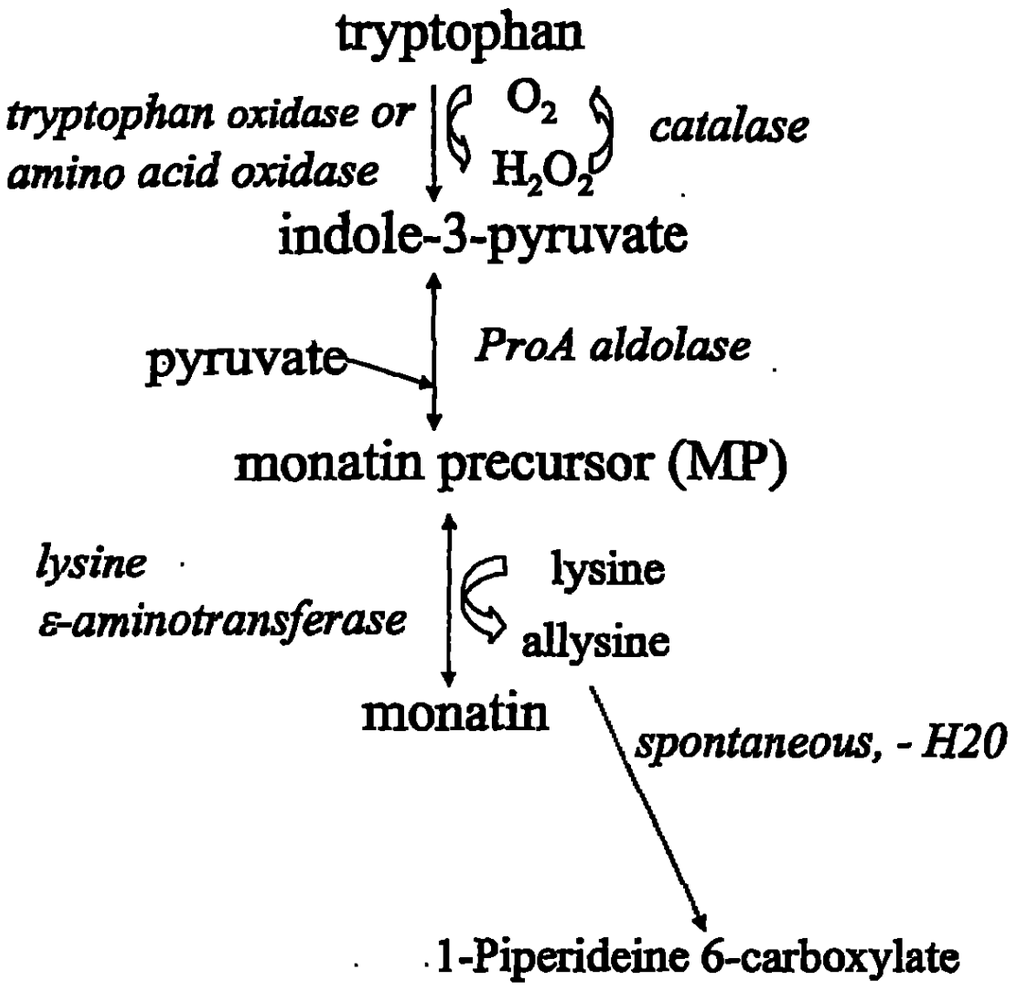
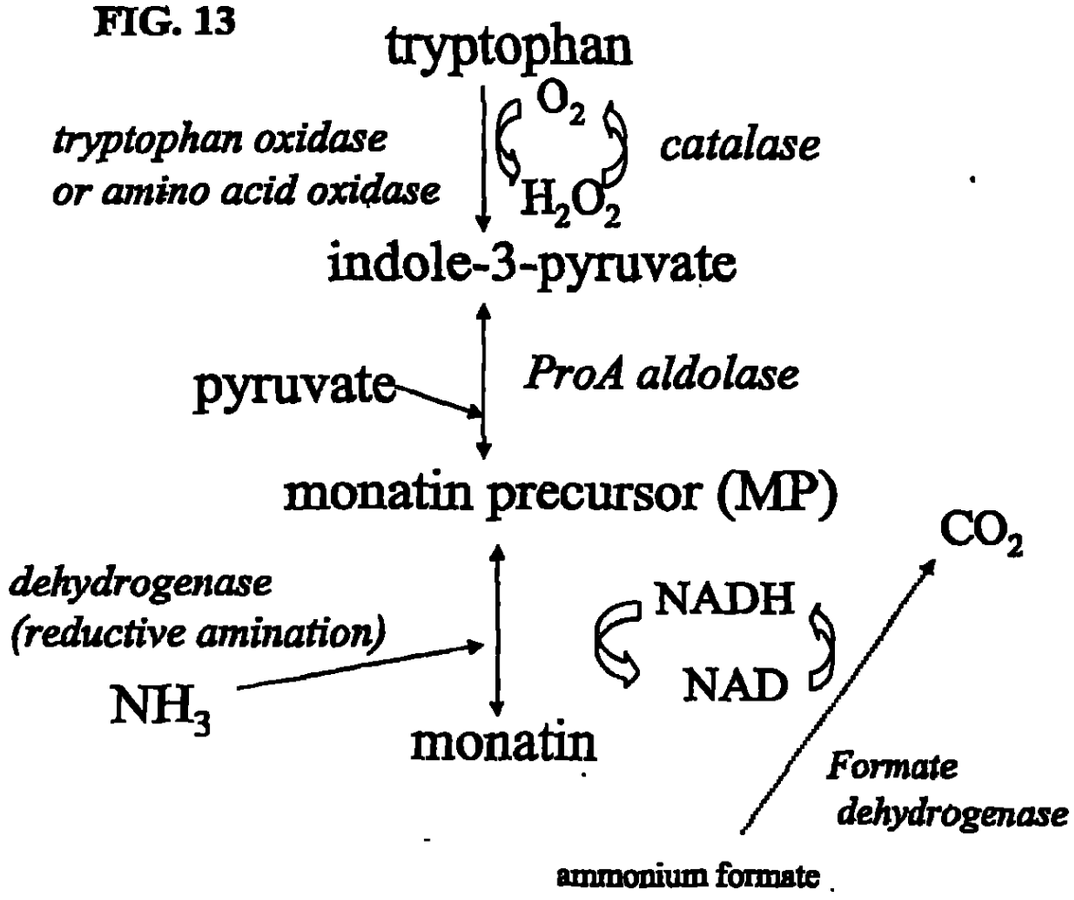

tryptophan
tryptophan oxidase or
amino acid oxidase
O2
H2O2
catalase
indole-3-pyruvate
pyruvate
ProA aldolase
monatin precursor (MP)
lysine
ε-aminotransferase
lysine
allysine
monatin
spontaneous, - H20
1-Piperideine 6-carboxylate

FIG. 13


tryptophan
O2
tryptophan oxidase
or amino acid oxidase
catalase
H2O2
indole-3-pyruvate
pyruvate
ProA aldolase
monatin precursor (MP)
CO2
dehydrogenase
(reductive amination)
NADH
NAD
NH3
monatin
Formate
dehydrogenase
ammonium formate

SEQUENCE LISTING

<110> Cargill, Incorporated

<120> Polypeptides and Biosynthetic Pathways

<130> 6682-64349

<150> 60/374,831
<151> 2002-04-23

<160> 74

<170> PatentIn version 3.1

<210> 1
<211> 1170
<212> DNA
<213> Sinorhizobium meliloti

<400> 1
atgttcgacg ccctcgcccg ccaagccgac gatcccttgc ttttcctgat cggcctgttc 60

aggaaggatg agcgccccgg aaaggtcgat ctcggcgtag gagtctatcg cgacgagacc 120

ggacgcacgc cgatcttccg ggccgtcaag gcggcggaaa agcggcttct cgaaacacag 180

gacagcaagg cctatatcgg ccccgaaggg gacctcgtct ttctcgatcg gctctgggaa 240

ctcgtcggcg gcgacacgat cgagcggagc catgttgcgg gcgtccagac gcccggcggc 300

tccggcgcgc tccgtttggc ggcggacctc atcgcccgca tgggcggccg aggcatctgg 360

ctcgggctgc cgagctggcc gaaccacgcg ccgatcttca aggcggccgg gctcgatatc 420

gccacctacg acttcttcga cattccgtcg cagtcggtca tcttcgataa tctggtgagc 480

gcgctggaag gcgccgcatc cggcgatgcg gtgctgctgc atgcaagctg ccacaacccg 540

accggcggcg tcctgagcga agcacaatgg atggagatcg ccgcgctggt ggccgagcgc 600

ggcctgctgc cgctcgtcga tctcgcctat caggggttcg gccgcggcct cgaccaggat 660

gtcgcgggcc tccggcatct tctcggcgtg gtcccggaag cgctcgtcgc ggtttcctgc 720

tcgaagtcct tcgggcttta tcgcgagcgc gcgggcgcga tcttcgcgcg gaccagctcg 780

actgcctcgg cggacagggt gcgctcaaac ctcgcgggcc tcgcacgcac cagctattcc 840

atgccgccgg atcacggcgc agccgtcgtg cggacgatcc ttgacgaccc ggaactcagg 900

cgcgactgga cggaggagct cgagacgatg cggctcagga tgacgggcct ccggcggtcg 960

cttgccgagg gactccgcac ccgctggcag agcctcggcg cagtcgccga tcaggagggc 1020

atgttctcca tgctgccgct ttccgaagcg gaggttatgc ggctcaggac cgagcacggc 1080

atctatatgc cggcatccgg ccgcatcaac atcgccgggc tgaagacggc ggaagccgcc 1140

gagattgccg gcaagttcac cagtctctga 1170

<210> 2
<211> 389
<212> PRT
<213> Sinorhizobium meliloti

<400> 2

Met Phe Asp Ala Leu Ala Arg Gln Ala Asp Asp Pro Leu Leu Phe Leu
1 5 10 15

Ile Gly Leu Phe Arg Lys Asp Glu Arg Pro Gly Lys Val Asp Leu Gly
20 25 30

Val Gly Val Tyr Arg Asp Glu Thr Gly Arg Thr Pro Ile Phe Arg Ala
35 40 45

Val Lys Ala Ala Glu Lys Arg Leu Leu Glu Thr Gln Asp Ser Lys Ala
50 55 60

Tyr Ile Gly Pro Glu Gly Asp Leu Val Phe Leu Asp Arg Leu Trp Glu
65 70 75 80

Leu Val Gly Gly Asp Thr Ile Glu Arg Ser His Val Ala Gly Val Gln
85 90 95

Thr Pro Gly Gly Ser Gly Ala Leu Arg Leu Ala Ala Asp Leu Ile Ala
100 105 110

Arg Met Gly Gly Arg Gly Ile Trp Leu Gly Leu Pro Ser Trp Pro Asn
115 120 125

His Ala Pro Ile Phe Lys Ala Ala Gly Leu Asp Ile Ala Thr Tyr Asp
130 135 140

Phe Phe Asp Ile Pro Ser Gln Ser Val Ile Phe Asp Asn Leu Val Ser
145 150 155 160

Ala Leu Glu Gly Ala Ala Ser Gly Asp Ala Val Leu Leu His Ala Ser
165 170 175

Cys His Asn Pro Thr Gly Gly Val Leu Ser Glu Ala Gln Trp Met Glu
180 185 190

Ile Ala Ala Leu Val Ala Glu Arg Gly Leu Leu Pro Leu Val Asp Leu
195 200 205

Ala Tyr Gln Gly Phe Gly Arg Gly Leu Asp Gln Asp Val Ala Gly Leu

210 215 220

Arg His Leu Leu Gly Val Val Pro Glu Ala Leu Val Ala Val Ser Cys
225 230 235 240

Ser Lys Ser Phe Gly Leu Tyr Arg Glu Arg Ala Gly Ala Ile Phe Ala
245 250 255

Arg Thr Ser Ser Thr Ala Ser Ala Asp Arg Val Arg Ser Asn Leu Ala
260 265 270

Gly Leu Ala Arg Thr Ser Tyr Ser Met Pro Pro Asp His Gly Ala Ala
275 280 285

Val Val Arg Thr Ile Leu Asp Asp Pro Glu Leu Arg Arg Asp Trp Thr
290 295 300

Glu Glu Leu Glu Thr Met Arg Leu Arg Met Thr Gly Leu Arg Arg Ser
305 310 315 320

Leu Ala Glu Gly Leu Arg Thr Arg Trp Gln Ser Leu Gly Ala Val Ala
325 330 335

Asp Gln Glu Gly Met Phe Ser Met Leu Pro Leu Ser Glu Ala Glu Val
340 345 350

Met Arg Leu Arg Thr Glu His Gly Ile Tyr Met Pro Ala Ser Gly Arg
355 360 365

Ile Asn Ile Ala Gly Leu Lys Thr Ala Glu Ala Ala Glu Ile Ala Gly
370 375 380

Lys Phe Thr Ser Leu
385

<210> 3
<211> 1260
<212> DNA
<213> Rhodobacter sphaeroides

<400> 3
atgcgctcta cgacggctcc tggtccgagt ggggcatgta tgacgatctc aaggtcgcga 60

aaggatgacg aaggaatgct gaccgccctg aagccgcagc ccgcggacaa gatcctgcaa 120

ctgatccaga tgttccgcga ggatgcgcgc gcggacaaga tcgatctggg cgtgggcgtc 180

tacaaggacc cgaccgggct caccccggtc atgcgggccg tgaaggcggc cgagaagcgg 240

ctctgggagg tcgagaccac caagacctac accggccttg ccgacgagcc ggcctacaat 300

gccgcgatgg cgaagctgat cctcgcgggc gcggtcccgg ccgaccgggt ggcctcggtc 360

gccacccccg gcggcacggg cgcggtgcgt caggcgctcg agctgatccg catggcctcg 420

cccgaggcca ccgtctggat ctcgaacccg acctggccga accatctgtc gatcgtgaaa 480

tatctcggca tcccgatgcg ggaataccgc tatttcgacg ccgagaccgg cgccgtcgat 540

gccgagggca tgatggagga tctggcccag gtgaaggcgg gcgacgtggt gctgctgcac 600

ggctgctgcc acaacccgac cggcgccaac ccgaacccgg tgcagtggct ggccatctgc 660

gagagcctgg cccggacagg cgcggtgccg ctgatcgacc tcgcctatca gggcttcggc 720

gacgggctcg agatggatgc ggcggcgacg cggcttctgg ccaccagact gcccgaggtg 780

ctgatcgcgg cctcctgctc gaagaacttc ggcatctacc gcgagcgcac gggcatcctg 840

atcgccatcg gcgaggcggc gggccggggc acggtgcagg ccaacctcaa cttcctgaac 900

cggcagaact actccttccc gccggaccat ggcgcgcggc tcgtgaccat gatcctcgag 960

gacgagacgc tgagcgccga ctggaaggcg gaactcgagg aggtgcggct caacatgctg 1020

acactgcgcc gccagcttgc cgatgcgctg caggccgaga ccggctcgaa ccgcttcggc 1080

ttcgtggccg agcatcgcgg catgttctcg cgcctcggga tcacgcccgc cgaggtggag 1140

cggctgcgga ccgagcacgg ggtctacatg gtgggcgatt cgcggctgaa catcgcgggg 1200

ctgaaccgga cgaccgtgcc ggtgctggcg cgcgcggtgg ccaaggtgct gcgcggctga 1260

<210> 4
<211> 419
<212> PRT
<213> Rhodobacter sphaeroides

<400> 4

Met Arg Ser Thr Thr Ala Pro Gly Pro Ser Gly Ala Cys Met Thr Ile
1 5 10 15

Ser Arg Ser Arg Lys Asp Asp Glu Gly Met Leu Thr Ala Leu Lys Pro
20 25 30

Gln Pro Ala Asp Lys Ile Leu Gln Leu Ile Gln Met Phe Arg Glu Asp
35 40 45

Ala Arg Ala Asp Lys Ile Asp Leu Gly Val Gly Val Tyr Lys Asp Pro
50 55 60

Thr Gly Leu Thr Pro Val Met Arg Ala Val Lys Ala Ala Glu Lys Arg
65 70 75 80

```
Leu Trp Glu Val Glu Thr Thr Lys Thr Tyr Thr Gly Leu Ala Asp Glu
                85                  90                  95

Pro Ala Tyr Asn Ala Ala Met Ala Lys Leu Ile Leu Ala Gly Ala Val
            100                 105                 110

Pro Ala Asp Arg Val Ala Ser Val Ala Thr Pro Gly Gly Thr Gly Ala
        115                 120                 125

Val Arg Gln Ala Leu Glu Leu Ile Arg Met Ala Ser Pro Glu Ala Thr
    130                 135                 140

Val Trp Ile Ser Asn Pro Thr Trp Pro Asn His Leu Ser Ile Val Lys
145                 150                 155                 160

Tyr Leu Gly Ile Pro Met Arg Glu Tyr Arg Tyr Phe Asp Ala Glu Thr
                165                 170                 175

Gly Ala Val Asp Ala Glu Gly Met Met Glu Asp Leu Ala Gln Val Lys
            180                 185                 190

Ala Gly Asp Val Val Leu Leu His Gly Cys Cys His Asn Pro Thr Gly
        195                 200                 205

Ala Asn Pro Asn Pro Val Gln Trp Leu Ala Ile Cys Glu Ser Leu Ala
    210                 215                 220

Arg Thr Gly Ala Val Pro Leu Ile Asp Leu Ala Tyr Gln Gly Phe Gly
225                 230                 235                 240

Asp Gly Leu Glu Met Asp Ala Ala Ala Thr Arg Leu Leu Ala Thr Arg
                245                 250                 255

Leu Pro Glu Val Leu Ile Ala Ala Ser Cys Ser Lys Asn Phe Gly Ile
            260                 265                 270

Tyr Arg Glu Arg Thr Gly Ile Leu Ile Ala Ile Gly Glu Ala Ala Gly
        275                 280                 285

Arg Gly Thr Val Gln Ala Asn Leu Asn Phe Leu Asn Arg Gln Asn Tyr
    290                 295                 300

Ser Phe Pro Pro Asp His Gly Ala Arg Leu Val Thr Met Ile Leu Glu
305                 310                 315                 320
```

98

Asp Glu Thr Leu Ser Ala Asp Trp Lys Ala Glu Leu Glu Glu Val Arg
325 330 335

Leu Asn Met Leu Thr Leu Arg Arg Gln Leu Ala Asp Ala Leu Gln Ala
340 345 350

Glu Thr Gly Ser Asn Arg Phe Gly Phe Val Ala Glu His Arg Gly Met
355 360 365

Phe Ser Arg Leu Gly Ile Thr Pro Ala Glu Val Glu Arg Leu Arg Thr
370 375 380

Glu His Gly Val Tyr Met Val Gly Asp Ser Arg Leu Asn Ile Ala Gly
385 390 395 400

Leu Asn Arg Thr Thr Val Pro Val Leu Ala Arg Ala Val Ala Lys Val
405 410 415

Leu Arg Gly

<210> 5
<211> 1260
<212> DNA
<213> Rhodobacter sphaeroides

<400> 5
atgcgctcta cgacggctcc tggtccgagt ggggcatgta tgacgatctc aaggtcgcga 60

aaggatgacg aaggaatgct gaccgccctg aagccgcagc ccgcggacaa gatcctgcaa 120

ctgatccaga tgttccgcga ggatgcgcgc gcggacaaga tcgatctggg cgtgggcgtc 180

tacaaggacc cgaccgggct caccccggtc atgcgggccg tgaaggccgc cgagaagcgg 240

ctctgggagg tcgagaccac caagacctac accggccttg ccggcgagcc cgcctacaat 300

gccgcgatgg cgaagctgat cctcgcaggc gcggtccccgg ccgaccgggt ggcctcggtc 360

gccacccccg gcggcacggg cgcggtgcgt caggcgctcg agctgatccg catggcctcg 420

cccgaggcca ctgtctggat ctcgaacccg acctggccga accatctgtc gatcgtgaaa 480

tatctcggca tccccgatgcg ggaataccgc tatttcgacg ccgagaccgg cgccgtcgat 540

gccgagggct tgatggagga tctggcccag gtgaaggcgg gcgacgtggt gctgctgcac 600

ggctgctgcc acaacccgac cggcgccaac ccgaacccgg tgcagtggct ggccgtctgc 660

gagagcctgg cccggacagg cgcggtgccg ctgatcgacc tcgcctatca gggcttcggc 720

gacgggctcg agatggatgc ggcggcgacg cggcttctgg ccaccagact gccccgaggtg 780

ctgatcgcgg cctcctgctc gaagaacttc ggcatctacc gcgagcgaac gggcatcctg 840

atcgccatcg gcgaggcggc gggccggggc acggtgcagg ccaacctcaa cttcctgaac 900

cggcagaact actccttccc gccggaccat ggcgcgcggc tcgtgaccat gatcctcgag 960

gacgagacgc tgagcgccga ctggaaggcg gaactcgagg aggtgcggct caacatgctg 1020

acgctgcgcc gccagcttgc cgatgcgctg caggccgaga ccggctcgaa ccgcttcggc 1080

ttcgtggccg agcatcgcgg catgttctcg cgcctcggga tcacgcccgc cgaggtggag 1140

cggctgcgga ccgagcacgg ggtctacatg gtgggcgatt cgcggctgaa catcgcgggg 1200

ctgaaccgga cgaccgtgcc ggtgctggcg cgcgcggtgg ccaaggtgct gcgcggctga 1260

<210> 6
<211> 419
<212> PRT
<213> Rhodobacter sphaeroides

<400> 6

Met Arg Ser Thr Thr Ala Pro Gly Pro Ser Gly Ala Cys Met Thr Ile
1 5 10 15

Ser Arg Ser Arg Lys Asp Asp Glu Gly Met Leu Thr Ala Leu Lys Pro
20 25 30

Gln Pro Ala Asp Lys Ile Leu Gln Leu Ile Gln Met Phe Arg Glu Asp
35 40 45

Ala Arg Ala Asp Lys Ile Asp Leu Gly Val Gly Val Tyr Lys Asp Pro
50 55 60

Thr Gly Leu Thr Pro Val Met Arg Ala Val Lys Ala Ala Glu Lys Arg
65 70 75 80

Leu Trp Glu Val Glu Thr Thr Lys Thr Tyr Thr Gly Leu Ala Gly Glu
85 90 95

Pro Ala Tyr Asn Ala Ala Met Ala Lys Leu Ile Leu Ala Gly Ala Val
100 105 110

Pro Ala Asp Arg Val Ala Ser Val Ala Thr Pro Gly Gly Thr Gly Ala
115 120 125

Val Arg Gln Ala Leu Glu Leu Ile Arg Met Ala Ser Pro Glu Ala Thr
130 135 140

Val Trp Ile Ser Asn Pro Thr Trp Pro Asn His Leu Ser Ile Val Lys
145 150 155 160

90

Tyr Leu Gly Ile Pro Met Arg Glu Tyr Arg Tyr Phe Asp Ala Glu Thr
165 170 175

Gly Ala Val Asp Ala Glu Gly Leu Met Glu Asp Leu Ala Gln Val Lys
180 185 190

Ala Gly Asp Val Val Leu Leu His Gly Cys Cys His Asn Pro Thr Gly
195 200 205

Ala Asn Pro Asn Pro Val Gln Trp Leu Ala Val Cys Glu Ser Leu Ala
210 215 220

Arg Thr Gly Ala Val Pro Leu Ile Asp Leu Ala Tyr Gln Gly Phe Gly
225 230 235 240

Asp Gly Leu Glu Met Asp Ala Ala Ala Thr Arg Leu Leu Ala Thr Arg
245 250 255

Leu Pro Glu Val Leu Ile Ala Ala Ser Cys Ser Lys Asn Phe Gly Ile
260 265 270

Tyr Arg Glu Arg Thr Gly Ile Leu Ile Ala Ile Gly Glu Ala Ala Gly
275 280 285

Arg Gly Thr Val Gln Ala Asn Leu Asn Phe Leu Asn Arg Gln Asn Tyr
290 295 300

Ser Phe Pro Pro Asp His Gly Ala Arg Leu Val Thr Met Ile Leu Glu
305 310 315 320

Asp Glu Thr Leu Ser Ala Asp Trp Lys Ala Glu Leu Glu Glu Val Arg
325 330 335

Leu Asn Met Leu Thr Leu Arg Arg Gln Leu Ala Asp Ala Leu Gln Ala
340 345 350

Glu Thr Gly Ser Asn Arg Phe Gly Phe Val Ala Glu His Arg Gly Met
355 360 365

Phe Ser Arg Leu Gly Ile Thr Pro Ala Glu Val Glu Arg Leu Arg Thr
370 375 380

Glu His Gly Val Tyr Met Val Gly Asp Ser Arg Leu Asn Ile Ala Gly
385 390 395 400

Leu Asn Arg Thr Thr Val Pro Val Leu Ala Arg Ala Val Ala Lys Val
405 410 415

Leu Arg Gly

<210> 7
<211> 1239
<212> DNA
<213> Leishmania major

<400> 7
atgtccatgc aggcggccat gaccacggcg gagcgctggc agaagattca ggcaoaagct 60

cccgatgtca tcttcgatct cgcaaaacgc gccgccgctg ccaagggccc caaggccaac 120

ctcgtcattg gtgcctaccg cgacgagcag ggccgtccct atccgotacg cgtggtccgc 180

aaggctgagc agcttctctt ggacatgaat ctcgactacg agtacctacc catctcgggc 240

taccagccct tcatcgatga ggcggtaaag attatctacg gcaataccgt cgagctggag 300

aacctggttg cggtgcagac gctgagcggg accggtgctg tctctctcgg ggcgaagctg 360

ctgactcgcg tcttcgacgc tgagacgacg cccatctacc tttccgaccc cacgtggccc 420

aaccactacg gcgtcgtgaa ggctgctggc tggaagaaca tctgcacgta cgcctactac 480

gaccccaaga cggtcagcct gaatttcgag ggcatgaaga aagacattct ggcggcgccg 540

gacggctccg tgttcattct gcaccagtgc gcgcacaacc ccaccggcgt ggacccgtcg 600

caggagcagt ggaacgagat cgcgtcactg atgctggcca agcaccatca ggtgttcttc 660

gactccgcct accaaggcta tgcgagcggc agcctcgaca cggacgcgta tgctgcccgc 720

ctgtttgccc gccgcggcat cgaggtactg ctggcgcagt cgttctccaa gaacatgggc 780

ttgtacagcg agcgtgcagg cacgctgtcg ctgctcctca aggacaagac gaagcgcgcg 840

gatgtaaaga gcgtgatgga ttcgctgatc cgtgaggagt acacgtgccc cccagcccac 900

ggtgcccgct tagcccacct aatcctgagc aacaacgaac tgcgaaagga gtgggaggca 960

gagctatcag ccatggcaga gcgcatccgt acgatgcgcc gcaccgtgta cgacgagctg 1020

ctgcgcctgc agacgcccgg gagctgggaa catgtcatta accagattgg catgttttcc 1080

ttcctcgggc tgtcaaaggc gcagtgcgaa tactgccaaa accacaacat cttcatcaca 1140

gtgtcgggcc gcgctaacat ggcaggtctg acgcatgaga cggcgctgat gctagcacag 1200

acgatcaacg atgctgtgcg caatgtgaat cgtgagtga 1239

<210> 8
<211> 412
<212> PRT

<213> Leishmania major

<400> 8

```
Met Ser Met Gln Ala Ala Met Thr Thr Ala Glu Arg Trp Gln Lys Ile
1               5                  10                  15

Gln Ala Gln Ala Pro Asp Val Ile Phe Asp Leu Ala Lys Arg Ala Ala
            20                  25                  30

Ala Ala Lys Gly Pro Lys Ala Asn Leu Val Ile Gly Ala Tyr Arg Asp
        35                  40                  45

Glu Gln Gly Arg Pro Tyr Pro Leu Arg Val Val Arg Lys Ala Glu Gln
    50                  55                  60

Leu Leu Leu Asp Met Asn Leu Asp Tyr Glu Tyr Leu Pro Ile Ser Gly
65                  70                  75                  80

Tyr Gln Pro Phe Ile Asp Glu Ala Val Lys Ile Ile Tyr Gly Asn Thr
                85                  90                  95

Val Glu Leu Glu Asn Leu Val Ala Val Gln Thr Leu Ser Gly Thr Gly
            100                 105                 110

Ala Val Ser Leu Gly Ala Lys Leu Leu Thr Arg Val Phe Asp Ala Glu
        115                 120                 125

Thr Thr Pro Ile Tyr Leu Ser Asp Pro Thr Trp Pro Asn His Tyr Gly
    130                 135                 140

Val Val Lys Ala Ala Gly Trp Lys Asn Ile Cys Thr Tyr Ala Tyr Tyr
145                 150                 155                 160

Asp Pro Lys Thr Val Ser Leu Asn Phe Glu Gly Met Lys Lys Asp Ile
                165                 170                 175

Leu Ala Ala Pro Asp Gly Ser Val Phe Ile Leu His Gln Cys Ala His
            180                 185                 190

Asn Pro Thr Gly Val Asp Pro Ser Gln Glu Gln Trp Asn Glu Ile Ala
        195                 200                 205

Ser Leu Met Leu Ala Lys His His Gln Val Phe Phe Asp Ser Ala Tyr
    210                 215                 220

Gln Gly Tyr Ala Ser Gly Ser Leu Asp Thr Asp Ala Tyr Ala Ala Arg
```

93

225 230 235 240

Leu Phe Ala Arg Arg Gly Ile Glu Val Leu Leu Ala Gln Ser Phe Ser
245 250 255

Lys Asn Met Gly Leu Tyr Ser Glu Arg Ala Gly Thr Leu Ser Leu Leu
260 265 270

Leu Lys Asp Lys Thr Lys Arg Ala Asp Val Lys Ser Val Met Asp Ser
275 280 285

Leu Ile Arg Glu Glu Tyr Thr Cys Pro Pro Ala His Gly Ala Arg Leu
290 295 300

Ala His Leu Ile Leu Ser Asn Asn Glu Leu Arg Lys Glu Trp Glu Ala
305 310 315 320

Glu Leu Ser Ala Met Ala Glu Arg Ile Arg Thr Met Arg Arg Thr Val
325 330 335

Tyr Asp Glu Leu Leu Arg Leu Gln Thr Pro Gly Ser Trp Glu His Val
340 345 350

Ile Asn Gln Ile Gly Met Phe Ser Phe Leu Gly Leu Ser Lys Ala Gln
355 360 365

Cys Glu Tyr Cys Gln Asn His Asn Ile Phe Ile Thr Val Ser Gly Arg
370 375 380

Ala Asn Met Ala Gly Leu Thr His Glu Thr Ala Leu Met Leu Ala Gln
385 390 395 400

Thr Ile Asn Asp Ala Val Arg Asn Val Asn Arg Glu
405 410

<210> 9
<211> 1182
<212> DNA
<213> Bacillus subtilis

<400> 9
atggaacatt tgctgaatcc gaaagcaaga gaggtcgaaa tttcaggaat acgcaaattc 60

tcgaatcttc tcgccgaacg cgaagacgtc atttcactta caatcggcca gcctgatttt 120

ttcacaccgc atcatgtgaa agctgccgca aaaaaagcca ttgatgaaaa cgtgacgtca 180

tatactccga atgccggcta cctggagctg agacaagctg tgcagcttta tatgaagaaa 240

aaagcggatt tcaactatga tgctgaatct gaaattatca tcacaacagg cgcaagccaa 300

gccattgatg ctgcattccg gacgatttta tctcccggtg atgaagtcat tatgccaggg 360

cctatttatc cgggctatga acctattatc aatttgtgcg gggccaagcc tgtcattgtt 420

gatactacgt cacacggctt taagcttacc gcccggctga ttgaagatgc tctgacaccc 480

aacaccaagt gtgtcgtgct tccttatccg tcaaacccta ccggcgtgac tttatctgaa 540

gaagaactga aaagcatcgc agctctctta aaaggcagaa atgtcttcgt attgtctgat 600

gaaatataca gtgaattaac atatgacaga ccgcattact ccatcgcaac ctatttgcgg 660

gatcaaacga ttgtcattaa cgggttgtca aaatcacaca gcatgaccgg ttggagaatt 720

ggatttttat ttgcaccgaa agacattgca aagcacattt taaaggttca tcaatacaat 780

gtgtcgtgcg cctcatccat ttctcaaaaa gccgcgcttg aagctgtcac aaacggcttt 840

gacgatgcat tgattatgag agaacaatac aaaaaacgtc tggactatgt ttatgaccgt 900

cttgtttcca tgggacttga cgtagttaaa ccgtccggtg cgttttatat cttcccttct 960

attaaatcat ttggaatgac ttcatttgat tttagtatgg ctcttttgga agacgctggc 1020

gtggcactcg tgccgggcag ctcgttctca acatatggtg aaggatatgt aaggctgtct 1080

tttgcatgct caatggacac gctgagagaa ggcctagacc gtttagaatt atttgtatta 1140

aaaaaacgtg aagcaatgca gacgataaac aacggcgttt aa 1182

<210> 10
<211> 393
<212> PRT
<213> Bacillus subtilis

<400> 10

Met Glu His Leu Leu Asn Pro Lys Ala Arg Glu Ile Glu Ile Ser Gly
1 5 10 15

Ile Arg Lys Phe Ser Asn Leu Val Ala Gln His Glu Asp Val Ile Ser
20 25 30

Leu Thr Ile Gly Gln Pro Asp Phe Phe Thr Pro His His Val Lys Ala
35 40 45

Ala Ala Lys Lys Ala Ile Asp Glu Asn Val Thr Ser Tyr Thr Pro Asn
50 55 60

Ala Gly Tyr Leu Glu Leu Arg Gln Ala Val Gln Leu Tyr Met Lys Lys
65 70 75 80

Lys Ala Asp Phe Asn Tyr Asp Ala Glu Ser Glu Ile Ile Ile Thr Thr

85 90 95

Gly Ala Ser Gln Ala Ile Asp Ala Ala Phe Arg Thr Ile Leu Ser Pro
100 105 110

Gly Asp Glu Val Ile Met Pro Gly Pro Ile Tyr Pro Gly Tyr Glu Pro
115 120 125

Ile Ile Asn Leu Cys Gly Ala Lys Pro Val Ile Val Asp Thr Thr Ser
130 135 140

His Gly Phe Lys Leu Thr Ala Arg Leu Ile Glu Asp Ala Leu Thr Pro
145 150 155 160

Asn Thr Lys Cys Val Val Leu Pro Tyr Pro Ser Asn Pro Thr Gly Val
165 170 175

Thr Leu Ser Glu Glu Glu Leu Lys Ser Ile Ala Ala Leu Leu Lys Gly
180 185 190

Arg Asn Val Phe Val Leu Ser Asp Glu Ile Tyr Ser Glu Leu Thr Tyr
195 200 205

Asp Arg Pro His Tyr Ser Ile Ala Thr Tyr Leu Arg Asp Gln Thr Ile
210 215 220

Val Ile Asn Gly Leu Ser Lys Ser His Ser Met Thr Gly Trp Arg Ile
225 230 235 240

Gly Phe Leu Phe Ala Pro Lys Asp Ile Ala Lys His Ile Leu Lys Val
245 250 255

His Gln Tyr Asn Val Ser Cys Ala Ser Ser Ile Ser Gln Lys Ala Ala
260 265 270

Leu Glu Ala Val Thr Asn Gly Phe Asp Asp Ala Leu Ile Met Arg Glu
275 280 285

Gln Tyr Lys Lys Arg Leu Asp Tyr Val Tyr Asp Arg Leu Val Ser Met
290 295 300

Gly Leu Asp Val Val Lys Pro Ser Gly Ala Phe Tyr Ile Phe Pro Ser
305 310 315 320

Ile Lys Ser Phe Gly Met Thr Ser Phe Asp Phe Ser Met Ala Leu Leu
325 330 335

96

Glu Asp Ala Gly Val Ala Leu Val Pro Gly Ser Ser Phe Ser Thr Tyr
340 345 350

Gly Glu Gly Tyr Val Arg Leu Ser Phe Ala Cys Ser Met Asp Thr Leu
355 360 365

Arg Glu Gly Leu Asp Arg Leu Glu Leu Phe Val Leu Lys Lys Arg Glu
370 375 380

Ala Met Gln Thr Ile Asn Asn Gly Val
385 390

<210> 11
<211> 1176
<212> DNA
<213> Lactobacillus amylovorus

<400> 11
atgccagaat tagctaatga tttaggatta agcaaaaaga tcactgatgt aaaagcttca 60

ggaattagaa tctttgataa caaagtttca gctattcctg gcattatcaa attgactttg 120

ggtgaaccag atatgaatac tcctgagcat gttaagcaag cggctattaa gaatattgca 180

gataatgatt cacactatgc tccacaaaag ggaaagcttg aattaagaaa agctatcagt 240

aaatatttga aaaagattac tggaattgaa tatgatccag aaacagaaat cgtagtaaca 300

gttggtgcaa ctgaagcaat taacgctacc ttgtttgcta ttactaatcc gggtgacaag 360

gttgcaattc ctacgccagt cttttctcta tattggcccg tggctacact tgctgatgcc 420

gattatgttt tgatgaatac tgcagaagat ggttttaagt taacacctaa gaagttagaa 480

gaaactatca aagaaaatcc aacaattaaa gcagtaattt tgaattatcc aactaaccca 540

actggtgttg aatatagcga agatgaaatt aaagctttgg ctaaggtaat taaagataat 600

catctgtacg taattaccga tgaaatttac agtactttga cttacggtgt aaaacacttt 660

tcaattgcca gcttaattcc agaaagagca atttatatct ctggtttatc taaatcacat 720

gcgatgactg gttatcgttt aggctatgtt gccggacctg caaaaattat ggcagaaatt 780

ggtaaagttc atggccttat ggtgacgact acgacggatt catcacaagc tgccgcaatt 840

gaagcacttg aacacggact tgatgaccct gagaaatata gggaagttta tgaaaagcgt 900

cgtgactatg ttttaaagga attagccgag atagagatgc aagcagttaa gccagaaggt 960

gcattttata tctttgctaa aattccagct aagtatggca aagacgatat gaaatttgcc 1020

ttggatttag cttttaaaga aaaagtgggt atcactccag gtagtgcatt tggtcctggt 1080

ggtgaaggtc atattagatt atcttatgca tcaagtgatg aaaacttgca tgaggcaatg 1140

aagcgaatga agaaagtttt acaagaggac gaataa 1176

<210> 12
<211> 391
<212> PRT
<213> Lactobacillus amylovorus

<400> 12

Met Pro Glu Leu Ala Asn Asp Leu Gly Leu Ser Lys Lys Ile Thr Asp
1 5 10 15

Val Lys Ala Ser Gly Ile Arg Ile Phe Asp Asn Lys Val Ser Ala Ile
20 25 30

Pro Gly Ile Ile Lys Leu Thr Leu Gly Glu Pro Asp Met Asn Thr Pro
35 40 45

Glu His Val Lys Gln Ala Ala Ile Lys Asn Ile Ala Asp Asn Asp Ser
50 55 60

His Tyr Ala Pro Gln Lys Gly Lys Leu Glu Leu Arg Lys Ala Ile Ser
65 70 75 80

Lys Tyr Leu Lys Lys Ile Thr Gly Ile Glu Tyr Asp Pro Glu Thr Glu
85 90 95

Ile Val Val Thr Val Gly Ala Thr Glu Ala Ile Asn Ala Thr Leu Phe
100 105 110

Ala Ile Thr Asn Pro Gly Asp Lys Val Ala Ile Pro Thr Pro Val Phe
115 120 125

Ser Leu Tyr Trp Pro Val Ala Thr Leu Ala Asp Ala Asp Tyr Val Leu
130 135 140

Met Asn Thr Ala Glu Asp Gly Phe Lys Leu Thr Pro Lys Lys Leu Glu
145 150 155 160

Glu Thr Ile Lys Glu Asn Pro Thr Ile Lys Ala Val Ile Leu Asn Tyr
165 170 175

Pro Thr Asn Pro Thr Gly Val Glu Tyr Ser Glu Asp Glu Ile Lys Ala
180 185 190

Leu Ala Lys Val Ile Lys Asp Asn His Leu Tyr Val Ile Thr Asp Glu
195 200 205

Ile Tyr Ser Thr Leu Thr Tyr Gly Val Lys His Phe Ser Ile Ala Ser
210 215 220

Leu Ile Pro Glu Arg Ala Ile Tyr Ile Ser Gly Leu Ser Lys Ser His
225 230 235 240

Ala Met Thr Gly Tyr Arg Leu Gly Tyr Val Ala Gly Pro Ala Lys Ile
245 250 255

Met Ala Glu Ile Gly Lys Val His Gly Leu Met Val Thr Thr Thr Thr
260 265 270

Asp Ser Ser Gln Ala Ala Ala Ile Glu Ala Leu Glu His Gly Leu Asp
275 280 285

Asp Pro Glu Lys Tyr Arg Glu Val Tyr Glu Lys Arg Arg Asp Tyr Val
290 295 300

Leu Lys Glu Leu Ala Glu Ile Glu Met Gln Ala Val Lys Pro Glu Gly
305 310 315 320

Ala Phe Tyr Ile Phe Ala Lys Ile Pro Ala Lys Tyr Gly Lys Asp Asp
325 330 335

Met Lys Phe Ala Leu Asp Leu Ala Phe Lys Glu Lys Val Gly Ile Thr
340 345 350

Pro Gly Ser Ala Phe Gly Pro Gly Gly Glu Gly His Ile Arg Leu Ser
355 360 365

Tyr Ala Ser Ser Asp Glu Asn Leu His Glu Ala Met Lys Arg Met Lys
370 375 380

Lys Val Leu Gln Glu Asp Glu
385 390

<210> 13
<211> 1413
<212> DNA
<213> R. sphaeroides

<400> 13
atgcgcgagc ctcttgccct cgagatcgac ccgggccacg gcggcccgct gttcctcgcc 60

atcgccgagg cgatcaccct cgacatcacc cgcgggcggc tgaggcccgg agcgagactg 120

cccggcacac gcgcgctggc gcgggcgctc ggcgtgcatc gcaacacggt ggatgccgcc 180

tatcaggagt tgctgaccca gggctggctg caggccgagc ccgcgcgggg caccttcgtg 240

gcgcaggatc tgccgcaggg gatgctggtg cacaggcccg cgcccgcgcc ggtcgagccg 300

gtcgcgatgc gcgcggggct cgccttctcc gatggcgcgc cggaccccga gctggtgccc 360

gacaaggcgc tggcgcgggc ctttcgccgg gcgctcctgt cgcccgcctt ccgcgccgga 420

gcggattacg gcgatgcccg cggcacctcc tcgctgcggg aggcgctggc agcctatctc 480

gcctcggacc ggggcgtggt cgcggatcct gcgcggctgc tgatcgcgcg gggcagccag 540

atggcgctgt tcctggtagc ccgggcggcg ctggcgccgg gagaggcgat cgcggtcgag 600

gagccgggct atccgctggc ctgggaggcg ttccgcgcag cgggagcgga ggtgcgcggc 660

gtgccggtgg atggcggcgg cctcaggatc gacgcgctcg aggccgcgct ggcccgggat 720

ccgcgaatcc gggcggtcta tgtcacgccc catcaccagt atccgacgac cgtcaccatg 780

ggcgcggcgc ggcggttgca gcttctggaa ctggcagagc gccaccggct cgcgctgatc 840

gaggacgact acgaccacga ataccgcttc gagggccgtc cggtgctgcc gctggctgcc 900

cgcgcgccgg aaggtctgcc gctgatctat gtgggctcgc tgtcgaaact gctctcgccc 960

ggtatccggc tgggatacgc gctggcgccc gagcggctgc tgacccgcat ggccgcggcg 1020

cgcgccgcca tcgaccggca gggcgacgcg ccgctcgagg cggcgctggc cgagctgatc 1080

cgcgacggcg atctgggccg tcatgcccgc aaggcgcgca gggtctaccg ggcgcggcgg 1140

gatctgctgg cggagcgtct cacggcgcag ctggccgggc gcgccgcctt cgatctgccg 1200

gccgggggcc tcgcgctgtg gctgcgctgc gcgggcgtct cggccgagac ctgggccgaa 1260

gccgcagggc aggcggggct cgccctgctg ccgggcacgc gcttcgcgct ggagagcccg 1320

gcgccgcagg ccttccggct gggctatgcg gcgctggacg aggggcagat cgccccgggcg 1380

gtggagatcc tcgccccgga cttccccggc tga 1413

<210> 14
<211> 470
<212> PRT
<213> R. sphaeroides

<400> 14

Met Arg Glu Pro Leu Ala Leu Glu Ile Asp Pro Gly His Gly Gly Pro
1 5 10 15

Leu Phe Leu Ala Ile Ala Glu Ala Ile Thr Leu Asp Ile Thr Arg Gly
20 25 30

Arg Leu Arg Pro Gly Ala Arg Leu Pro Gly Thr Arg Ala Leu Ala Arg
35 40 45

Ala Leu Gly Val His Arg Asn Thr Val Asp Ala Ala Tyr Gln Glu Leu
50 55 60

Leu Thr Gln Gly Trp Leu Gln Ala Glu Pro Ala Arg Gly Thr Phe Val
65 70 75 80

Ala Gln Asp Leu Pro Gln Gly Met Leu Val His Arg Pro Ala Pro Ala
85 90 95

Pro Val Glu Pro Val Ala Met Arg Ala Gly Leu Ala Phe Ser Asp Gly
100 105 110

Ala Pro Asp Pro Glu Leu Val Pro Asp Lys Ala Leu Ala Arg Ala Phe
115 120 125

Arg Arg Ala Leu Leu Ser Pro Ala Phe Arg Ala Gly Ala Asp Tyr Gly
130 135 140

Asp Ala Arg Gly Thr Ser Ser Leu Arg Glu Ala Leu Ala Ala Tyr Leu
145 150 155 160

Ala Ser Asp Arg Gly Val Val Ala Asp Pro Ala Arg Leu Leu Ile Ala
165 170 175

Arg Gly Ser Gln Met Ala Leu Phe Leu Val Ala Arg Ala Ala Leu Ala
180 185 190

Pro Gly Glu Ala Ile Ala Val Glu Glu Pro Gly Tyr Pro Leu Ala Trp
195 200 205

Glu Ala Phe Arg Ala Ala Gly Ala Glu Val Arg Gly Val Pro Val Asp
210 215 220

Gly Gly Gly Leu Arg Ile Asp Ala Leu Glu Ala Ala Leu Ala Arg Asp
225 230 235 240

Pro Arg Ile Arg Ala Val Tyr Val Thr Pro His His Gln Tyr Pro Thr
245 250 255

Thr Val Thr Met Gly Ala Ala Arg Arg Leu Gln Leu Leu Glu Leu Ala
260 265 270

Glu Arg His Arg Leu Ala Leu Ile Glu Asp Asp Tyr Asp His Glu Tyr
275 280 285

Arg Phe Glu Gly Arg Pro Val Leu Pro Leu Ala Ala Arg Ala Pro Glu
290 295 300

Gly Leu Pro Leu Ile Tyr Val Gly Ser Leu Ser Lys Leu Leu Ser Pro
305 310 315 320

Gly Ile Arg Leu Gly Tyr Ala Leu Ala Pro Glu Arg Leu Leu Thr Arg
325 330 335

Met Ala Ala Ala Arg Ala Ala Ile Asp Arg Gln Gly Asp Ala Pro Leu
340 345 350

Glu Ala Ala Leu Ala Glu Leu Ile Arg Asp Gly Asp Leu Gly Arg His
355 360 365

Ala Arg Lys Ala Arg Arg Val Tyr Arg Ala Arg Arg Asp Leu Leu Ala
370 375 380

Glu Arg Leu Thr Ala Gln Leu Ala Gly Arg Ala Ala Phe Asp Leu Pro
385 390 395 400

Ala Gly Gly Leu Ala Leu Trp Leu Arg Cys Ala Gly Val Ser Ala Glu
405 410 415

Thr Trp Ala Glu Ala Ala Gly Gln Ala Gly Leu Ala Leu Leu Pro Gly
420 425 430

Thr Arg Phe Ala Leu Glu Ser Pro Ala Pro Gln Ala Phe Arg Leu Gly
435 440 445

Tyr Ala Ala Leu Asp Glu Gly Gln Ile Ala Arg Ala Val Glu Ile Leu
450 455 460

Ala Arg Ser Phe Pro Gly
465 470

<210> 15
<211> 35
<212> DNA
<213> Artificial sequence

<220>
<223> primer

<400> 15
ggtattgagg gtcgcatgaa ggttttagtc aatgg 35

<210> 16
<211> 37
<212> DNA
<213> Artificial sequence

<220>
<223> primer

<400> 16
agaggagagt tagagcctta tgaaatgcta gcagcct 37

<210> 17
<211> 35
<212> DNA
<213> Artificial sequence

<220>
<223> primer

<400> 17
ggtattgagg gtcgcatgtt cgacgccctc gcccg 35

<210> 18
<211> 37
<212> DNA
<213> Artificial sequence

<220>
<223> primer

<400> 18
agaggagagt tagagcctca gagactggtg aacttgc 37

<210> 19
<211> 35
<212> DNA
<213> Artificial sequence

<220>
<223> primer

<400> 19
ggtattgagg gtcgcatgga acatttgctg aatcc 35

<210> 20
<211> 37
<212> DNA
<213> Artificial sequence

<220>
<223> primer

<400> 20
agaggagagt tagagcctta aacgccgttg tttatcg 37

<210> 21

<211> 35
<212> DNA
<213> Artificial sequence

<220>
<223> primer

<400> 21
ggtattgagg gtcgcatgcg cgagcctctt gccct 35

<210> 22
<211> 37
<212> DNA
<213> Artificial sequence

<220>
<223> primer

<400> 22
agaggagagt tagagcctca gccggggaag ctccggg 37

<210> 23
<211> 35
<212> DNA
<213> Artificial sequence

<220>
<223> primer

<400> 23
ggtattgagg gtcgcatgtc cacgcaggcg gccat 35

<210> 24
<211> 37
<212> DNA
<213> Artificial sequence

<220>
<223> primer

<400> 24
agaggagagt tagagcctca ctcacgattc acattgc 37

<210> 25
<211> 35
<212> DNA
<213> Artificial sequence

<220>
<223> primer

<400> 25
ggtattgagg gtcgcatgcc agaattagct aatga 35

<210> 26
<211> 37

<212> DNA
<213> Artificial sequence

<220>
<223> primer

<400> 26
agaggagagt tagagcctta ttcgtcctct tgtaaaa 37

<210> 27
<211> 35
<212> DNA
<213> Artificial sequence

<220>
<223> primer

<400> 27
ggtattgagg gtcgcatgcg ctctacgacg gctcc 35

<210> 28
<211> 37
<212> DNA
<213> Artificial sequence

<220>
<223> primer

<400> 28
agaggagagt tagagcctca gccgcgcagc accttgg 37

<210> 29
<211> 35
<212> DNA
<213> Artificial sequence

<220>
<223> primer

<400> 29
ggtattgagg gtcgcatgtt tgagaacatt accgc 35

<210> 30
<211> 37
<212> DNA
<213> Artificial sequence

<220>
<223> primer

<400> 30
agaggagagt tagagcctta cagcactgcc acaatcg 37

<210> 31
<211> 1194
<212> DNA

<213> E. coli

<400> 31
gtgtttcaaa aagttgacgc ctacgctggc gacccgattc ttacgcttat ggagcgtttt 60

aaagaagacc ctcgcagcga caaagtgaat ttaagtatcg gtctgtacta caacgaagac 120

ggaattattc cacaactgca agccgtggcg gaggcggaag cgcgcctgaa tgcgcagcct 180

catggcgctt cgctttattt accgatggaa gggcttaact gctatcgcca tgccattgcg 240

ccgctgctgt ttggtgcgga ccatccggta ctgaaacaac agcgcgtagc aaccattcaa 300

acccttggcg gctccggggc attgaaagtg ggcgcggatt tcctgaaacg ctacttcccg 360

gaatcaggcg tctgggtcag cgatcctacc tgggaaaacc gcgtagcaat attcgccggg 420

gctggattcg aagtgagtac ttacccctgg tatgacgaag cgactaacgg cgtgcgcttt 480

aatgacctgt tggcgacgct gaaaacatta cctgcccgca gtattgtgtt gctgcatcca 540

tgttgccaca acccaacggg tgccgatctc actaatgatc agtgggatgc ggtgattgaa 600

attctcaaag cccgcgagct tattccattc ctcgatattg cctatcaagg atttggtgcc 660

ggtatggaag aggatgccta cgctattcgc gccattgcca gcgctggatt acccgctctg 720

gtgagcaatt cgttctcgaa aattttctcc ctttacggcg agcgcgtcgg cggactttct 780

gttatgtgtg aagatgccga agccgctggc cgcgtactgg ggcaattgaa agcaacagtt 840

cgccgcaact actccagccc gccgaatttt ggtgcgcagg tggtggctgc agtgctgaat 900

gacgaggcat tgaaagccag ctggctggcg gaagtagaag agatgcgtac tcgcattctg 960

gcaatgcgtc aggaattggt gaaggtatta agcacagaga tgccagaacg caatttcgat 1020

tatctgctta atcagcgcgg catgttcagt tataccggtt taagtgccgc tcaggttgac 1080

cgactacgtg aagaatttgg tgtctatctc atcgccagcg gtcgcatgtg tgtcgccggg 1140

ttaaatacgg caaatgtaca acgtgtggca aaggcgtttg ctgcggtgat gtaa 1194

<210> 32
<211> 397
<212> PRT
<213> E. coli

<400> 32

Val Phe Gln Lys Val Asp Ala Tyr Ala Gly Asp Pro Ile Leu Thr Leu
1 5 10 15

Met Glu Arg Phe Lys Glu Asp Pro Arg Ser Asp Lys Val Asn Leu Ser
20 25 30

Ile Gly Leu Tyr Tyr Asn Glu Asp Gly Ile Ile Pro Gln Leu Gln Ala
35 40 45

106

Val Ala Glu Ala Glu Ala Arg Leu Asn Ala Gln Pro His Gly Ala Ser
50 55 60

Leu Tyr Leu Pro Met Glu Gly Leu Asn Cys Tyr Arg His Ala Ile Ala
65 70 75 80

Pro Leu Leu Phe Gly Ala Asp His Pro Val Leu Lys Gln Gln Arg Val
85 90 95

Ala Thr Ile Gln Thr Leu Gly Gly Ser Gly Ala Leu Lys Val Gly Ala
100 105 110

Asp Phe Leu Lys Arg Tyr Phe Pro Glu Ser Gly Val Trp Val Ser Asp
115 120 125

Pro Thr Trp Glu Asn Arg Val Ala Ile Phe Ala Gly Ala Gly Phe Glu
130 135 140

Val Ser Thr Tyr Pro Trp Tyr Asp Glu Ala Thr Asn Gly Val Arg Phe
145 150 155 160

Asn Asp Leu Leu Ala Thr Leu Lys Thr Leu Pro Ala Arg Ser Ile Val
165 170 175

Leu Leu His Pro Cys Cys His Asn Pro Thr Gly Ala Asp Leu Thr Asn
180 185 190

Asp Gln Trp Asp Ala Val Ile Glu Ile Leu Lys Ala Arg Glu Leu Ile
195 200 205

Pro Phe Leu Asp Ile Ala Tyr Gln Gly Phe Gly Ala Gly Met Glu Glu
210 215 220

Asp Ala Tyr Ala Ile Arg Ala Ile Ala Ser Ala Gly Leu Pro Ala Leu
225 230 235 240

Val Ser Asn Ser Phe Ser Lys Ile Phe Ser Leu Tyr Gly Glu Arg Val
245 250 255

Gly Gly Leu Ser Val Met Cys Glu Asp Ala Glu Ala Ala Gly Arg Val
260 265 270

Leu Gly Gln Leu Lys Ala Thr Val Arg Arg Asn Tyr Ser Ser Pro Pro
275 280 285

Asn Phe Gly Ala Gln Val Val Ala Ala Val Leu Asn Asp Glu Ala Leu
290 295 300

Lys Ala Ser Trp Leu Ala Glu Val Glu Glu Met Arg Thr Arg Ile Leu
305 310 315 320

Ala Met Arg Gln Glu Leu Val Lys Val Leu Ser Thr Glu Met Pro Glu
325 330 335

Arg Asn Phe Asp Tyr Leu Leu Asn Gln Arg Gly Met Phe Ser Tyr Thr
340 345 350

Gly Leu Ser Ala Ala Gln Val Asp Arg Leu Arg Glu Glu Phe Gly Val
355 360 365

Tyr Leu Ile Ala Ser Gly Arg Met Cys Val Ala Gly Leu Asn Thr Ala
370 375 380

Asn Val Gln Arg Val Ala Lys Ala Phe Ala Ala Val Met
385 390 395

<210> 33
<211> 35
<212> DNA
<213> Artificial sequence

<220>
<223> primer

<400> 33
ggtattgagg gtcgcgtgtt tcaaaaagtt gacgc 35

<210> 34
<211> 37
<212> DNA
<213> Artificial sequence

<220>
<223> primer

<400> 34
agaggagagt tagagccttа catcaccgca gcaaacg 37

<210> 35
<211> 35
<212> DNA
<213> Artificial sequence

<220>
<223> primer

108

<400> 35
ggtattgagg gtcgcatgga gtccaaagtc gttga 35

<210> 36
<211> 37
<212> DNA
<213> Artificial sequence

<220>
<223> primer

<400> 36
agaggagagt tagagcctta cacttggaaa acagcct 37

<210> 37
<211> 35
<212> DNA
<213> Artificial sequence

<220>
<223> primer

<400> 37
ggtattgagg gtcgcatgaa aaactggaaa acaag 35

<210> 38
<211> 37
<212> DNA
<213> Artificial sequence

<220>
<223> primer

<400> 38
agaggagagt tagagcctta cagcttagcg ccttcta 37

<210> 39
<211> 35
<212> DNA
<213> Artificial sequence

<220>
<223> primer

<400> 39
ggtattgagg gtcgcatgcg aggggcatta ttcaa 35

<210> 40
<211> 36
<212> DNA
<213> Artificial sequence

<220>
<223> primer

<400> 40

agaggagagt tagagcctca gcccttgagc gcgaag 36

<210> 41
<211> 1416
<212> DNA
<213> E. coli

<400> 41
atggaaaact ttaaacatct ccctgaaccg ttccgcattc gtgttattga gccagtaaaa 60

cgtaccaccc gcgcttatcg tgaagaggca attattaaat ccggtatgaa cccgttcctg 120

ctggatagcg aagatgtttt tatcgattta ctgaccgaca gcggcaccgg ggcggtgacg 180

cagagcatgc aggctgcgat gatgcgcggc gacgaagcct acagcggcag tcgtagctac 240

tatgcgttag ccgagtcagt gaaaaatatc tttggttatc aatacaccat tccgactcac 300

cagggccgtg gcgcagagca aatctatatt cggtactga ttaaaaaacg cgagcaggaa 360

aaaggcctgg atcgcagcaa aatggtggcg ttctctaact atttctttga taccacgcag 420

ggccatagcc agatcaacgg ctgtaccgtg cgtaacgtct atatcaaaga agccttcgat 480

acgggcgtgc gttacgactt taaaggcaac tttgaccttg agggattaga acgcggtatt 540

gaagaagttg gtccgaataa cgtgccgtat atcgttgcaa ccatcaccag taactctgca 600

ggtggtcagc cggtttcact ggcaaactta aaagcgatgt acagcatcgc gaagaaatac 660

gatattccgg tggtaatgga ctccgcgcgc tttgctgaaa acgcctattt catcaagcag 720

cgtgaagcag aatacaaaga ctggaccatc gagcagatca cccgcgaaac ctacaaatat 780

gccgatatgc tggcgatgtc cgccaagaaa gatgcgatgg tgccgatggg cggcctgctg 840

tgcatgaaag acgacagctt ctttgatgtg tacaccgagt gcagaaccct ttgcgtggtg 900

caggaaggct tcccgacata tggcggcctg gaaggcggcg cgatggagcg tctggcggta 960

ggtctgtatg acggcatgaa tctcgactgg ctggcttatc gtatcgcgca ggtacagtat 1020

ctggtcgatg gtctggaaga gattggcgtt gtctgccagc aggcgggcgg tcacgcggca 1080

ttcgttgatg ccggtaaact gttgccgcat atcccggcag accagttccc ggcacaggcg 1140

ctggcctgcg agctgtataa agtcgccggt atccgtgcgg tagaaattgg ctctttcctg 1200

ttaggccgcg atccgaaaac cggtaaacaa ctgccatgcc cggctgaact gctgcgttta 1260

accattccgc gcgcaacata tactcaaaca catatggact tcattattga agcctttaaa 1320

catgtgaaag agaacgcggc gaatattaaa ggattaacct ttacgtacga accgaaagta 1380

ttgcgtcact tcaccgcaaa acttaaagaa gtttaa 1416

<210> 42
<211> 1371
<212> DNA

<213> Citrobacter freundii

<400> 42
atgaattatc cggcagaacc cttccgtatt aaaagcgttg aaactgtatc tatgatcccg 60

cgtgatgaac gcctcaagaa aatgcaggaa gcgggttaca atactttcct gttaaattcg 120

aaagatattt atattgacct gctgacagac agtggcacta acgcaatgag cgacaagcag 180

tgggccggaa tgatgatggg tgatgaagcg tacgcgggca gcgaaaactt ctatcatctg 240

gaaagaaccg tgcaggaact gtttggcttt aaacatatcg ttccgactca ccagggcgt 300

ggcgcagaaa acctgttatc gcagctggct attaaacctg ggcaatatgt tgccgggaat 360

atgtatttca ctaccacccg ttatcaccag gaaaaaaatg gtgcggtgtt tgtcgatatc 420

gttcgtgacg aagcgcacga tgccggtctg aatattgcgt ttaaaggtga tatcgatctt 480

aaaaaattac aaaagctgat tgatgaaaaa ggcgcagaga atattgcgta tatctgcctg 540

gcggtgacgg ttaacctcgc gggcggccaa ccggtctcga tggctaacat gcgtgcggtg 600

cgtgaactga cagaagcgca tggcattaaa gtgttctacg acgctacccg ctgcgtagaa 660

aacgcctact ttatcaaaga gcaagagcag ggctttgaga acaagagcat cgccgagatc 720

gtgcatgaga tgttcagcta cgccgacggt tgtaccatga gtggtaaaaa agactgtctg 780

gtgaacatcg gcggcttcct gtgcatgaac gatgacgaaa tgttctcttc tgccaaagag 840

ttagtcgtgg tctacgaagg gatgccatct tacggcggcc tggccggacg tgatatggaa 900

gcgatggcga ttggcctgcg tgaagccatg cagtacgaat atattgagca ccgcgtgaag 960

caggttcgct acctgggcga taagctgaaa gccgctggcg taccgattgt tgaaccggta 1020

ggcggtcacg cggtattcct cgatgcgcgt cgcttctgcg agcatctgac gcaagatgag 1080

ttcccggcac aaagtctggc tgccagcatc tatgtggaaa ccggcgtgcg cagtatggag 1140

cgcggaatta tctctgcggg ccgtaataac gtgaccggtg aacaccacag accgaaactg 1200

gaaaccgtgc gtctgactat tccacgtcgt gtttatacct acgcacatat ggatgttgtg 1260

gctgacggta ttattaaact ttaccagcac aaagaagata ttcgcgggct gaagtttatt 1320

tacgagccga agcagttgcg tttctttact gcacgctttg attacatcta a 1371

<210> 43
<211> 35
<212> DNA
<213> Artificial sequence

<220>
<223> primer

<400> 43
ggtattgagg gtcgcatgga aaactttaaa catct 35

<210> 44
<211> 37
<212> DNA
<213> Artificial sequence

<220>
<223> primer

<400> 44
agaggagagt tagagcctta aacttcttta agttttg 37

<210> 45
<211> 35
<212> DNA
<213> Artificial sequence

<220>
<223> primer

<400> 45
ggtattgagg gtcgcatgaa ttatccggca gaacc 35

<210> 46
<211> 37
<212> DNA
<213> Artificial sequence

<220>
<223> primer

<400> 46
agaggagagt tagagcctta gatgtaatca aagcgtg 37

<210> 47
<211> 31
<212> DNA
<213> Artificial sequence

<220>
<223> primer

<400> 47
ccagggcacc ggcgcagagc aaatctatat t 31

<210> 48
<211> 30
<212> DNA
<213> Artificial sequence

<220>
<223> primer

<400> 48
tgcgccggtg ccctggtgag tcggaatggt 30

<210> 49
<211> 32
<212> DNA
<213> Artificial sequence

<220>
<223> primer

<400> 49
tcctgcacgc ggcaaagggt tctgcactcg gt 32

<210> 50
<211> 30
<212> DNA
<213> Artificial sequence

<220>
<223> primer

<400> 50
ctttgccgcg tgcaggaagg cttcccgaca 30

<210> 51
<211> 29
<212> DNA
<213> Artificial sequence

<220>
<223> primer

<400> 51
aggggaccgg cgcagaaaac ctgttatcg 29

<210> 52
<211> 32
<212> DNA
<213> Artificial sequence

<220>
<223> primer

<400> 52
tctgcgccgg tcccctggtg agtcggaaca at 32

<210> 53
<211> 29
<212> DNA
<213> Artificial sequence

<220>
<223> primer

<400> 53
gttagtccgc gtctacgaag ggatgccat 29

<210> 54

<211> 29
<212> DNA
<213> Artificial sequence

<220>
<223> primer

<400> 54
gtagacgcgg actaactctt tggcagaag 29

<210> 55
<211> 35
<212> DNA
<213> Artificial sequence

<220>
<223> primer

<400> 55
ggtattgagg gtcgcatgta cgaactggga gttgt 35

<210> 56
<211> 37
<212> DNA
<213> Artificial sequence

<220>
<223> primer

<400> 56
agaggagagt tagagcctta gtcaatatat ttcaggc 37

<210> 57
<211> 35
<212> DNA
<213> Artificial sequence

<220>
<223> primer

<400> 57
ggtattgagg gtcgcatgtc cggcatcgtt gtcca 35

<210> 58
<211> 37
<212> DNA
<213> Artificial sequence

<220>
<223> primer

<400> 58
agaggagagt tagagcctca gacatatttc agtccca 37

<210> 59
<211> 35

<212> DNA
<213> Artificial sequence

<220>
<223> primer

<400> 59
ggtattgagg gtcgcatgcg actgaacaac ctcgg 35

<210> 60
<211> 37
<212> DNA
<213> Artificial sequence

<220>
<223> primer

<400> 60
agaggagagt tagagcctca gttctccacg tattcca 37

<210> 61
<211> 35
<212> DNA
<213> Artificial sequence

<220>
<223> primer

<400> 61
ggtattgagg gtcgcatgag cgtggttcac cggaa 35

<210> 62
<211> 37
<212> DNA
<213> Artificial sequence

<220>
<223> primer

<400> 62
agaggagagt tagagcctca atcgatatat ttcagtc 37

<210> 63
<211> 35
<212> DNA
<213> Artificial sequence

<220>
<223> primer

<400> 63
ggtattgagg gtcgcatgag cctggttaat atgaa 35

<210> 64
<211> 37
<212> DNA

<213> Artificial sequence

<220>
<223> primer

<400> 64
agaggagagt tagagcctta tgactttaac gcgttga 37

<210> 65
<211> 684
<212> DNA
<213> C. testosteroni

<400> 65
atgtacgaac tgggagttgt ctaccgcaat atccagcgcg ccgaccgcgc tgctgctgac 60

ggcctggccg ccctgggctc cgccaccgtg cacgaggcca tgggccgcgt cggtctgctc 120

aagccctata tgcgccccat ctatgccggc aagcaggtct cgggcaccgc cgtcacggtg 180

ctgctgcagc ccggcgacaa ctggatgatg catgtggctg ccgagcagat tcagcccggc 240

gacatcgtgg tcgcagccgt caccgcagag tgcaccgacg gctacttcgg cgatctgctg 300

gccaccagct tccaggcgcg cggcgcacgt gcgctgatca tcgatgccgg cgtgcgcgac 360

gtgaagacgc tgcaggagat ggactttccg gtctggagca aggccatctc ttccaagggc 420

acgatcaagg ccaccctggg ctcggtcaac atccccatcg tctgcgccgg catgctggtc 480

acgcccggtg acgtgatcgt ggccgacgac gacggcgtgg tctgcgtgcc cgccgcgcgt 540

gccgtggaag tgctggccgc cgcccagaag cgtgaaagct tcgaaggcga aaagcgcgcc 600

aagctggcct cgggcatcct cggcctggat atgtacaaga tgcgcgagcc cctggaaaag 660

gccggcctga aatatattga ctaa 684

<210> 66
<211> 227
<212> PRT
<213> C. testosteroni

<400> 66

Met Tyr Glu Leu Gly Val Val Tyr Arg Asn Ile Gln Arg Ala Asp Arg
1 5 10 15

Ala Ala Ala Asp Gly Leu Ala Ala Leu Gly Ser Ala Thr Val His Glu
20 25 30

Ala Met Gly Arg Val Gly Leu Leu Lys Pro Tyr Met Arg Pro Ile Tyr
35 40 45

Ala Gly Lys Gln Val Ser Gly Thr Ala Val Thr Val Leu Leu Gln Pro
50 55 60

116

Gly Asp Asn Trp Met Met His Val Ala Ala Glu Gln Ile Gln Pro Gly
65 70 75 80

Asp Ile Val Val Ala Ala Val Thr Ala Glu Cys Thr Asp Gly Tyr Phe
85 90 95

Gly Asp Leu Leu Ala Thr Ser Phe Gln Ala Arg Gly Ala Arg Ala Leu
100 105 110

Ile Ile Asp Ala Gly Val Arg Asp Val Lys Thr Leu Gln Glu Met Asp
115 120 125

Phe Pro Val Trp Ser Lys Ala Ile Ser Ser Lys Gly Thr Ile Lys Ala
130 135 140

Thr Leu Gly Ser Val Asn Ile Pro Ile Val Cys Ala Gly Met Leu Val
145 150 155 160

Thr Pro Gly Asp Val Ile Val Ala Asp Asp Asp Gly Val Val Cys Val
165 170 175

Pro Ala Ala Arg Ala Val Glu Val Leu Ala Ala Ala Gln Lys Arg Glu
180 185 190

Ser Phe Glu Gly Glu Lys Arg Ala Lys Leu Ala Ser Gly Ile Leu Gly
195 200 205

Leu Asp Met Tyr Lys Met Arg Glu Pro Leu Glu Lys Ala Gly Leu Lys
210 215 220

Tyr Ile Asp
225

<210> 67
<211> 42
<212> DNA
<213> Artificial sequence

<220>
<223> primer

<400> 67
actcggatcc gaaggagata tacatatgta cgaactggga ct 42

<210> 68
<211> 33
<212> DNA

<213> Artificial sequence

<220>
<223> primer

<400> 68
cggctgtcga ccgttagtca atatatttca ggc 33

<210> 69
<211> 31
<212> DNA
<213> Artificial sequence

<220>
<223> primer

<400> 69
cgcggatcca taatggttga gaacattacc g 31

<210> 70
<211> 30
<212> DNA
<213> Artificial sequence

<220>
<223> primer

<400> 70
acgcgtcgac ttacagcact gccacaatcg 30

<210> 71
<211> 32
<212> DNA
<213> Artificial sequence

<220>
<223> primer

<400> 71
ccggaattca taatggtcga actgggagtt gt 32

<210> 72
<211> 33
<212> DNA
<213> Artificial sequence

<220>
<223> primer

<400> 72
gaatgcggcc gcttagtcaa tatatttcag gcc 33

<210> 73
<211> 15
<212> DNA
<213> Artificial sequence

<220>
<223> primer

<400> 73
ggtattgagg gtcgc 15

<210> 74
<211> 17
<212> DNA
<213> Artificial sequence

<220>
<223> primer

<400> 74
agaggagagt tagagcc 17

(12) SOLICITUD INTERNACIONAL PUBLICADA BAJO EL TRATADO DE COOPERACIÓN DE PATENTES (PCT)

(19) Organización Mundial de Propiedad Intelectual
Oficina Internacional

(43) Fecha de Publicación Internacional
6 de Noviembre de 2003 (06.11.2003)

PCT

(10) Número de Publicación Internacional
WO 03/091396 A2

(51) Clasificación Internacional de Patentes: C12N

(21) Numero de Solicitud Internacional:
PCT/US03/12588

(22) Fecha de Entrega Internacional:
23 de Abril de 2003 (23.04.2003)

(25) Lenguaje de entrega: Inglés

(26) Lenguaje de Publicación: Inglés

(30) Datos de la Prioridad:
60/374,831 23 de Abril de 2002 (23.04.2002) US

(71) Solicitante: CARGILL, INCORPORATED [US/US]; P. O. Box 5624, Minneapolis, MN 5544-5624 (US).

(75) Inventores: ABRAHAM, Timothy, W.; 18507 Apple Tree Court, Minnetonka, MN 55345 (US). CAMERON, Douglas, C.; 3590 Ranier Lane, N. Plymouth, MN 55447 (US). DALLUGE, Joseph; 3516 County Road 30 S. W., Waverly, MN 55390 (US). HICKS, Paula, M.; 11601 Tanglewood Drive, Eden Praire, MN 55347 (US). HOBSON, Russell, J.; 316 Vinewood Lane North, Plymouth, MN 55441 (US). MCFARLAN, Sara, C.; 2151 Bayard Avenue, St. Paul, MN 55116 (US). MILLS, Jim; 22360 Yuma Lane North, Plymouth, MN 55447 (US). ROSAZZA, John; 453 Linder Road N. E., Iowa City, IA 55240 (US).

(74) Apoderado: SKELTON, Jeffrey J.; Cargill Incorporated, P. O. Box 5624, Minneapolis, MN 55440-5624 (US)

(81) Estados Designados (*nacional*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FIM GB, GD, GE, GH, GM, HR, HY, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UAM UG, UZ, VN, YU, ZA, ZM, ZW.

(84) Estados Designados (regional): patente ARIPO (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Patente Euroasiática (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), Patente europea (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), Patente OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Publicado:

— *sin el informe de búsqueda internacional y a ser vuelto a publicar al recibir ese informe*

Para los códigos de dos letras y otras abreviaturas, referirse a las "Notas de Guía sobre Códigos y Abreviaturas" que aparece al comienzo de cada emisión regular de la Gaceta del PCT.

(54) Título: POLIPÉPTIDOS Y RUTAS BIOSINTÉTICAS

(57) Resumen: Se proporcionan métodos y composiciones que pueden utilizarse para hacer monatina a partir de glucosa, triptófano, ácido Indole-3-láctico, Indole-3-piruvato, y ácido 2-hidroxi-2-(iridol-3-ilmetil)-4-ceto glutámico. También se divulgan métodos para producir intermedios de indole-3-piruvato y ácido 2-hidroxi-2-(iridol-3-ilmetil)-4-ceto glutámico. Se proporcionan composiciones que incluyen moléculas de ácido nucleico, polipéptidos, estructuras químicas, y células. Los métodos incluyen proceso *in vitro* e *in vivo*, y los métodos *in vitro* incluyen reacciones químicas.

120

121

CAMPO DE LA INVENCIÓN

Esta descripción proporciona los polipéptidos y las vías biosintéticas que son útiles en la producción de indol-3-piruvato, ácido 2-hidroxi-2-(indol-3-ilmetil)-4-cetoglutárico (MP) y/o monatina.

ANTECEDENTES DE LA INVENCIÓN

Piruvato de indol

El indol-3-piruvato es un antioxidante fuerte que se cree contra-ataca la tensión oxidativa en tejidos con altas concentraciones de oxígeno (Politi et al. "Recent advances in Tryptophan Research", editado por G.A. Filippini et al. Plenum Press, Nueva York, 1996, pp. 291-8). El indol-piruvat o piruvato de indol también es un intermediario en una vía para producir el ácido indol acético (IAA), la hormona auxina principal para el desarrollo de las plantas (factor de promoción de crecimiento difundible). IAA es activo en cantidades de submicrogramos en una gama de procesos fisiológicos que incluyen dominancia apical, tropismos, alargamiento de las raíces, inducción de la división de las células cambiales e iniciación de las raíces. Las auxinas sintéticas son utilizadas en horticultura para inducir enraizamiento y para promover el establecimiento y desarroll del fruto. A altas concentraciones las auxinas sintéticas son

herbicidas efectivos contra plantas de hoja ancha. Las auxinas naturales producidas por la fermentación pueden ser consideradas más ambientalmente amigables que los herbicidas químicamente producidos. Los reguladores del crecimiento tuvieron ventas mundiales en 1999 de 0.4 billones de libras (1.4 mil millones de dólares).

Algunos ejemplos de patentes sobre el ácido indolacético y derivados del mismo incluyen: la Patente de los Estados Unidos No. 5,843,782: Micropropagación de plantas de rosa, auxina utilizada en medio de cultivo; y la Patente de los Estados Unidos No. 5,952,231: Micropropagación de plantas de rosa.

Además de las utilidades relacionadas a las plantas, el ácido indolacético es útil en aplicaciones farmacéuticas. Por ejemplo, la Patente de los Estados Unidos No. 5,173,497: "Método para la preparación de los ácidos alfa-oxopirrolo[2,3-B]indol-acéticos y derivados" propone el uso de estos compuestos en el tratamiento del deterioro de la memoria tal como aquel asociado con enfermedad de Alzheimer y la demencia senil. El mecanismo propuesto en la Patente de los Estados Unidos No. 5,173,497 es que estos compuestos inhiben el polipéptido acetilcolinesterasa e incrementa los niveles de acetilcolina en el cerebro.

El indol-3-carbinol es producido a partir del ácido

indol-3-acético por la oxidación catalizada por la peroxidasa, y puede ser fácilmente convertido a diindolilmetano. Se reportan que ambos compuestos eliminan las toxinas y promueven la promoción de hormonas benéficas para la salud de las mujeres.

Derivados de triptofano

D-triptofano clorado ha sido identificado como un endulzante no nutritivo, y existe interés creciente en conseguir también otros derivados. La monatina es un endulzante natural que es similar en composición al aminoácido triptofano. Este puede ser extraído de las cortezas de las raíces de un arbusto de Sudáfrica, *Sclerochiton ilicifolius*, y es promisorio en la industria de los alimentos y bebidas como un endulzante de alta intensidad. Algunos ejemplos de patentes sobre monatina incluye: la Patente de los Estados Unidos No. 5994559 Síntesis de endulzante natural de alta intensidad monatina-A, Patente de los Estados Unidos No. 4975298 compuestos de 3-(1-amino-1,3-dicarboxi-3-hidroxi-but-4-il)-indol, patente de los Estados Unidos No. 5128164 composición para el consumo humano que contiene los compuestos 3-(1-amino-1,3-dicarboxi-3-hidroxi-but-4-il)-indol; y la patente de los Estados Unidos No. 5128482 proceso para la producción de 3-1(1-amino-1,3-dicarboxi-3-hidroxi-but-4-il)indol.

Algunos de los precursores de la monatina descritos en

la presente pueden también ser útiles como endulzantes sintéticos o como intermediarios en la síntesis de derivados de monatina.

BREVE DESCRIPCIÓN DE LA INVENCIÓN

La descripción proporciona varias rutas biosintéticas para la elaboración de monatina a partir de glucosa, triptofano, ácido indol-3-láctico y/o a través de los precursores de monatina tales como el indol-2-piruvato y el 2-hidroxi-2-(indol-3-ilmetil)-4-ceto-glutárico. Las secuencias de polipéptidos y de ácidos nucleicos que pueden ser utilizadas para elaborar la monatina, el indol-3-piruvato, y el ácido 2-hidroxi-2-(indol-3-ilmetil)-4-ceto-glutárico son también descritas.

La monatina puede ser producida a través del indol-3-piruvato, el ácido 2-hidroxi-2-(indol-3-ilmetil)-4-ceto-glutárico (precursor de monatina, MP, la forma alfa-ceto de la monatina), el ácido indol-3-láctico, triptofano, y/o glucosa (FIG. 1). Se describen los métodos para producir o elaborar la monatina o sus intermediarios mostrados en las FIGS. 1-3 y 11-13 que involucran la conversión de un sustrato a un primer producto, y luego la conversión del primer producto a un segundo producto y así sucesivamente, hasta que se crea el producto final deseado.

Las FIGS. 1-3 y 11-13 muestran productos intermediarios potenciales y productos finales en cajas. Por

ejemplo, una conversión de un producto a otro, tal como glucosa a triptofano, triptofano a indol-3-piruvato, indol-3-piruvato a MP, MP a monatina o ácido indol-3-láctico (indol-lactato) a indol-3-piruvato, puede ser realizado mediante el uso de estos métodos. Estas conversiones pueden ser facilitadas química o biológicamente. El término "convertir" se refiere al uso ya sea de medios químicos o de polipéptidos en una reacción que cambia un primer intermediario a un segundo intermediario. El término "conversión química" se refiere a las reacciones que no son activamente facilitadas por los polipéptidos. El término "conversión biológica" se refiere a las reacciones que son activamente facilitadas por los polipéptidos. Las conversiones pueden tener lugar *in vivo* o *in vitro*. Cuando se utilizan conversiones biológicas, los polipéptidos y/o las células pueden ser inmovilizados sobre soportes tales como mediante acoplamiento químico sobre soportes poliméricos. La conversión puede ser lograda utilizando cualquier reactor conocido para una persona de experiencia ordinaria en la técnica, por ejemplo en un reactor por lotes o en un reactor continuo.

Son proporcionados también los métodos que incluyen poner en contacto un primer polipéptido con un sustrato y elaborando un primer producto, y luego poniendo en contacto el primer producto creado con un segundo polipéptido y creando un segundo producto, y luego poniendo en contacto el

segundo creado, con un tercer polipéptido y creando un tercer producto, por ejemplo la monatina. Los polipéptidos utilizados y los productos producidos son mostrados en las FIGS. 1-3 y 11-13.

Los polipéptidos, y sus secuencias de codificación, que pueden ser utilizados para revisar las conversiones mostradas en las FIGS. 1-3 y 11-13, son también descritos. En algunos ejemplos, los polipéptidos que tienen una o más mutaciones puntuales que permiten especificidad de sustrato y/o actividad de los polipéptidos que van a ser modificados, se utilizan para elaborar la monatina.

Las células aisladas y recombinantes que producen la monatina son también descritas. Estas células pueden ser cualquier célula tal como una planta, animal, bacteria, levadura, alga, arqueobacterias o células de hongos.

En un ejemplo particular, las células descritas incluyen una o más de las siguientes actividades, por ejemplo dos o más o tres o más de las siguientes actividades: triptofano-aminotransferasa (EC 2.6.1.27), tirosina(aromático)-aminotransferasa (EC 2.6.1.5), aminotransferasa de sustratos múltiples (EC 2.6.1.-), aspartato-aminotransferasa (EC 2.6.1.1), triptofano-deshidrogenasa (EC 1.4.1.19), triptofano-fenilpiruvato-transaminasa (EC 2.6.1.28), oxidasa de L-aminoácido (EC 1.4.3.2), triptofano-oxidasa (sin número EC, Hadar et al., *J.*

126

Bacteriol 125:1096-1104, 1976 y Furuya et al., *Biosci Biotechnol Biochem* 64:1486-93, 2000), D-aminoácido deshidrogenasa (EC 1.4.99.1), D-aminoácido-oxidasa (EC 1.4.3.3), D-alanina aminotransferasa (EC 2.6.1.21), sintasa/liasa (EC 4.1.3-), tal como 4-hidroxi-4-metil-2-oxoglutarato-aldolasa (EC 4.1.3.17) o 4-hidroxi-2-oxoglutarato-aldolasa (EC 4.1.3.16), sintasa/liasa (4.1.2-), D-triptofano aminotransferasa (Kohiba y Mito, Proceedings of the 8th International Symposium on Vitamin B_6 and Carbonyl Catalysis, Osaka, Japón 1990), fenilalanina-deshidrogenasa (EC 1.4.1.20) y/o glutamato-deshidrogenasa (EC 1.4.1.2, 1.4.1.3, 1.4.1.4).

En otro ejemplo más, las células incluyen uno o más, por ejemplo, dos o más, o tres o más, de las siguientes actividades: indol-lactato-deshidrogenasa (EC 1.1.1.1110), R-4-hidroxifenil-lactato-deshidrogenasa (EC 1.1.1.222), 3-(4)-hidroxifenilpiruvato-reductasa (EC 1.1.1.237), lactato-deshidrogenasa (EC 1.1.1.27, 1.1.1.28, 1.1.2.3), (3-imidazol-5-il)lactato-deshidrogenasa (EC 1.1.1.111), lactato oxidasa (EC 1.1.3-), sintasa/liasa (4.1.3-) tal como 4-hidroxi-4-metil-2-oxoglutarato-aldolasa (EC 4.1.3.17) o 4-hidroxi-2-oxoglutarato-aldolasa (EC 4.1.3.16), sintasa/liasa (4.1.2.-), triptofano-deshidrogenasa (EC 1.4.1.19), triptofano-fenilpiruvato-transaminasa (EC 2.6.1.28), triptofano-aminotransferasa (EC 2.6.1.27), tirosina(aromático)-

aminotransferasa (EC 2.6.1.5), aminotransferasa de sustratos múltiples (EC 2.6.1-), aspartato-aminotransferasa (EC 2.6.1.1), fenilalanina-deshidrogenasa (EC 1.4.1.20), glutamato-deshidrogenasa (EC 1.4.1.2, 1.4.1.3, 1.4.1.4), D-aminoácido-deshidrogenasa (EC 1.4.99.1), D-triptofano-aminotransferasa y/o D-alanina-aminotransferasa (EC 2.6.1.21).

Además, las células descritas pueden incluir uno o más de las siguientes actividades, por ejemplo dos o más o tres o más de las siguientes actividades: triptofano-aminotransferasa (EC 2.6.1.27), tirosina(aromático)-aminotransferasa (EC 2.6.1.5), aminotransferasa de sustratos múltiples (EC 2.6.1-), aspartato-aminotransferasa (EC 2.6.1.1), triptofano-deshidrogenasa (EC 1.4.1.19), triptofano-fenilpiruvato-transaminasa (EC 2.6.1.28), L-aminoácido-oxidasa (EC 1.4.3.2), triptofano-oxidasa (sin número EC), D-aminoácido-deshidrogenasa (EC 1.4.99.1), D-aminoácido-oxidasa (EC 1.4.3.3), D-alanina-aminotransferasa (EC 2.6.1.21), indol-lactato-deshidrogenasa (EC 1.1.1.110), R-4-hidroxifenil-lactato-deshidrogenasa (EC 1.1.222), 3-(4)-hidroxifenilpiruvato-reductasa (EC 1.1.1.237), lactato-deshidrogenasa (EC 1.1.1.27, 1.1.1.28, 1.1.2.3), (3-imidazol-5-il)lactato-deshidrogenasa (EC 1.1.1.111), lactato-oxidasa (EC 1.1.3.-), sintasa/liasa (4.1.3-), tal como 4-hidroxi-4-metil-2-oxoglutarato-aldolasa (EC 4.1.3.17) o 4-hidroxi-2-

128

oxoglutarato-aldolasa (EC 4.1.3.16), sintasa/liasa (4.2.1-), glutamato-deshidrogenasa (EC 1.4.1.2, 1.4.1.3, 1.4.1.4), fenilalanina-deshidrogenasa (EC 1.4.1.20), y/o D-triptofano-aminotransferasa.

La monatina puede ser producida mediante un método que incluye poner en contacto el triptofano y/o el ácido indol-3-láctico con un primer polipéptido, en donde el primer polipéptido convierte el triptofano y/o el ácido indol-3-láctico a indol-3-piruvato (ya sea la forma D o la forma L del triptofano o el ácido indol-3-láctico puede ser utilizado como el sustrato que es convertido a indol-3-piruvato; una persona de experiencia ordinaria en la técnica apreciará que los polipéptidos elegidos para este paso muestran idealmente la especificidad apropiada), poniendo en contacto el indol-3-piruvato resultante con un segundo polipéptido, en donde el segundo polipéptido convierte el indol-3-piruvato a ácido 2-hidroxi-2-(indol-3-ilmetil)-4-cetoglutárico (MP), y poniendo en contacto el MP con un tercer polipéptido, en donde el tercer polipéptido convierte MP a monatina. Los polipéptidos ejemplares que pueden ser utilizados para estas conversiones son mostrados en las FIGS. 2 y 3.

Otro aspecto más de la invención proporciona las composiciones tales como MP, las células que contienen al menos dos polipéptidos, o algunas veces al menos tres o al menos cuatro polipéptidos, que son codificados sobre al menos

una secuencia exógena de ácido nucleico.

Estos y otros aspectos de la descripción son aparentes a partir de la siguiente descripción detallada y de los ejemplos ilustrativos.

BREVE DESCRIPCIÓN DE LAS FIGURAS

La FIG.1 muestra las vías biosintéticas utilizadas para producir la monatina y/o el indol-3-piruvato. Una vía produce el indol-3-piruvato vía el triptofano, mientras que otra más produce el indol 3-piruvato vía el ácido indol-3-láctico. La monatina es subsecuentemente producida vía el intermediario de ácido 2-hidroxi-2-(indol-3-ilmetil)-4-cetoglutárico (MP).

Los compuestos mostrados en recuadros son sustratos y productos producidos en las vías biosintéticas.

Las composiciones adyacentes a las flechas son cofactores, o reactivos que pueden ser utilizados durante la conversión de un sustrato a un producto. El cofactor o el reactivo utilizado dependerá del polipéptido utilizado para el paso particular de la vía biosintética. El cofactor PLP (piridoxal-5'-fosfato) puede catalizar las reacciones independientemente de un polipéptido, y por lo tanto, al proporcionar meramente PLP puede permitir la progresión del sustrato hasta el producto.

La FIG. 2 es un diagrama más detallado de la vía biosintética que utiliza el intermediario de MP. Los

sustratos para cada paso en las vías son mostrados en los recuadros. Los polipéptidos que permiten la conversión entre los sustratos son listados adyacentemente a las flechas entre los sustratos. Cada polipéptido es descrito por su nombre común y en un número de clase enzimática (EC).

La FIG. 3 muestra un diagrama más detallado de la vía biosintética de la conversión del ácido indol-3-láctico a indol-3-piruvato. Los sustratos son mostrados en recuadros, y los polipéptidos que permiten la conversión entre los sustratos son listados adyacentes a la flecha entre los sustratos. Cada polipéptido es descrito por su nombre común y un número de la clase enzimática (EC).

La FIG. 4 muestra una posible reacción para elaborar MP por medios químicos.

Las FIGS. 5A y 5B son cromatogramas que muestran la identificación por LC/MS de la monatina producida enzimáticamente.

La FIG. 6 es un espectro de masa de electrorrocío de la monatina enzimáticamente sintetizada.

Las FIGS. 7A y 7B muestran cromatogramas de los análisis de iones hijos por LC/MS/MS de la monatina producida en una mezcla enzimática.

La FIG. 8 es un cromatograma que muestra la medición de masa de alta resolución de la monatina producida enzimáticamente.

Las FIGS. 9A-9C son cromatogramas que muestran la separación quiral de (A) R-triptofano, (B) S-triptofano, y (C) monatina producida enzimáticamente.

La FIG. 10 es un diagrama de barras que muestra la cantidad relativa de la monatina producida en células bacterianas después de la inducción con IPTG. El signo (-) indica una falta de adición de sustrato (no se agregó triptofano o piruvato).

Las FIGS. 11-12 son diagramas esquemáticos que muestran las vías utilizadas para incrementar el rendimiento de la monatina producida a partir del triptofano o del indol-3-piruvato.

La FIG. 13 es un diagrama esquemático que muestra una vía que puede ser utilizada para incrementar el rendimiento de la monatina producida a partir del triptofano o del indol-3-piruvato.

LISTADO DE SECUENCIAS

Las secuencias de ácidos nucleicos y de aminoácidos listadas en el listado de secuencias anexo, se muestran utilizando las abreviaturas de una sola letra estándares, para las bases nucleotídicas, y el código de tres letras para los aminoácidos. Únicamente una hebra de cada secuencia de ácido nucleido es mostrada, pero se entiende que la hebra complementaria es incluida por cualquier referencia a la hebra mostrada.

132

SEQ ID NOS.: 1 y 2 muestran las secuencias de ácidos nucleicos y de aminoácidos de una aminotransferasa de *Sinorhizobium meliloti*, respectivamente (el gen tatA, llamado una aminotransferasa de tirosina o aromático en la literatura).

SEQ ID NOS.: 3 y 4 muestran las secuencias de ácidos nucleicos y de aminoácidos de una tirosina-aminotransferasa de *Rhodobacter sphaeroides* (2.4.1), respectivamente (por homología con tatA (SEQ ID NOS.: 1 y 2) que se predice es una "aspartato-aminotransferasa" por software genómico.

SEQ ID NOS.: 5 y 6 muestran las secuencias de ácidos nucleicos y de aminoácidos de una aminotransferasa de *Rhodobacter sphaeroides* (35053), respectivamente (nuevo, clonado con base en 2.4.1 secuencia SEQ ID NOS.: 3 y 4).

SEQ ID NOS.: 7 y 8 muestran las secuencias de ácidos nucleicos y de aminoácidos de una aminotransferasa de amplios sustratos (bsat) proveniente de *Leishmania major*, respectivamente.

SEQ ID NOS.: 9 y 10 muestran las secuencias de ácidos nucleicos y de aminoácidos de una aminotransferasa de aromático (araT) de *Bacillus subtilis*, respectivamente.

SEQ ID NOS.: 11 y 12 muestran nuevas secuencias de ácido nucleico y de aminoácidos de una aromático-aminotransferasa (araT) de *Lactobacillus amylovorus*, respectivamente (por homología identificada como una

aromático-aminotransferasa).

SEQ ID NOS.: 13 y 14 muestran las secuencias de ácidos nucleicos y de aminoácidos de una aminotransferasa de sustratos múltiples (msa) proveniente de *R. Sphaeroides* (35053), respectivamente (identificada como una aminotransferasa de sustratos múltiples por homología al No. de Acceso AAAE01000093.1, pb 14743-16155 y No. de Acceso ZP00005082.1).

SEQ ID NOS.: 15 y 16 muestran cebadores utilizados para clonar la secuencia de D-alanina-aminotransferasa (dat) de *B. subtilis*.

SEQ ID NOS.: 17 y 18 muestran cebadores utilizados para clonar la secuencia de tatA de *S. meliloti*.

SEQ ID NOS.: 19 y 20 muestran cebadores utilizados para clonar la secuencia de araT de *B. subtilis*.

SEQ ID NOS.: 21 y 22 muestran cebadores utilizados para clonar las secuencias de aminotransferasa de sustratos múltiples de *Rhodobacter sphaeroides* (2.4.1. y 35053).

SEQ ID NOS.: 23 y 24 muestran cebadores utilizados para clonar la secuencia bsat de *Leishmania major*.

SEQ ID NOS.: 25 y 26 muestran cebadores utilizados para clonar la secuencia de araT de *Lactobacillus amylovorus*.

SEQ ID NOS.: 27 y 28 muestran cebadores utilizados para clonar la secuencia de tatA de *R. Sphaeroides* (ambas 2.4.1 y 35053).

SEQ ID NOS.: 29 y 30 muestran cebadores utilizados para clonar la secuencia de aspC de *E. coli* (secuencia de genes Genbank No. de Acceso: AE000195.1, secuencia de proteína Genbank No. de Acceso: AAC74014).

SEQ ID NOS.: 31 y 32 muestran las secuencias de ácidos nucleicos y de aminoácidos de la aromático-aminotransferasa (tyrB) de *E. coli*, respectivamente.

SEQ ID NOS.: 33 y 34 muestran cebadores utilizados para clonar la secuencia de tyrB de *E. coli*.

SEQ ID NOS.: 35-40 muestran cebadores utilizados para clonar los polipéptidos con actividad de 4-hidroxi-2-oxoglutarato aldolasa (KHG) (EC 4.1.3.16).

SEQ ID NOS.: 41 y 42 muestran las secuencias de ácidos nucleicos de la triptofanasa (tna) de *E. coli* y tirosina-fenol-liasa (tpl) de *Citrobacter freundii* que codifican para las proteínas P00913 (GI:401195) y P31013 (GI:401201), respectivamente.

SEQ ID NOS.: 43-46 muestran cebadores utilizados para clonar los polipéptidos de triptofanasa y los polipéptidos de β-tirosinasa (tirosina-fenol-liasa).

SEQ ID NOS.: 47-54 muestran cebadores utilizados para mutar los polipéptidos de triptofanasa y los polipéptidos de β-tirosinasa.

SEQ ID NOS.: 55-64 muestran cebadores utilizados para

clonar los polipéptidos con actividad de 4-hidroxi-4-metil-2-oxoglutarato-aldolasa (EC 4.1.3.17).

SEQ ID NOS.: 65 y 66 muestran las secuencias de ácidos nucleicos y de aminoácidos de la 4-hidroxi-4-metil-2-oxoglutarato-aldolasa (proA) de *C. testosteroni*, respectivamente.

SEQ ID NOS.: 67-68 muestran cebadores utilizados para clonar la 4-hidroxi-4-metil-2-oxoglutarato-aldolasa (proA) de *C. testosteroni* en un operón con aspC de *E. coli* en pET30 Xa/LIC.

SEQ ID NOS.: 69-72 muestran cebadores utilizados para clonar aspC de *E. coli* y proA de *C. testosteroni* en pESC-his.

SEQ ID NOS.: 73-74 muestran las secuencias agregadas al extremo 5' de los cebadores utilizados para clonar los genes descritos en la presente.

DESCRIPCIÓN DETALLADA DE LAS DIVERSAS MODALIDADES

Abreviaturas y términos

Las siguientes explicaciones de los términos y métodos se proporcionan para describir mejor la presente descripción, y para guiar a aquellos de experiencia ordinaria en la técnica en la práctica de la presente descripción. Como se utiliza en la presente, "que comprende" significa "que incluye" y las zonas singulares "un" o "uno" o "una" o "el, la, los, las" incluyen las referencias plurales a no ser que el contexto lo dicte claramente de otro modo. Por ejemplo, la

136

referencia a "que comprende una proteína" incluye una o una pluralidad de tales proteínas, y la referencia a "que comprende las células" incluye la referencia a una o más células equivalentes de las mismas conocidas por aquellos expertos en la técnica, y así sucesivamente.

A no ser que se explique de otro modo, todos los términos técnicos y científicos utilizados en la presente tienen los mismos significados que son comúnmente comprendidos por una persona de experiencia ordinaria en la técnica a la cual pertenece esta descripción. Aunque los métodos y materiales similares o equivalentes a aquellos descritos en la presente pueden ser utilizados en la práctica o prueba de la presente descripción, se describen en seguida los métodos y materiales adecuados. Los materiales, métodos y ejemplos son ilustrativos únicamente y no se pretende que sean limitantes. Otras características y ventajas de la descripción son aparentes a partir de la siguiente descripción detallada y de las reivindicaciones.

ADNc (ADN complementario): Una pieza de ADN que carece de los segmentos de no codificación, internos (intrones) y las secuencias reguladoras que determinan la transcripción. El ADNc puede ser sintetizado en el laboratorio mediante transcripción inversa del ARN mensajero extraído de las células.

Sustitución conservadora: Una o más sustituciones de

aminoácidos (por ejemplo, 2,5 ó 10 residuos) para los residuos de aminoácidos que tienen similares propiedades bioquímicas. Típicamente, las sustituciones conservadoras tienen poco o ningún impacto sobre la actividad de un polipéptido resultante. Por ejemplo, idealmente, un polipéptido de triptofano-aminotransferasa que incluye una o más sustituciones conservadoras conserva la actividad de triptofano-aminotransferasa. Puede ser producido un polipéptido para contener una o más sustituciones conservadoras mediante la manipulación de la secuencia de nucleótidos que codifica para ese polipéptido utilizando, por ejemplo, procedimientos estándares tales como la mutagénesis dirigida al sitio o PCR.

Las variantes sustitucionales son aquellas en las cuales al menos un residuo en la secuencia de aminoácidos ha sido removido y en su lugar ha sido insertado un residuo diferente. Los ejemplos de aminoácidos que pueden ser sustituidos por un aminoácido original en una proteína, y que son considerados como sustituciones conservadoras incluyen: Ala sustituida con ser o thr; arg sustituido con gln, his, o lys; asn sustituido con glu, gln, lys, his, asp; asp sustituido con asn, glu, o gln; cys sustituido con ser o ala; gln sustituido con asn, glu, lys, his, asp, o arg; glu sustituido con asn, gln lys o asp; gly sustituido con pro; his sustituido con asn, lys, gln, arg, tyr; ile sustituido

con leu, met, val, phe; leu sustituido con ile, met, val, phe; lys sustituido con asn, glu, gln, his, arg; met sustituido con ile, leu, val, phe; phe sustituido con trp, tyr, met, ile o leu; ser sustituido con thr, ala; thr sustituido con ser o ala; trp sustituido con phe, tyr; tyr sustituido con his, phe, o trp; y val sustituido con met, ile, leu.

La información adicional respecto a las sustituciones conservadoras pueden ser encontradas en, entre otros sitios, Ben-Bassat et al., (*J. Bacteriol*, 169:751-7, 1987), O'Regan et al., (*Gene* 77:237-51, 1989), Sahin-Toth et al., (*Protein Sci*. 3:240-7, 1994), Hochuli et al., (Bio/Technology 6:1321-5, 1988, (WO 00/67796 (Curd et al), y en los libros de textos estándares de genética y biología molecular.

Exógenos: El término "exógeno(a)" como se utiliza en la presente, con referencia al ácido nucleico y una célula particular, se refiere a cualquier ácido nucleico que no se origina de una célula particular como se encuentra en la naturaleza. De este modo, el ácido nucleico de origen natural es considerado como exógeno para una célula una vez que se introduce dentro de la célula. El ácido nucleico que es de origen natural puede ser también exógeno para una célula particular. Por ejemplo, un cromosoma completo aislado de una célula de la persona X es un ácido nucleico exógeno con respecto a una célula de la persona Y una vez que el

cromosoma es introducido dentro de la célula de Y.

Funcionalmente Equivalente: Que tiene una función equivalente. En el contexto de una enzima, las moléculas funcionalmente equivalentes incluyen diferentes moléculas que conservan la función de la enzima. Por ejemplo, los equivalentes funcionales pueden ser proporcionados por alteraciones en la secuencia en una secuencia enzimática, en donde el péptido con una o más alteraciones secuenciales conserva una función del péptido no alterado, tal que éste conserva su actividad enzimática. En un ejemplo particular, una triptofano-aminotransferasa funcionalmente equivalente conserva la habilidad para convertir el triptofano a indol-3-piruvato.

Ejemplos de alteraciones de secuencias incluyen, pero no están limitados a, sustituciones conservadoras, supresiones, mutaciones, desplazamientos estructurales e inserciones. En un ejemplo, un polipéptido dado se enlaza a un anticuerpo, y un equivalente funcional es un polipéptido que se enlaza al mismo anticuerpo. De este modo, un equivalente funcional incluye los péptidos que tienen la misma especificidad de enlace que un polipéptido, y que puede ser utilizado como un reactivo en lugar del polipéptido. En un ejemplo, un equivalente funcional incluye un polipéptido donde la secuencia de enlace es discontinua, en donde el anticuerpo se enlaza a un epitopo lineal. De este modo, si la

140

secuencia peptídica es MPELANDLGL (aminoácidos 1-10 de la SEQ ID No.: 12) un equivalente funcional incluye epitopos discontinuos, que pueden aparecer como sigue (** = cualquier número de aminoácidos de intervención): NH2-**-M**P**E**L**A**N**D**L**G**L-COOH. En este ejemplo, el polipéptido es funcionalmente equivalente a los aminoácidos 1-10 de la SEQ ID No.: 12 si la estructura tridimensional del polipéptido es tal que éste puede enlazarse a un anticuerpo monoclonal que se enlaza a los aminoácidos 1-10 de la SEQ ID No.: 12.

Hibridación: El término "hibridación" como se utiliza en la presente se refiere a un método de prueba para la complementariedad en la secuencia nucleotídica de dos moléculas de ácido nucleico, con base en la habilidad de ADN y/o el ARN complementarios de una sola hebra, para formar una molécula dúplex. Las técnicas de hibridación en ácidos nucleicos pueden ser utilizados para obtener un ácido nucleico aislado dentro del alcance de la invención. En resumen, cualquier ácido nucleico que tenga alguna homología para una secuencia descrita en la SEQ ID No.: 11, puede ser utilizado como una sonda para identificar un ácido nucleico similar mediante hibridación bajo condiciones de exigencia moderada a alta. Una vez identificado, el ácido nucleico puede ser luego purificado, secuenciado y analizado para determinar si éste está dentro del alcance de la presente

descripción.

La hibridación puede ser realizada mediante el análisis de Southern o de Nothern, para identificar la secuencia de ADN o ARN respectivamente, que se híbrida a una sonda. La sonda puede ser marcada con una biotina, digoxigenina, un polipéptido, o un radioisótopo tal como ^{32}P. El ADN o ARN que va a ser analizado puede ser electroforéticamente separado sobre un gel de agarosa o de poliacrilamida, transferido a nitrocelulosa, nailon u otra membrana adecuada, e hibridado con la sonda utilizando técnicas estándares bien conocidas en la técnica tales como aquellas descritas en las secciones 7.39-7.52 de Sambrook et al, (1989) Molecular Cloning, segunda edición, Cold Spring Harbor Laboratory, Plainview, NY. Típicamente, una sonda es al menos de aproximadamente 20 nucleótidos de longitud. Por ejemplo, una sonda correspondiente a una secuencia contigua de 20 nucleótidos descrita en la SEQ ID No.: 11, puede ser utilizada para identificar un ácido nucleico idéntico o similar. Además, pueden ser utilizadas sondas más largas o más cortas de 20 nucleótidos.

La descripción también proporciona las secuencias aisladas de ácido nucleico que son al menos de aproximadamente 12 bases de longitud (por ejemplo, al menos aproximadamente 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 40, 50, 60, 100, 250, 500, 750, 1000, 1500, 2000, 3000, 4000, ó

5000 bases de longitud) y se hibridan, bajo condiciones de hibridación, a la hebra en sentido o anti-sentido de un ácido nucleico que tiene la secuencia descrita en la SEQ ID No.: 11. Las condiciones de hibridación pueden ser condiciones de hibridación moderada o altamente exigentes.

Para fines de esta descripción, las condiciones de hibridación moderadamente exigentes significan que la hibridación es realizada aproximadamente a 42°C en una solución de hibridación que contiene KPO_4 25 mM (pH 7.4), 5X SSC, 5X de solución de Denhart, 50 µg/mL de ADN sonicado desnaturalizado de esperma de salmón, 50% de formamida, 10% de sulfato de dextrano, y 1-15 ng/mL de la sonda (aproximadamente 5 x 10^7 cpm/µg), mientras que los lavados son realizados aproximadamente a 50°C con una solución de lavado que contiene 2X SSC y 0.1% de dodecilsulfato de sodio.

Las condiciones de hibridación altamente exigentes significan que la hibridación es realizada aproximadamente a 42°C en una solución de hibridación que contiene KPO_4 25 mM (pH 7.4), 5X SSC, 5X solución de Denhart, 50 µg/mL de ADN de esperma de salmón, sonicado, desnaturalizado, 50% de formamida, 10% de sulfato de dextrano, y 1-15 ng/mL de la sonda (aproximadamente 5 x 10^7 cpm/µg), mientras que los lavados son realizados aproximadamente a 65°C con una solución de lavado que contiene 0.2X SSC y 0.1% de

dodecilsulfato de sodio.

Aislado: El término "aislado" como se utiliza en la presente, con referencia al ácido nucleico, se refiere a un ácido nucleico de origen natural que no está inmediatamente contiguo con ambas de las secuencias con las cuales éste está inmediatamente contiguo (uno sobre el extremo 5' y uno sobre el extremo 3') en el genoma de origen natural del organismo del cual se deriva. Por ejemplo, un ácido nucleico aislado puede ser, sin limitación, una molécula de ADN recombinante de cualquier longitud, con la condición de que una de las secuencias de ácido nucleico normalmente encontradas inmediatamente flanqueando esa molécula de ADN recombinante en un genoma de origen natural, es removida o es ausente. De este modo, un ácido nucleico aislado incluye, sin limitación un ADN recombinante que existe como una molécula separada (por ejemplo, un ADNc o un fragmento de ADN genómico producido mediante PCR o tratamiento con endonucleasa de restricción) independientemente de otra secuencias, así como el ADN recombinante que es incorporado dentro de un vector, un plásmido de replicación autónoma, un virus (por ejemplo un retrovirus, adenovirus o herpes virus), o dentro del ADN genómico de un procariote o eucariote. Además, un ácido nucleico aislado puede incluir una molécula de ADN recombinante que es parte de una secuencia de ácido nucleico híbrida o de fusión.

El término "aislado" como se utiliza en la presente con referencia al ácido nucleico, también incluye cualquier ácido nucleico de origen no natural, ya que las secuencias de ácido nucleico de origen no natural no son encontradas en la naturaleza y no tienen secuencias inmediatamente contiguas en un genoma de origen natural. Por ejemplo, el ácido nucleico de origen natural tal como un ácido nucleico manipulado por ingeniería genética es considerado como ácido nucleico aislado. El ácido nucleico manipulado por ingeniería genética puede ser elaborado utilizando clonación molecular común o técnicas de síntesis químicas de ácidos nucleicos. El ácido nucleico de origen no natural, aislado, puede estar independiente de otras secuencias, o estar incorporado dentro de un vector, un plásmido de replicación autónoma, un virus (por ejemplo, un retrovirus, adenovirus, o herpes virus) o el ADN genómico de un procariote o eucariote. Además, el ácido nucleico de origen no natural puede incluir una molécula de ácido nucleico que es parte de una secuencia de ácido nucleico híbrida o de fusión.

Será aparente para aquellos expertos en la técnica que un ácido nucleico existente entre los cientos a millones de otras moléculas de ácido nucleico, dentro, por ejemplo, de las bibliotecas de ADNc o bibliotecas genómicas, o rebanadas de gel que contienen una digestión de restricción de ADN genómico, no ha de ser considerado un ácido nucleico aislado.

Ácido nucleico: El término "ácido nucleico" como se utiliza en la presente abarca el ARN y el ADN incluyendo, sin limitación, ADNc, ADN genómico, y ADN sintético (por ejemplo, químicamente sintetizado). El ácido nucleico puede ser de doble hebra o de hebra simple. Donde es de hebra simple, el ácido nucleico puede ser la hebra en sentido o la hebra anti-sentido. Además, el ácido nucleico puede ser circular o lineal.

Operablemente enlazada: Una primera secuencia de ácido nucleico está "operablemente enlazada" con una segunda secuencia de ácido nucleico siempre que la primera secuencia de ácido nucleico sea colocado en una relación funcional con la segunda secuencia de ácido nucleico. Por ejemplo, un promotor está operablemente enlazado a una secuencia de codificación si el promotor afecta la transcripción de la secuencia de codificación. En general, las secuencias de ADN operablemente enlazada están contiguas y, donde sea necesaria unen dos regiones de codificación de polipéptidos, en la misma estructura de lectura.

Modificaciones de Péptido: La presente descripción incluye enzimas, así como las modalidades sintéticas de las mismas. Además, los análogos (moléculas orgánicas no peptídicas), derivados (moléculas peptídicas químicamente funcionalizadas obtenidas comenzando con las secuencias peptídicas descritas) y las variantes (homólogos) que tienen

la actividad enzimática deseada pueden ser utilizadas en los métodos descritos en la presente. Los péptidos descritos en la presente incluyen una secuencia de aminoácidos, que puede ser ya sea los L- y/o D-aminoácidos, de origen natural y de otro modo.

Los péptidos pueden ser modificados por una variedad de técnicas químicas para producir derivados peptídicos que tienen esencialmente la misma actividad que los péptidos no modificados y que tienen opcionalmente otras propiedades deseables. Por ejemplo, los grupos de ácidos carboxílico de la proteína, ya sea carboxilo-terminal o de cadena lateral, pueden ser proporcionados en la forma de una sal de un catión farmacéuticamente aceptable, o esterificados para formar un éster de 1 a 16 átomos de carbono, o convertidos a una amida de la fórmula NR1R2 en donde R1 y R2 son cada uno independientemente hidrógeno o alquilo de 1 a 16 átomos de carbono, o combinados para formar un anillo heterocíclico, tal como un anillo de 5 ó 6 miembros. Los grupos amino del péptido, ya sea amino-terminal o de cadena lateral, pueden estar en la forma de una sal por adición del ácido farmacéuticamente aceptable, tal como las sales orgánicas de los ácidos acético, benzoico, toluensulfónico, maleico, tartárico, y otras sales orgánicas, o pueden ser modificados a alquilo de 1 a 6 átomos de carbono o dialquilamino o convertidos posteriormente a una amida.

Los grupos hidroxilo de las cadenas laterales de los péptidos pueden ser convertidos a alcoxi de 1 a 16 átomos de carbono o un éster de 1 a 16 átomos de carbono utilizando técnicas bien reconocidas. Los anillos del fenilo y fenólicos de las cadenas laterales de los péptidos pueden ser sustituidos con uno o más átomos de halógeno, tales como flúor, cloro, bromo o yodo, o con alquilo de 1 a 16 átomos de carbono, alcoxi de 1 a 16 átomos de carbono, ácidos carboxílicos y ésteres de los mismos, o amidas de tales ácidos carboxílicos. Los grupos metileno de las cadenas laterales peptídicas pueden ser extendidos a alquilenos homólogos de 2 a 4 átomos de carbono. Los tioles pueden ser protegidos con cualquiera de un número de grupos protectores bien reconocidos tales como los grupos acetamida. Aquellos expertos en la técnica también reconocen los métodos para introducir estructuras cíclicas dentro de los péptidos de esta descripción para seleccionar y proporcionar constreñimientos conformacionales a la estructura, que dan como resultado estabilidad aumentada. Por ejemplo, una cisteína C- o N-terminal puede ser agregada al péptido, de modo que cuando se oxida el péptido contendrá un enlace disulfuro, generando un péptido cíclico. Otros métodos de ciclización de péptidos incluyen la formación de tioéteres y amidas y ésteres carboxilo- y amino-terminales.

Las modalidades peptidomiméticas y organomiméticas

148

están también dentro del alcance de la presente descripción, con lo cual el arreglo tridimensional de los constituyentes químicos de tales péptidos- y organomiméticos imitan el arreglo tridimensional de la cadena principal peptídica y las cadenas laterales de los aminoácidos componentes, dando como resultado tales péptidos- y organomiméticos de las proteínas de esta descripción, que tienen actividad enzimática detectable. Para aplicaciones de modelación en computadora, un farmacóforo es una definición tridimensional idealizada de los requerimientos estructurales para la actividad biológica. Los péptidos- y organomiméticos pueden ser diseñados para ajustar cada farmacóforo con el software de modulación por computadora actual (utilizando el diseño de fármacos ayudado por computadora o CADD). Ver Walters, "Computer-Assisted Modeling of Drugs", en Klegerman & Gorves (eds), Pharmaceutical Biotechnology, 1993, Interpharm Press: Buffalo Grove, IL, pp. 165-74 y Ch. 102 en Munson (ed), Principles of Pharmacology, 1995, Chapman & Hall, para descripciones de las técnicas utilizadas en CADD. También incluidos dentro del alcance de las descripciones están los miméticos preparados utilizando tales técnicas. En un ejemplo, un mimético imita la actividad enzimática generada por una enzima o una variante, fragmento o fusión de la misma.

Sondas y cebadores: Las sondas y cebadores de ácido nucleico pueden ser preparadas fácilmente con base en las

secuencias de aminoácidos y en las secuencias de ácido nucleico proporcionadas en la presente. Una "sonda" incluye un ácido nucleico aislado que contiene un marcador detectable o una molécula reportera. Los marcadores ejemplares incluyen, pero no están limitados a, isótopos radioactivos, ligandos, agentes quimioluminiscentes, y polipéptidos. Los métodos para la marcación y guía en la elección de los marcadores apropiados para diversos propósitos en, por ejemplo, Sambrook et al. (ed.), Molecular Cloning: A Laboratory Manual 2ª.ed., vol. 1-3, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989, y Ausubel et al (ed) Current Protocols in Molecular Biology, Greene Publishing and Wiley-Interscience, Nueva York (con actualizaciones periódicas), 1987.

Los "cebadores" son típicamente moléculas de ácido nucleico que tienen diez o más nucleótidos (por ejemplo, moléculas de ácido nucleico que tienen entre aproximadamente 10 nucleótidos y aproximadamente 100 nucleótidos). Un cebador puede ser recocido a una hebra complementaria de ácido nucleico objetivo mediante hibridación de ácido nucleico para formar un híbrido entre el cebador y la hebra de ácido nucleico objetivo y luego extendido a lo largo de la hebra de ácido nucleico objetivo mediante, por ejemplo, una ADN-polimerasa polipeptídica. Los pares de cebadores pueden ser utilizados para la amplificación de una secuencia de ácido nucleico, por ejemplo, por la reacción en cadena de

polimerasa (PCR) y otros métodos de amplificación de ácido nucleico conocidos en la técnica.

Los métodos para preparar y utilizar las sondas y los cebadores se describen, por ejemplo, en las referencias tales como Sambrook et al. (ed), Molecular Cloning: A Laboratory Manual, 2ª.ed., vol. 1-3, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989; Ausubel et al. (ed), Current Protocols in Molecular Biology, Greene Publishing and Wiley-Interscience, Nueva York (con actualizaciones periódicas), 1987; e Innis et al. (eds.), PCR Protocols: A Guide to Methods and Applications, Academic Press: San Diego, 1990. Los pares de cebadores de PCR pueden ser derivados de una secuencia conocida, por ejemplo, mediante le uso de programas de computadoras destinados para ese propósito tales como Primer (Versión 0.5, © 1991, Whitehead Institute for Biomedical Research, Cambridge, Mass). Una persona de experiencia en la técnica apreciará que la especificidad de una sonda o cebador particular se incrementa con la longitud, pero que una sonda o cebador puede estar en el intervalo de tamaño de una secuencia de longitud completa hasta secuencias tan cortas como de cinco nucleótidos consecutivos. De este modo, por ejemplo, un cebador de 20 nucleótidos consecutivos puede recocerse a un objetivo con una mayor especificidad que un cebador correspondiente únicamente de 15 nucleótidos. De este modo, con el fin de obtener mayor especificidad, las

sondas y cebadores pueden ser seleccionados los cuales comprenden, por ejemplo, 10, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1250, 1300, 1350, 1400, 1450, 1500, 1550, 1600, 1650, 1700, 1750, 1800, 1850, 1900, 2000, 2050, 2100, 2150, 2200, 2250, 2300, 2350, 2400, 2450, 2500, 2550, 2600, 2650, 2700, 2750, 2800, 2850, 2900, 3000, 3050, 3100, 3150, 3200, 3250, 3300, 3350, 3400, 3450, 3500, 3550, 3600, 3650, 3700, 3750, 3800, 3850, 3900, 4000, 4050, 4100, 4150, 4200, 4250, 4300, 4350, 4400, 4450, 4500, 4550, 4600, 4650, 4700, 4750, 4800, 4850, 4900, 5000, 5050, 5100, 5150, 5200, 5250, 5300, 5350, 5400, 5450, o más nucleótidos consecutivos.

Promotor: Un arreglo de secuencias de control de ácido nucleico que dirigen la transcripción de ácido nucleico. Un promotor incluye las secuencias necesarias de ácido nucleico cercanas al sitio de inicio de la transcripción, tales como, en el caso de un promotor tipo polimerasa II, un elemento TATA. Un promotor puede incluir el aumentador distal o los elementos represores que pueden ser localizados tanto como a varios miles de pares de bases del sitio de inicio de la transcripción.

Purificado: El término "purificado(a)" como se utiliza en la presente no requiere pureza absoluta; más bien, se pretende que éste sea un término relativo. De este modo, por

ejemplo, una preparación purificada de polipéptido o de ácido nucleico puede ser una en la cual el polipéptido o ácido nucleido de interés, respectivamente, está a una mayor concentración que el polipéptido o ácido nucleico estaría en su ambiente natural dentro de un organismo. Por ejemplo, una preparación polipeptídica puede ser considerada purificada si el contenido de polipéptido en la preparación representa al menos 50%, 60%, 70%, 80%, 85%, 90%, 92%, 95%, 98% ó 99% del contenido total de proteína de la preparación.

Recombinante: Un ácido nucleico "recombinante" es uno que tiene (1) una secuencia que no es de origen natural en el organismo en el cual se expresa o (2) una secuencia elaborada mediante una combinación artificial de dos secuencias más cortas, de otro modo separadas. Esta combinación artificial es frecuentemente lograda mediante síntesis química o más comúnmente, por manipulación artificial de los segmentos aislados de ácidos nucleicos, por ejemplo, por técnicas de ingeniería genética. "Recombinante" se utiliza también para describir las moléculas de ácido nucleico que han sido artificialmente manipuladas, pero que contienen las mismas secuencias reguladoras y las regiones de codificación que son encontradas en el organismo a partir del cual se aisló el ácido nucleico.

Identidad Secuencial: La similitud entre las secuencias de aminoácidos es expresada en términos de la

similitud entre las secuencias, de otro modo denominada como la identidad secuencial. La identidad secuencial es frecuentemente medida en términos de la identidad porcentual (o similitud u homología); entre mayor es el porcentaje, más similares son las dos secuencias. Los homólogos o variantes de un péptido, tal como SEQ ID No.: 12, poseen un grado relativamente alto de identidad secuencial cuando se alinean utilizando métodos estándares.

Los métodos de alineamiento de las secuencias para comparación son bien conocidos en la técnica. Diversos programas y algoritmos de alineamiento se describe en: Smith y Waterman, *Adv. Appl. Math.* 2:482, 1981; Needleman y Munsch, *J. Mol. Biol.* 48:443-53, 1970; Pearson y Lipman, *Proc. Natl. Acad. Sci.* E.U.A. 85:2444-8, 1988; Higgins y Sharp, *Gene* 73:237-44, 1988; Higgins y Sharp, *CABIOS* 5:151-3, 1989; Corpet et al., *Nucleic Acids Research* 16:10881-90, 1988; y Altschul et al., *Nature Genet.* 6:119-29, 1994.

La Herramienta de Búsqueda de Alineación Local Básica de NCBI (BLAST[MR]) (Altschul et al., *J. Mol. Biol.* 215:403-10, 1990) está disponible de varias fuentes, incluyendo el Centro Nacional para la Información en Biotecnología (NCI, Bethesda, MD) y en la Internet, para el uso en conexión con los programas de análisis secuencial, blastp, blastn, blastx, tbalstn y tblastx.

Las variantes de un péptido son típicamente

caracterizadas por posesión de al menos 50% de identidad secuencial contada sobre el alineamiento de longitud completa con la secuencia de aminoácidos utilizando NCBI Blast 2.0, blastp con espacios vacíos, ajustado a los parámetros por omisión. Para comparaciones de las secuencias de aminoácidos de más de 30 aminoácidos, la función de secuencias Blast 2 es empleada utilizando la matriz BLOSUM62 por omisión, ajustada a los parámetros por emisión (existencia de espacio vacío costo de 11, y por un espacio vacío de residuo costo de 1). Cuando se alinean los péptidos cortos (menos de alrededor de 30 aminoácidos) el alineamiento es realizado utilizando la función de secuencias Blast 2, empleando la matriz PAM30 ajustada a los parámetros por omisión (penalidades de 9 por espacio vacío abierto, de 1 por espacio vacío de extensión). Las proteínas con similitud aún mayor a las secuencias de referencia mostrarán identidades porcentuales cada vez mayores cuando se evalúen mediante este método, tal como al menos 80%, al menos 90%, al menos 95%, al menos 98%, o incluso al menos 99% de identidad secuencial. Cuando menos de la secuencia completa está siendo comparada para la identidad secuencial, los homólogos y variantes típicamente poseerán al menos 80% de identidad secuencial sobre ventanas cortas de 10-20 aminoácidos, y pueden poseer identidades secuenciales de al menos 85%, al menos 90%, al menos 95% ó 98%, dependiendo de su similitud a la secuencia de referencia. Los

155

métodos para determinar la identidad secuencial sobre tales ventanas cortas, se describen en el sitio de la red que es mantenido por el Centro Nacional para la Información en Biotecnología en Bethesda, Maryland. Una persona de experiencia en la técnica apreciará que estos intervalos de identidad secuencial son proporcionados para guía únicamente; es completamente posible que podrían ser obtenidos homólogos fuertemente significativos que caigan fuera de los intervalos proporcionados.

Pueden ser utilizados métodos similares para determinar la identidad porcentual de la secuencia de una secuencia de ácido nucleido. En un ejemplo particular, una secuencia homóloga es alineada a una secuencia nativa, y el número de concordancias es determinado por el conteo de número de posiciones donde un nucleótido o residuo de aminoácido idéntico es presentado en ambas frecuencias. La identidad secuencial porcentual es determinada al dividir el número de concordancias ya sea entre la longitud de la secuencia descrita en la secuencia identificada (por ejemplo, SEQ ID No.: 11), o por una longitud articulada (por ejemplo, 100 nucleótidos consecutivos o residuos de aminoácidos provenientes de una secuencia descrita en una secuencia identificada), seguido por la multiplicación del valor resultante por 100. Por ejemplo, una secuencia de ácidos nucleicos que tienen 1166 concordancias cuando se alinea con

la secuencia descrita en la SEQ ID No.: 11 es 75.0 por ciento idéntica a la secuencia descrita en la SEQ ID No.: 11 (por ejemplo, 1166+1554*100=75.0). Se nota que el valor de identidad secuencial porcentual es redondeado al décimo más cercano. Por ejemplo, 75.11, 75.12, 75.13 y 75.14 se redondean a 75.1, mientras que 75.15, 75.16, 75.17, 75.18 y 75.19 se redondean hasta 75.2. Se nota también que el valor de longitud será siempre un número entero. En otro ejemplo más, una secuencia objetivo que contiene una región de 20 nucleótidos que se alinea con 20 nucleótidos consecutivos provenientes de una secuencia identificada como sigue, contiene una región que comparte 75% de identidad secuencial a aquella secuencia identificada (por ejemplo, 15+20*100-75).

	1	20
Secuencia objetivo:	AGGTCGTGTACTGTCAGTCA	
	| || ||| |||| |||| |	
Secuencia de identidad:	ACGTGGTGAACTGCCAGTGA	

Agente de enlace específico: Un agente que es capaz de enlazarse específicamente a cualquier polipéptido descrito en la presente. Los ejemplos incluyen, pero no están limitados a, anticuerpos policlonales, anticuerpos monoclonales (incluyendo anticuerpos monoclonales humanizados), y fragmentos de anticuerpos monoclonales tales como Fab, $F(ab')_2$, y Fv así como cualquier agente capaz de enlazarse

específicamente a un epitopo de tales polipéptidos.

Los anticuerpos para los polipéptidos proporcionados en la presente (o fragmentos, variantes o fusiones de los mismos) pueden ser utilizados para purificar o identificar tales polipéptidos. Las secuencias de aminoácidos y de ácidos nucleicos proporcionadas en la presente permiten la producción de agentes de enlace basados en anticuerpo específico, que reconocen los polipéptidos descritos en la presente.

Los anticuerpos monoclonales o policlonales pueden ser producidos para los polipéptidos, porciones de los polipéptidos o variantes de los mismos. Óptimamente, los anticuerpos producidos contra uno o más epitopos sobre un antígeno polinucleotídico detectarán específicamente ese polipéptido. Es decir, los anticuerpos producidos contra un polipéptido particular, podrían reconocer y enlazarse a ese polipéptido particular, y podrían sustancialmente no reconocer o no enlazarse a otros polipéptidos. La determinación de que un anticuerpo se enlaza específicamente a un polipéptido particular, es realizada mediante cualquiera de un número de métodos de inmunoensayos estándares; por ejemplo, transferencia de Western (Ver, por ejemplo, Sambrook et al. (ed.), Molecular Cloning: A Laboratory Manual, 2ª.ed., vol. 1-3, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989).

Para determinar que una preparación dada de anticuerpo (tal como una preparación producida en un ratón contra un polipéptido que tiene la secuencia de aminoácidos descrita en la SEQ ID No.: 12) detecta específicamente el polipéptido apropiado (por ejemplo, un polipéptido que tiene la secuencia de aminoácidos descrita en la SEQ ID No.: 12) mediante transferencia de Western, la proteína celular total puede ser extraída de las células y separada mediante electroforesis en gel de SDS-poliacrilamida.

La proteína celular total separada puede ser luego transferida a una membrana (por ejemplo, nitrocelulosa), y la preparación de anticuerpo incubada con la membrana. Después de lavar la membrana para eliminar los anticuerpos no específicamente enlazados, la presencia de los anticuerpos específicamente enlazados puede ser detectada utilizando un anticuerpo secundario apropiado (por ejemplo, un anticuerpo anti-ratón) conjugado a un polipéptido tal como la fosfatasa alcalina (ya que la aplicación de fosfato de 5-bromo-4-cloro-3-indolilo/nitro azul de tetrazolio da como resultado la producción de un compuesto densamente coloreado de azul por la fosfatasa alcalina inmunolocalizada.

Los polipéptidos sustancialmente puros, adecuados para el uso como un inmunógeno pueden ser obtenidos de las células transfectadas, las células transformadas o las células de tipo silvestre. Las concentraciones de polipéptidos en la

preparación final pueden ser ajustadas, por ejemplo, mediante concentración sobre un dispositivo de filtro Amicon, al nivel de unos pocos microgramos por mililitro. Además, los polipéptidos en el intervalo de tamaño de los polipéptidos de longitud completa hasta polipéptidos que tienen tan pocos como nueve residuos de aminoácidos, pueden ser utilizados como inmunógenos. Tales como polipéptidos pueden ser producidos en el cultivo celular, pueden ser sintetizados químicamente utilizando métodos estándares, o pueden ser obtenidos mediante la escisión de los polipéptidos grandes en polipéptidos más pequeños que pueden ser purificados. Los polipéptidos que tienen tan pocos como nueve residuos de aminoácidos de longitud pueden ser inmunogénicos cuando son presentados a un sistema inmune en el contexto de una molécula de Complejo Mayor de Histocompatibilidad (MHC) tal como la molécula MHC de la clase I o MHC de la clase II. En consecuencia, los polipéptidos que tienen al menos, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 900, 1000, 1050, 1100, 1150, 1200, 1250, 1300, 1350, o más residuos de aminoácidos consecutivos de cualquier secuencia de aminoácidos descrita en la presente pueden ser utilizados como inmunógenos para producir anticuerpos.

Los anticuerpos monoclonales para cualquiera de los

polipéptidos descritos en la presente, pueden ser preparados a partir de hibridomas murinos de acuerdo al método clásico de Kohler y Milstein (Nature 256:495-7, 1975) o un método derivado del mismo.

El antisuero policlonal que contienen anticuerpos para los epitopos heterogéneos de cualquier polipéptido descrito en la presente, puede ser preparado mediante la inmunización de animales adecuados con el polipéptido (o fragmento del mismo), que puede ser no modificado o modificado para aumentar la inmunogenicidad. Un protocolo efectivo de inmunización para conejos puede ser encontrado en Vaitukaitis et al. (*J. Clin. Endocrinol. Metab.* 33:988-91, 1971).

Se puede utilizar fragmentos de anticuerpo en lugar de anticuerpos completos, y pueden ser fácilmente expresados en células hospederas procarióticas. Los métodos para elaborar y utilizar porciones inmunológicamente efectivas de anticuerpos monoclonales, también denominadas como "fragmentos de anticuerpo" son bien conocidos e incluyen aquellos descritos en Better y Horowitz Metods Enzymol. 178:476-96, 1989), Glockshuber et al., (*Biochemistry* 29:1362-7, 1990), Patente de los Estados Unidos No. 5,648,237 ("Expresión de Fragmentos Funcionales de Anticuerpo"), Patente de los Estados Unidos No. 4,946,778 ("moléculas que se Enlazan a la Cadena Polipeptídica Simple"), Patente de los Estados Unidos No. 5,455,030 ("inmunoterapia Utilizando Moléculas que se Enlazan

161

a Polipéptidos de Cadena Simple"), y referencias citadas éstas.

Transformada: Una célula "transformada" es una célula dentro de la cual ha sido introducida una molécula de ácido nucleïco, por ejemplo, mediante técnicas de biología molecular. La transformación abarca todas las técnicas mediante las cuales una molécula de ácido nucleico pueda ser introducida dentro de tal célula incluyendo, sin limitación, la transfección con un vector viral, la conjugación, la transformación con un vector plasmídico, y la introducción de ADN desnudo mediante electroporación, lipofección y aceleración de pistolas de partículas.

Variantes, fragmentos o proteínas de fusión: Las proteínas descritas, incluyen variantes, fragmentos y fusiones de las mismas. Las secuencias de ADN (por ejemplo SEQ ID No.: 11) que codifican para una proteína (por ejemplo SEQ ID No.: 12), la proteína de fusión o un fragmento o variante de una proteína, pueden ser manipuladas mediante ingeniería genética para permitir que la proteína sea expresada en células eucarióticas, bacterias, insectos y/o plantas. Para obtener la expresión, la secuencia de ADN puede ser alterada y operablemente enlazada a otras secuencias reguladoras. El producto final, el cual contiene las secuencias reguladoras y la proteína, es denominado como un vector. Este vector puede ser introducido dentro de células

eucarióticas, bacterianas, de insecto y/o células vegetales. Una vez dentro de la célula el vector permite que la proteína sea producida.

Una proteína de fusión que incluye una proteína, tal como una triptofano-aminotransferasa (o variante, polimorfismo, mutante o fragmento de la misma), por ejemplo, la SEQ ID No.: 12, enlazada a otras secuencias de aminoácidos que no inhiben la actividad deseada de la proteína, por ejemplo la habilidad para convertir el triptofano a indol-3-piruvato. En un ejemplo, las otras secuencias de aminoácidos no son mayores de aproximadamente 10, 12, 15, 20, 25, 30 ó 50 aminoácidos de longitud.

Una persona de experiencia ordinaria en la técnica apreciará que una secuencia de ADN puede ser alterada de numerosas formas sin afectar la actividad biológica de la proteína conjugada. Por ejemplo, puede ser utilizada PCR para producir variaciones en la secuencia de ADN que codifica para una proteína. Tales variantes pueden ser variantes optimizadas para preferencia de codón en una célula hospedera utilizada para expresar la proteína, u otros cambios en la secuencia que faciliten la expresión.

Vector: Una molécula de ácido nucleico como es introducida dentro de una célula, con lo cual se produce una célula transformada. Un vector puede incluir secuencias de ácido nucleico que le permitan replicarse dentro de la

célula, tal como un origen de replicación. Un vector puede también incluir uno o más genes marcadores seleccionables y otros elementos genéticos conocidos en la técnica.

Panorama General de la Vías Biosintéticas

Como se muestra en las FIGS. 1-3 y 11-13, pueden ser utilizadas muchas vías biosintéticas para producir la monatina o sus intermediarios tales como el indol-3-piruvato o el MP. Para la conversión, de cada sustrato (glucosa, triptofano, ácido indol-3-acético, indol-3-piruvato, y MP) a cada producto (triptofano, indol-3-piruvato, MP y monatina) pueden ser utilizados varios polipéptidos diferentes. Además, estas reacciones pueden ser llevadas a cabo *in vivo*, *in vitro* o a través de una combinación de reacciones *in vivo* y reacciones *in vitro* tales como las reacciones *in vitro* que incluyen las reacciones químicas no enzimáticas. Por lo tanto, las FIGS. 1-3 y 11-13 son ejemplares y muestran diferentes y múltiples vías que pueden ser utilizadas para obtener los productos deseados.

Glucosa a Triptofano

Muchos organismos pueden sintetizar triptofano a partir de glucosa. La o las construcciones que contienen el o los genes necesarios para producir monatina, MP, y/o indol-3-piruvato a partir de la glucosa y/o el triptofano, pueden ser clonadas dentro de tales organismos. Se muestra en la presente que el triptofano puede ser convertido a monatina.

En otros ejemplos, un organismo es manipulado mediante ingeniería genética utilizando los polipéptidos conocidos para producir triptofano, o sobreproducir triptofano. Por ejemplo, la Patente de los Estados Unidos No. 4,371,614 (incorporada por referencia en la presente) describe una cepa de *E. coli* transformada con un plásmido que contiene un operón de triptofano de tipo silvestre.

Los títulos máximos del triptofano descritos en la Patente de los Estados Unidos No. 4,371,614 son de aproximadamente 230 ppm. Similarmente, el documento WO 8701130 (incorporado por referencia en la presente) describe una cepa de *E. coli* que ha sido genéticamente manipulada por ingeniería para producir triptofano y discute la producción fermentativa cada vez mayor del L-triptofano. Aquellos expertos en la técnica reconocerán que los organismos capaces de producir triptofano a partir de la glucosa son también capaces de utilizar otras fuentes de carbono y energía que pueden ser convertidos a glucosa o fructosa-6-fosfato, con resultados similares. Las fuentes de carbono y de energía ejemplares incluyen, pero no están limitadas a, sucrosa, fructosa, almidón, celulosa o glicerol.

Triptofano a Indol-3-piruvato

Se pueden utilizar varios polipéptidos para convertir el triptofano a indol-3-piruvato. Los polipéptidos ejemplares incluyen los miembros de las clases de enzimas (EC) 2.6.1.27,

1.4.1.19, 1.4.99.1, 2.6.1.28, 1.4.3.2, 1.4.3.3, 2.6.1.5, 2.6.1.-, 2.6.1.1. y 2.6.1.21. Estas clases incluyen los polipéptidos denominados triptofano-aminotransferasa (también denominada L-fenilalanina-2-oxoglutarato-aminotransferasa, triptofano-transaminasa, transaminasa 5-hidroxitriptofan-cetoglutárica, hidroxitriptofano-aminotransferasa, L-triptofano-aminotransferasa, L-triptofano-transaminasa, y L-triptofano-2-oxoglutarato-aminotransferasa) que convierte en L-triptofano y el 2-oxoglutarato a indol-3-piruvato y L-glutamato; la D-triptofano-aminotransferasa que convierte el D-triptofano y un 2-oxoácido a indol-3-piruvato y un aminoácido; la triptofano-deshidrogenasa (también denominada NAD(P)-L-triptofano-deshidrogenasa, L-triptofano-deshidrogenasa, L-Trp-deshidrogenasa, TDH y L-triptofano:NAD(P)oxidorreductasa (desaminación) que convierte el L-triptofano y el NAD(P) a indol-3-piruvato y NH_3 y NAD(P)H; la deshidrogenasa del D-aminoácido que convierte los D-aminoácidos y FAD a indol-3-piruvato y NH_3 y $FADH_2$; la triptofano-fenilpiruvato-transaminasa (también denominada L-triptofano-α-cetoisocaproato-aminotransferasa y L-triptofano-fenilpiruvato-aminotransferasa) que convierte el L-triptofano y el fenilpiruvato a indol-3-piruvato y L-fenilalanina; la oxidasa del L-aminoácido (también denominada como oxidasa de ofio-amino-ácido, y oxidorreductasa de L-aminoácido:oxígeno

(desaminación)) que convierte un L-aminoácido y H_2O y O_2 a un 2-oxoácido y NH_3 y H_2O_2; La oxidasa de D-aminoácido (también denominada como oxidasa de ofio-aminoácido y oxidorreductasa de D-aminoácido:oxígeno (desaminación)) que convierte un D-aminoácido y H_2O y O_2 a un 2-oxoácido y NH_3 y H_2O_2 y la triptofano-oxidasa que convierte el L-triptofano y el H_2O y O_2 a indol-3-piruvato y NH_3 y H_2O_2. Estas clases también contienen la tirosina(aromático)-aminotransferasa, aspartato-aminotransferasa, D-aminoácido-(o D-alanina)-aminotransferasa, y la aminotransferasa amplia (de sustrato múltiple) las cuales tienen múltiples actividades de aminotransferasa, algunas de las cuales pueden convertir el triptofano y un 2-oxoácido a indol-3-piruvato y un aminoácido.

Once miembros de la clase de aminotransferasa que tienen tal actividad son descritos más adelante en el Ejemplo 1, incluyendo una novedosa aminotransferasa mostrada en la SEQ ID Nos.: 11 y 12. Por lo tanto, esta descripción proporciona las secuencias aisladas de ácido nucleico y de proteína que tienen al menos 80%, al menos 85%, al menos 90%, al menos 95%, al menos 98%, o incluso al menos 99%, de identidad secuencial a la SEQ ID No.: 11 y 12. También abarcados por esta descripción están los fragmentos y las fusiones de la SEQ ID Nos.: 11 y 12 que conservan actividad de aminotransferasa o codifican para una proteína que tiene

actividad de aminotransferasa. Los fragmentos ejemplares incluyen, pero no están limitados al menos a 10, 12, 15, 20, 25, 50, 100, 200, 500 ó 100 nucleótidos contiguos de la SEQ ID No.: 11 o al menos 6, 10, 15, 20, 25, 50, 75, 100, 200, 300 ó 350 aminoácidos contiguos de la SEQ ID No.: 12. Las secuencias descritas (y las variantes, fragmentos y fusiones de las mismas) pueden ser parte de un vector. El vector puede ser utilizado para transformar células hospederas, con lo cual se producen células recombinantes que pueden producir indol-3-piruvato a partir de triptofano, y en algunos ejemplos pueden producir además MP y/o monatina.

Son conocidas las L-aminoácido-oxidasas (1.4.3.2) y las secuencias pueden ser aisladas a partir de varias diferentes fuentes, tales como *Vipera lebetine* (sp P81375), *Ophiophagus hannah* (sp P81383), *Agkistrodon rhodostoma* (sp P81382), *Crotalus atrox* (sp P56742), *Burkholderia cepacia*, *Arabidopsis thaliana*, *Caulobacter cresentus*, *Chlamydomonas renhardtii*, *Mus musculus*, *Pseudomonas syringae*, y *Rhodococcus str.* Además, las triptofano-oxidasa son descritas en la literatura y pueden ser aisladas, por ejemplo, de *Coprinus sp.* SF-1, la col china con enfermedad de raíz en forma de cachiporra, *Arabidopsis thaliana*, e hígado de mamífero. Un miembro de la clase de L-aminoácido-oxidasa que puede convertir el triptofano a indol-3-piruvato se discute más adelante en el Ejemplo 3, así como las fuentes alternativas

para la clonación molecular. Muchos genes de la D-aminoácido-oxidasa son disponibles en las bases de datos para la clonación molecular.

Son conocidas las triptofano-deshidrogenasas, y pueden ser aisladas por ejemplo, de la espinaca, *Pisum sativum*, *Prosopis juliflora*, chícharo, mesquite, trigo, maíz, tomate, tabaco, *Chromobacterium violaceum*, y lactobacilos. Son conocidas muchas secuencias del gen de la D-aminoácido-deshidrogenasa.

Como se muestra en las FIGS. 11-13, si se utiliza una aminoácido-oxidasa tal como la triptofano-oxidasa para convertir el triptofano a indol-3-piruvato, puede ser agregada catalasa para reducir o incluso eliminar la presencia del peróxido de hidrógeno.

Indol-3-lactato a Indol-3-piruvato

La reacción que convierte el indol-3-lactato a indol-3-piruvato puede ser catalizada por una variedad de polipéptidos, tales como los miembros de las clases 1.1.1.110, 1.1.1.27, 1.1.1.28, 1.1.2.3, 1.1.1.222, 1.1.1.237, 1.1.3.-, o 1.1.1.111. La clase 1.1.1.110 de los polipéptidos incluye las indol-lactato-deshidrogenasas (también denominado ácido indolacético:NAD^+-oxidorreductasa). Las clases 1.1.1.27, 1.1.1.28 y 1.1.2.3 incluyen las lactato-deshidrogenasas (también denominadas deshidrogenasas de ácido láctico, lactato:NAD^+-oxidorreductasa). La clase 1.1.1.222 que contiene

69

la (R)-4-hidroxifenil-lactato-deshidrogenasa (también denominada deshidrogenasa de lactato-D-aromático, deshidrogenasa de lactato-R-aromático y R-3-(4-hidroxifenil)lactato-NAD(P)$^+$-2-oxidorreductasa) y la clase 1.1.1.237 contiene 3-(4-hidroxifenilpiruvato)reductasa (también denominada como hidroxifenilpiruvato)reductasa y 4-hidroxifenil-lactato:NAD$^+$-oxidorreductasa). La clase 1.1.3.- contiene lactato-oxidasas y la clase 1.1.1.111 contiene (3-imidazol-5-il)lactato-deshidrogenasas (también denominadas (S)-3-(imidazol-5-il)lactato:NAD(P)$^+$-oxidorreductasa). Es probable que varios de los polipéptidos en estas clases permitan la producción del indol-3-piruvato a partir del ácido indol-2-láctico. Los ejemplos de esta conversión son proporcionados en el Ejemplo 2.

Las relaciones químicas pueden también ser utilizadas para convertir el ácido indol-3-láctico a indol-3-piruvato. Tales reacciones químicas incluyen un paso de oxidación que puede ser logrado utilizando varios métodos, por ejemplo: oxidación con aire utilizando el catalizador B2 (China Chemical Reporter, v. 13, no. 28, p. 18(1), 2002), perclorato y permanganato diluidos, o peróxido de hidrógeno en presencia de catalizadores metálicos.

Indol-3-piruvato a ácido 2-hidroxi-2-(indol-3-ilmetil)-4-cetoglutárico (MP)

Varios polipéptidos conocidos pueden ser utilizados

para convertir el indol-3-piruvato a MP. Las clases de polipéptidos ejemplares incluyen 4.1.3-, 4.1.3.16, 4.1.3.17 y 4.1.2.-. Estas clases incluyen las sintasas-liasas carbono-carbono, tales como las aldolasas que catalizan la condensación de dos sustratos de ácido carboxílico. Los péptidos de la clase EC 4.1.3- son las sintasas-liasas que forman enlaces carbono-carbono utilizando sustratos de oxoácido (tales como el indol-3-piruvato) como el electrófilo, mientras que EC 4.1.2- son sintasas/liasas que forman enlaces carbono-carbono utilizando sustratos de aldehído (tales como benzaldehído) como el electrófilo.

Por ejemplo, el polipéptido descrito en la Patente Europea EP 1045-029 (EC 4.1.3.16, 4-hidroxi-2-oxoglutarato-glioxilato-liasa también denominada 4-hidroxi-2-oxoglutarato-aldolasa, 2-oxo-4-hidroxiglutarato-aldolasa o KHG-aldolasa) convierte el ácido glioxílico y el piruvato al ácido 4-hidroxi-2-cetoglutárico y la polipéptido-4-hidroxi-metil-2-oxoglutarato-aldolasa (EC 4.1.3.17 también denominada como 4-hidroxi-4-metil-2-oxoglutarato-piruvato-liasa o aldolasa ProA), condensa dos cetoácidos tales como dos piruvatos a 4-hidroxi-4-metil-2-oxoglutarato. Las reacciones que utilizan estas liasas son descritas en la presente.

Las FIGS. 1-2 y 11-13 muestran diagramas esquemáticos de estas reacciones en las cuales una molécula de 3 átomos de carbono es combinada con el indol-3-piruvato. Muchos miembros

de EC 4.1.2- y 4.1.3-, particularmente polipéptidos que utilizan PLP, pueden utilizar moléculas de 3 átomos de carbono que son aminoácidos tales como serina, cisteína, y alanina, o derivadas de las mismas. Las condensaciones de aldol catalizadas por representantes de EC 4.1.2- y 4.1.3- requieren la molécula de tres átomos de carbono de esta vía para ser piruvato o un derivado de piruvato. No obstante, otros compuestos pueden servir como la fuente de 3 átomos de carbono y ser convertidos a piruvato. La alanina pueden ser transaminada por muchas transaminasas que utilizan PLP incluyendo muchas de aquellas mencionadas anteriormente, para producir piruvato. El piruvato y el amoniaco pueden ser obtenidos mediante reacciones de beta-eliminación (tales como aquellas catalizadas por la triptofanasa o β-tirosinasa) de L-serina, L-cisteína, y derivados de serina y de cisteína con suficientes grupos salientes, tales como O-metil-L-serina, O-bencil-L-serina, S-metilcisteína, S-bencilcisteína, S-alquil-L-cisteína, O-acil-L-serina y 3-cloro-L-alanina. El aspartato puede servir como una fuente de piruvato en las reacciones de beta-liasa mediadas por PLP tales como aquellas catalizadas por la triptofanasa (EC 4.1.99.1) y/o la β-tirosinasa (EC 4.1.99.2, también denominada tirosina-fenol-liasa). La velocidad de la reacción de beta-liasa puede ser incrementada al realizar la mutagénesis dirigida al sitio sobre los

172

polipéptidos (4.1.99.1-2) como se describe por Mouratou et al (*J. Biol. Chem.* 274:1320-5, 1999) y en el Ejemplo 8. Estas modificaciones permiten que los polipéptidos acepten sustratos de aminoácido dicarboxílico. El lactato puede también servir como una fuente de piruvato, y es oxidado a piruvato por la adición de lactato-deshidrogenasa y un cofactor oxidado o lactato-oxidasa y oxígeno. Los ejemplos de estas reacciones se describen más adelante. Por ejemplo, como se muestra en la FIG. 2 y en las FIGS. 11-13, la aldolasa ProA puede ser puesta en contacto con el indol-3-piruvato cuando el piruvato es utilizado como la molécula de 3 átomos de carbono.

El MP puede también ser generado utilizando reacciones químicas, tales como las condensaciones de aldol proporcionadas en el Ejemplo 5.

MP a Monatina

La conversión de MP a monatina puede ser catalizada por uno o más de: triptofano-aminotransferasas (2.6.1.27), triptofano-deshidrogenasas (1.4.1.19), D-aminoácido-deshidrogenasas (1.4.99.1), glutamato-dehidrogenasas (1.4.1.2-4), fenilalanina-deshidrogenasa (EC 1.4.1.20), triptofano-fenilpiruvato-transaminasas (2.6.1.28), o más en general miembros de la familia de aminotransferasas (2.6.1.-) tales como la aspartato-aminotransferasa (EC 2.6.1.1), tirosina(aromático)aminotransferasa (2.6.1.5), D-triptofano-

aminotransferasa o D-alanina (2.6.1.21) aminotransferasa (FIG. 2). Once miembros de la clase de las aminotransferasas se describen en seguida (Ejemplo 1), incluyendo un nuevo miembro de la clase mostrada en la SEQ ID Nos.: 11 y 12, y las reacciones que demuestran la actividad de la aminotransferasa y de las enzimas deshidrogenasas, se proporcionan en el Ejemplo 7.

Esta reacción puede ser también realizada utilizando reacciones químicas. La aminación del cetoácido (MP) es realizada por aminación reductiva utilizando amoniaco y cianoborohidruro de sodio.

Las FIGS. 11-13 muestran polipéptidos adicionales que pueden ser utilizados para convertir MP a monatina, así como la provisión de rendimientos incrementados de monatina a partir de indol-3-piruvato o triptofano. Por ejemplo, si se utiliza aspartato como entrenador de amino, la aspartato-aminotransferasa puede ser utilizada para convertir en el aspartato oxaloacetato (FIG. 11). El oxaloacetato es convertido a piruvato y dióxido de carbono por una descarboxilasa, tal como oxaloacetato de descarboxilasa (FIG. 11). Además, si se utiliza lisina como el donador de amino, la lisina-epsilon-aminotransferasa puede ser utilizada para convertir la lisina a alisina (FIG. 12). La alisina es espontáneamente convertida a 1-piperideina-6-carboxilato (FIG. 12). Si un polipéptido capaz de catalizar las

reacciones de aminación reductiva (por ejemplo, glutamato-deshidrogenasa) se utiliza para convertir MP a monatina, un polipéptido que puede reciclar NAD(P)H y/o producir un producto volátil (FIG. 13) puede ser utilizado, tal como la formiato-deshidrogenasa.

Consideraciones Adicionales en el Diseño de la Vías Biosintéticas

Dependiente de cuáles polipéptidos sean utilizados para generar el indol-3-piruvato, MP y/o monatina, cofactores, sustratos y/o polipéptidos adicionales pueden ser proporcionados para que la célula productora aumente la formación del producto.

Eliminación de peróxido de hidrógeno

El peróxido de hidrógeno (H_2O_2) es un producto que, si es generado, puede ser tóxico para las células productoras y puede dañar a los polipéptidos o los intermediarios producidos. La L-aminoácido-oxidasa descrita anteriormente genera H_2O_2 como un producto. Por lo tanto, si la L-aminoácido-oxidasa es utilizado, el H_2O_2 resultante puede ser eliminado o sus niveles disminuidos para disminuir el daño potencial a la célula o al producto.

Las catalasas pueden ser utilizadas para reducir el nivel de H_2O_2 en la célula (FIGS. 11-13). La célula productora puede expresar un gen o secuencia de ADNc que codifica para una catalasa (EC 1.11.1.6), que cataliza la descomposición

del peróxido de hidrógeno a agua y oxígeno gaseoso. Por ejemplo, una catalasa puede ser expresada a partir de un vector transfectado en la célula productora. Los ejemplos de catalasas que pueden ser utilizadas incluyen, pero no están limitadas a: tr/Q9EV50 (*Staphylococcus xylosus*), tr/Q9KBE8 (*Bacillus halodurans*), tr/Q9URJ7 (*Candida albicans*), tr/P77948 (*Streptomyces coelicolor*), tr/Q9RBJ5 (*Xanthomomas campestris*) (SwissProt Accesos Nos.). Los reactores biocatalíticos que utilizan la L-aminoácido-oxidasa, D-aminoácido-oxidasa, o triptofano-oxidasa, pueden también contener un polipéptido de catalasa.

Modulación de la Disponibilidad de PLP (piridoxal-5'-fosfato)

Como se muestra en la FIG. 1, el PLP puede ser utilizado en uno o más de los pasos biosintéticos descritos en la presente. La concentración de PLP puede ser suplementada de modo que PLP no se vuelva una limitación sobre la eficiencia total de la reacción.

La vía biosintética para vitamina B_6 (la precursora de PLP) ha sido completamente estudiada en *E. coli* y algunas de las proteínas han sido cristalizadas (Laber et al., *FEBS Letter*, 449:45-8, 1999). Dos de los genes (epd o gapB y serC) son requeridos en otras vías metabólicas, mientras que tres genes (*pdxA*, *pdxB*, y *pdxj*) son únicos para la biosíntesis del fosfato de piridoxal. Uno de los materiales iniciales en la vía de *E. coli* es el 1-desoxi-D-xilulosa-5-fosfato (DXP). La

síntesis de este precursor a partir de metabolitos central comunes de 2 y 3 átomos de carbono, es catalizada por el polipéptido 1-desoxi-D-xilulosa-5-fosfato-sintasa (DSX). El otro precursor es un derivado de treonina formado a partir del azúcar de 4 átomos de carbono, del D-eritrosa-4-fosfato. Los genes requeridos para la conversión a fosfo-4-hidroxil-L-treonina (HTP) son *epd*, *pdxB*, y *serC*. La última reacción para la formación de PLP es una condensación intramolecular compleja y una reacción de cierre de anillo entre DXP y HTP, catalizada por los productos génicos de *pdxA* y *pdxJ*.

Si PLP se vuelve un nutriente limitante durante la fermentación para producir monatina, la expresión incrementada de uno o más de los genes de la vía en una célula hospedera productora, puede ser utilizada para incrementar el rendimiento de la monatina. Un organismo hospedero puede contener múltiples copias de sus genes de la vía nativa o copias de los genes de la vía no nativa pueden ser incorporados dentro del genoma del organismo. Adicionalmente, múltiples copias de los genes de la vía salvaje pueden ser clonados dentro del organismo hospedero.

Una vía salvaje que es conservada en todos los organismos, recicla los diversos derivados de la vitamina B_6 a la forma PLP activa. Los polipéptidos involucrados en esta vía son la cinasa *pdxK*, la oxidasa *pdxH*, y la cinasa *pdxY*. La sobre-expresión de una o más de estos genes puede incrementar

la disponibilidad de PLP.

Los niveles de la vitamina B_6 pueden ser elevados por la eliminación o represión de la regulación metabólica de los genes de la vía biosintética nativa en el organismo hospedero. PLP reprime los polipéptidos involucrados en la biosíntesis del derivado precursor de treonina en la bacteria *Flavobacterium sp.* Cepa 238-7. Esta cepa bacteriana, liberada del control metabólico, sobreproduce derivados de piridoxal y puede excretar hasta 20 mg/L de PLP. La manipulación genética del organismo hospedero que produce monatina de una manera similar permitirá la producción incrementada de PLP sin sobre-expresión de los genes de la vía biosintética.

Utilización de Amonio

Las reacciones de triptofanasa pueden ser impulsadas hacia la dirección sintética (producción de triptofano a partir de indol) al hacer el amoniaco más disponible o mediante eliminación del agua. Las reacciones de aminación reductiva tales como aquellas catalizadas por la glutamato-deshidrogenasa, pueden también ser impulsadas hacia delante por un exceso de amoniaco.

El amoniaco puede ser hecho disponible como una sal de carbamato de amonio o fosfato de amonio en un sistema amortiguado con carbonato o con fosfato. El amoniaco puede también ser proporcionado como piruvato de amonio o formiato de amonio. Alternativamente, el amoniaco puede ser

suministrado si la reacción es acoplada con una reacción que genere amoniaco, tal como la glutamato-deshidrogenasa o triptofano-deshidrogenasa. El amoniaco puede ser generado por la adición de los sustratos naturales de EC 4.1.99- (tirosina o triptofano), que serán hidrolizados a fenol o indol, piruvato y NH_3. Esto permite también un rendimiento incrementado del producto sintético sobre la cantidad normal en equilibrio al permitir que la enzima hidrolice su sustrato preferido.

Eliminación de productos y subproductos

La conversión de triptofano a indol-3-piruvato vía un triptofano-aminotransferasa puede afectar de manera adversa la velocidad de producción del indol-3-piruvato debido a que la reacción produce glutamato y requiere el co-sustrato 2-oxoglutarato (α-cetoglutarato). El glutamato puede provocar la inhibición de la aminotransferasa, y la reacción consumirá grandes cantidades del co-sustrato. Además, las altas concentraciones de glutamato son dañinas para los procesos de separación corriente abajo (3').

El polipéptido-glutamato-deshidrogenasa (GLDH) convierte el glutamato a 2-oxoglutarato, con lo cual se recicla el co-sustrato en la reacción catalizada por la triptofano-aminotransferasa. GLDH también genera equivalentes reductores (NADH o NADPH) que pueden ser utilizados para

generar energía para la célula (ATP) bajo condiciones aeróbicas. La utilización del glutamato por GLDH también reduce la formación de subproducto. Adicionalmente, la reacción genera amoniaco, que puede servir como una fuente de nitrógeno para la célula o como un sustrato en una aminación reductora para el paso final mostrado en la FIG. 1. Por lo tanto, una célula productora que sobre-expresa un polipéptido GLDH puede ser utilizada para incrementar el rendimiento y reducir el costo de los medios y/o los procesos de separación.

En la vía del triptofano a la monatina, el donador del grupo amino del paso tres (por ejemplo, glutamato o aspartato) puede ser convertido nuevamente al aceptor de amino requerido para el paso 1 (por ejemplo, 2-oxo-glutarato u oxaloacetato), si se utiliza una aminotransferasa proveniente de las clases de enzimas apropiadas. La utilización de dos transaminasas separadas para esta vía, en la cual el sustrato de una transaminasa no inhibe competitivamente la actividad de la otra transaminasa, puede incrementar la eficiencia de esta vía.

Muchas de las reacciones en las vías descritas son reversibles y, por lo tanto, alcanzarán un equilibrio entre los sustratos y los productos. El rendimiento de la vía puede ser incrementado por la eliminación continua de los productos a partir de los polipéptidos. Por ejemplo, la secreción de

monatina hacia el caldo de fermentación utilizando una permeasa u otra proteína de transporte, o la cristalización selectiva de la monatina a partir de una corriente del reactor biocatalítico con el reciclamiento concomitante de los sustratos, incrementará el rendimiento de la reacción.

La eliminación de los subproductos por reacciones enzimáticas adicionales o por sustitución de los grupos donadores de amino es otra manera más para incrementar el rendimiento de la reacción. Son discutidos varios ejemplos en el Ejemplo 13 y mostrados en las FIGS. 11-13. Idealmente, es producido un sub-producto que no es disponible para reaccionar en la dirección inversa, ya sea por cambio de fase (evaporación) o por conversión espontánea a un producto final no reactivo, tal como dióxido de carbono.

Modulación de los Combinados de Sustrato

El combinado de indol puede ser modulado al incrementar la producción de los precursores de triptofano y/o por alteración de las vías catabólicas que involucran el indol-3-piruvato y/o el triptofano. Por ejemplo, la producción de ácido indol-3-acético a partir de indol-3-piruvato puede ser reducida o eliminada al suprimir funcionalmente el gen que codifica para EC 4.1.74.1 en la célula hospedera. La producción de indol a partir de triptofano puede ser reducida o eliminada al suprimir funcionalmente el gen que codifica para EC 4.1.99.1 en la

célula hospedera. Alternativamente, un exceso de indol puede ser utilizado como un sustrato en un proceso *in vitro* o *in vivo* en combinación con cantidades incrementadas del gen que codifica para EC 4.1.99 (Kawasaki et al., *J. Ferm. And Bioeng.*, 82:604-6, 1996). Pueden ser realizadas modificaciones genéticas para incrementar el nivel de los intermediarios tales como D-eritrosa-4-fosfato y corismato.

La producción de triptofano es regulado en la mayoría de los organismos. Un mecanismo es vía la inhibición de la retroalimentación de ciertas enzimas en la vía; conforme se incrementan los niveles de triptofano, disminuye la velocidad de producción del triptofano. De este modo, cuando se utiliza una célula hospedera manipulada por ingeniería genética para producir monatina vía un intermediario de triptofano, puede ser utilizado un organismo que no es sensible a las concentraciones de triptofano. Por ejemplo, una cepa de *Catharanthus roseus* que es resistente a la inhibición del desarrollo por diversos análogos de triptofano, fue seleccionada por la exposición repetida a altas concentraciones de 5-metiltriptofano (Schallenberg y Berlin, *Z. Naturforsch* 34:541-5, 1979). La actividad resultante de la triptofano-sintasa de la cepa, fue menos afectada por la inhibición del producto, probablemente debido a mutaciones en el gen. Similarmente, una célula hospedera utilizada para la producción de monatina puede ser optimizada.

182

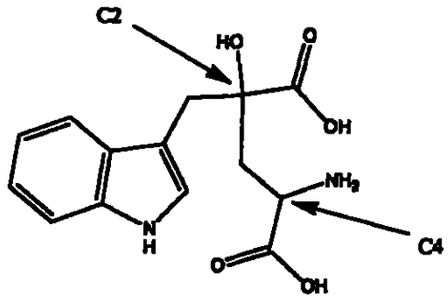
La producción de triptofano puede ser optimizada a través del uso de la evolución dirigida para evolucionar los polipéptidos que son menos sensibles a la inhibición del producto. Por ejemplo, la selección puede ser realizada sobre placas que no contienen triptofano en el medio, pero con altos niveles de análogos de triptofano no metabolizables. Las Patentes de los Estados Unidos Nos. 5,756,345; 4,742,007; y 4,371,614 describen los métodos utilizados para incrementar la productividad de triptofano en un organismo de fermentación. El último paso de la biosíntesis del triptofano es la adición de serina a indol. Por lo tanto, la disponibilidad de la serina puede ser incrementada para incrementar la producción de triptofano.

La cantidad de monatina producida por un organismo de fermentación puede ser incrementada al incrementar la cantidad de piruvato producido por el organismo hospedero. Ciertas levaduras, tales como *Trichosporon cutaneum* (Wang et al., Lett. Appl. Microbiol. 35:338-42, 2002) y *Torulopsis glabrata* (Li et al., *Appl. Microbiol. Biotechnol.* 57:451-9, 2001) sobre-producen piruvato y pueden ser utilizadas para practicar los métodos descritos aquí. Además, pueden ser realizadas modificaciones genéticas a los organismos para promover la producción de ácido pirúvico, tales como aquellas en la cepa W1485*lip2* de *E. coli* (Kawasaki et al., *J. Ferm. And Bioeng.* 82:604-6, 1996).

183

Control de la Quiralidad

El perfil de sabor de la monatina puede ser alterado al controlar la estereoquímica (quiralidad) de la molécula. Por ejemplo, pueden ser deseados diferentes isómeros de la monatina en diferentes mezclas de concentraciones para diferentes sistemas alimenticios. La quiralidad puede ser controlada vía una combinación del pH y de los polipéptidos.



La racemización en la posición C-4 de la monatina (ver molécula numerada arriba) puede ocurrir mediante desprotonación y reprotonación por el carbono alfa, lo cual puede ocurrir por un desplazamiento en el pH o por reacción con el co-factor PLP. En un microorganismo, es improbable que el pH se desplace lo suficiente para provocar la racemización pero PLP es abundante. Los métodos para controlar la quiralidad con los polipéptidos dependen de la ruta biosintética utilizada para la producción de monatina.

Cuando la monatina es formada utilizando la vía mostrada en la FIG. 2, puede ser considerado lo siguiente. En una reacción biocatalítica, la quiralidad del carbono 2 es

determinada por la enzima que convierte el indol-3-piruvato a MP. Múltiples enzimas (por ejemplo de EC 4.1.2, 4.1.3-) pueden convertir el indol-3-piruvato a MP, de este modo, se puede elegir la enzima que forme el isómero deseado. Alternativamente, la enantioespecificidad de la enzima que convierte el indol-3-piruvato a MP puede ser modificada a través del uso de la evolución directa o los anticuerpos catalíticos pueden ser manipulados mediante ingeniería genética para catalizar la reacción deseada. Una vez que se produce MP (ya sea enzimáticamente o mediante condensación química) el grupo amino puede ser agregado estereoespecíficamente utilizando una transaminasa, tales como aquellas descritas en la presente. Puede ser generada ya sea la configuración R o la configuración S del carbono 4 dependiendo de si se utiliza la aminotransferasa del ácido D- o L-aromático. La mayoría de las aminotransferasas son específicas para el isómero L, no obstante, existen D-triptofano-aminotransferasa en ciertas plantas (Kohiba y Mito, Proceedings of the 8th International Symposium on Vitamin B_6 and Carbonyl Catalysis, Osaka, Japón 1990). Además, las D-alanina-aminotransferasas (2.6.1.21), las D-metionina-piruvato-aminotransferasas (2.6.1.41) y la (R)-3-amino-2-metilpropanoato-aminotransferasa (2.6.1.61) y la (S)-3-amino-2-metilpropanoato-aminotransferasa (2.6.1.22) han sido identificadas. Ciertas aminotransferasas pueden únicamente

aceptar el sustrato para esta reacción con una configuración particular en el carbono C2. Por lo tanto, incluso si la conversión a MP no es estereoespecífica, la estereoquímica del producto final puede ser controlada a través de la selección apropiada de una transaminasa. Ya que las reacciones son reversibles, el MP sin reaccionar (isómero no deseado) puede ser reciclado nuevamente a sus constituyentes, y puede ser reformada una mezcla racémica de MP.

Activación de los Sustratos

Pueden ser utilizados sustratos fosforilados, tales como fosfoenolpiruvato (PEP), en las reacciones descritas en la presente. Los sustratos fosforilados pueden ser más energéticamente favorables y, por lo tanto, pueden ser utilizados para incrementar las velocidades de reacción y/o los rendimientos. En las condensaciones de aldol, la adición de un grupo fosfato estabiliza el tautómero de enol del sustrato nucleofílico, haciéndolo más reactivo. En otras reacciones, un sustrato fosforilado frecuentemente proporciona un mejor grupo saliente. Similarmente, los sustratos pueden ser activados mediante conversión a derivados de CoA o derivados de pirofosfato.

Ejemplo 1

Clonación y Expresión de Triptofano-Aminotransferasas

Este ejemplo describe los métodos que fueron utilizados para clonar las triptofano-aminotransferasas, que

pueden ser utilizadas para convertir el triptofano a indol-3-piruvato.

Panorama Experimental

Once genes que codifican para las aminotransferasas fueron clonados en *E. coli*. Estos genes fueron la D-alanina-aminotransferasa de *Bacillus subtilis* (dat, Genbank No. de Acceso Y14082, 1 pb 28622-29470) y Genbank No. de Acceso NP_388848.1, las secuencias de ácidos nucleicos y la secuencia de aminoácidos, respectivamente. La tirosina-aminotransferasa de *Sinorhizobium meliloti* (también denominado *Rhizobium meliloti*) (tatA, SEQ ID Nos.: 1 y 2, la secuencia de ácidos nucleicos y la secuencia de aminoácidos, respectivamente), la tirosina-aminotransferasa de la cepa 2.4.1 de *Rhodobacter sphaeroides* (tatA validada por homología, SEQ ID Nos.: 3 y 4, la secuencia de ácidos nucleicos y la secuencia de aminoácidos, respectivamente), la tirosina-aminotransferasa de *R. sphaeroides* 35053 (validada por homología, SEQ ID Nos.: 5 y 6, la secuencia de ácido nucleico y la secuencia de aminoácidos, respectivamente), la aminotransferasa de amplio sustrato de la *Leishmania major* (*bsat*, validada por homología a los fragmentos peptídicos provenientes de *L. mexicana*, SEQ ID Nos.: 7 y 8, la secuencia de ácidos nucleicos y la secuencia de aminoácidos, respectivamente), la aminotransferasa de aromáticos de *Bacillus subtilis* (*araT*, validada por homología, SEQ ID Nos.:

9 y 10, la secuencia de ácidos nucleicos y la secuencia de aminoácidos, respectivamente), la aminotransferasa de aromáticos de *Lactobacillus amylovorus* (*araT* validada por homología, SEQ ID Nos.: 11 y 12, la secuencia de ácidos nucleicos y la secuencia de aminoácidos, respectivamente), la aminotransferasa de sustratos múltiples de *R. Sphaeroides* 35053 (validada por homología SEQ ID Nos.: 13 y 14, la secuencia de ácidos nucleicos y la secuencia de aminoácidos, respectivamente), la aminotransferasa de sustratos múltiples de *Rhodobacter sphaeroides* cepa 2.4.1 (*msa* validada por homología, Genbank No. de Acceso AAAE01000093.1, pb 14743-16155 y Genbank No. de Acceso ZP00005082, la secuencia de ácidos nucleicos y la secuencia de aminoácidos, respectivamente), aspartato-aminotransferasa de *Escherichia coli* (*aspC*, Genbank No. de Acceso AE000195.1, pb 2755-1565 y Genbank No. de Acceso AAC74014.1, la secuencia de ácidos nucleicos y la secuencia de aminoácidos, respectivamente), y tirosina-aminotransferasa de *E. coli* (*tirB*, SEQ ID Nos.: 31 y 32, la secuencia de ácidos nucleicos y la secuencia de aminoácidos, respectivamente).

Los genes fueron clonados, expresados, y probados para la actividad en la conversión de triptofano a indol-3-piruvato, junto con las enzimas comercialmente disponibles. Todos los once clones tuvieron actividad.

Identificación de Cepas Bacterianas que Pueden Contener Polipéptidos con la Actividad Deseada

Ninguno de los genes en la base de datos de NCBI (National Center for Biotechnology Information) fueron designados como triptofano-aminotransferasas. No obstante, los organismos que tienen esta actividad enzimática han sido identificados. La actividad de L-triptofano-aminotransferasa (TAT) ha sido medida en los extractos celulares o a partir de proteína purificada proveniente de las siguientes fuentes: aislado Rhizobacteriano de *Festuca octoflora*, mitocondrias y citosol de chícharo, células de agalla de corona de girasol, *Rhizobium leguminosarum* biovar *trifoli*, Erwinia herbicola pv *gypsophiale*, Pseudomonas syringae *pv. savastanoi*, *Agrobacterium tumefaciens*, *Azospirrillum lipferum & brasilense*, *Enterobacter cloacae*, *Enterobacter agglomerans*, *Bradyhizobium elkanii*, *Candida maltosa*, *Azotobacter vinelandii*, cerebro de rata, hígado de rata, *Sinorhizobium meliloti*, *Pseudomonas fluorescens* CHA0, *Lactococcus lactis*, *Lactobacillus casei*, *Lactobacillus helveticus*, plántulas de trigo, cebada, *Phaseoulus aureus* (fríjol mung), *Saccharomyces uvarum* (carlsbergensis), *Leishmania sp.*, maíz, brotes de tomate, plantas de chícharo, tabaco, cerdo, *Clostridium sporogenes* y *Streptomyces griseus*.

Aislamiento de ADN Genómico para la Clonación

Se desarrolló *S. meliloti* (ATCC número 9930) en medio

TY a 25°C, pH 7.2. Las células se desarrollaron a una densidad óptica de 600 nm (OD_{600}) de 1.85 y se utilizó un inoculo de 2% para las preparaciones de ADN genómico. El equipo Qiagen Genomic tipo 20/G (Valencia, CA) fue utilizado para el aislamiento del ADN genómico.

El *Bacillus subtilis* 6051 (ATCC) se desarrolló a 30°C en Caldo Nutritivo Bereto (Difco; Detroit, MI). Se utilizó protocolo de Qiagen Genomic tipo 20/G para aislar el ADN genómico con los siguientes cambios: las concentraciones de la proteinasa K y la lisozima fueron duplicadas y los tiempos de incubación fueron incrementados de 2-3 veces.

Se suministró ADN genómico de *Leishmania major* ATCC 50122 por IDI, Inc. (Québec, Canadá) en amortiguador TE pH 8.01, 17 ng/µl.

Rhodobacter sphaeroides 2.4.1 (proporcionado por el Profesor Sam Kaplan de la Universidad de Texas, Houston), *R. sphaeroides* 35053 (ATCC número), y el ADN genómico de *L. amylovorus* se preparó mediante extracción estándar con fenol. Las células fueron cosechadas en fase logarítmica tardía, resuspendidas en amortiguador TEN (Tris-HCl 10 mM, pH 7.5, EDTA 1 mM, cloruro de sodio 100 mM), y se lisaron por la adición de 0.024 mL de urilsarcosina de sodio por mL de suspensión celular. Después de extraer al menos tres veces con un volumen igual de fenol saturado con el amortiguador TE (Tris-HCl 10 mM, pH 7.5, EDTA 1 mM), la solución de ADN fue

190

extraída una vez con cloroformo:octanol 9:1 y tres veces con cloroformo. El ADN fue precipitado por la adición de 0.1 volumen de acetato de sodio 3M, pH 6.8 y 2 volúmenes de etanol. El precipitado se recolectó mediante centrifugación y se lavó una vez con etanol al 70%. Finalmente, el ADN se disolvió en 0.10 mL de agua destilada.

El ADN genómico de *Escherichia coli* fue aislado de la cepa DH10B (Invitrogen) y preparada utilizando el equipo Qiagen Genomic-tip^{MR} (500/G). A partir de 30 mL de esta cepa desarrollada en LB a una DO_{650} de 1.87, se obtuvieron 0.3 mg de ADN purificado. El ADN purificado fue disuelto en amortiguador de elusión Qiagen (EB) a una concentración de 0.37 µg/µL.

Protocolo de la Reacción en Cadena de Polimerasa

Se diseñaron cebadores con proyecciones compatibles para el vector Xa/LIC de pET 30 (Novagen, Madison, WI). El vector pET tiene una proyección de una sola hebra de 12 bases sobre el lado 5′ del sitio Xa/LIC y una proyección de una sola hebra de 15 bases sobre el lado 3′ del sitio Xa/LIC. El plásmido es diseñado para la clonación independiente de la ligadura, con los marcadores de His y S-N-terminales y un marcador de His-S-terminal opcional. El sitio de reconocimiento de la proteasa Xa (IEGR) se asienta directamente enfrente del codón de inicio del gen de interés, tal que pueden ser removidos los marcadores de la proteína de

fusión.

La siguiente secuencia fue agregada a los extremos 5' de las secuencias específicas del organismo cuando se diseñan los cebadores: cebador delantero, 5'GGTATTGAGGGTCGC (SEQ ID No.: 73); cebador inverso, 5'AGAGGAGAGTTAGAGCC (SEQ ID No.: 74).

Cebadores *dat* de *Bacillus subtilis*: Extremo N: 5'-GGTATTGAGGGTCGCATGAAGGTTTTAGTCAATGG-3' y Extremo C 5'-AGAGGAGAGTTAGAGCCTTATGAAATGCTAGCAGCCT-3' (SEQ ID Nos.: 15 y 16).

Cebadores *tatA* de *Sinorhizobium meliloti*: Extremo N: 5'-GGTATTGAGGGTCGCATGTTCGACGCCCTCGCCCG y Extremo C 5'-AGAGGAGAGTTAGAGCCTCAGAGACTGGTGAACTTGC (SEQ ID Nos.: 17 y 18).

Cebadores *araT* de *Bacillus subtilis*: Extremo N: 5'-GGTATTGAGGGTCGCATGGAACATTTGCTGAATCC y Extremo C 5'-AGAGGAGAGTTAGAGCCTTAAACGCCGTTGTTTATCG (SEQ ID Nos.: 19 y 20).

Cebadores *msa* de *Rhodobacter sphaeroides* (ambos 2.4.1 y 35053): Extremo N: 5'-GGTATTGAGGGTCGCATGCGCGAGCCTCTTGCCCT y Extremo C 5'-AGAGGAGAGTTAGAGCCTCAGCCGGGGAAGCTCCGGG (SEQ ID Nos.: 21 y 22).

Cebadores *bsat* de *Leishmania major*: Extremo N: 5'-GGTATTGAGGGTCGCATGTCCACGCAGGCGGCCAT y Extremo C 5'-AGAGGAGAGTTAGAGCCTCACTCACGATTCACATTGC (SEQ ID Nos.: 23 y 24).

Cebadores *araT* de *Lactobacillus amylovorus*: Extremo N: 5'-GGTATTGAGGGTCGCATGCCAGAATTAGCTAATGA y Extremo C 5'-

AGAGGAGAGTTAGAGCCTTATTCGTCCTCTTGTAAAA (SEQ ID Nos.: 25 y 26).

Cebadores *tatA* de *Rhodobacter sphaeroides* (ambos 2.4.1 y 35053): Extremo N: 5'-GGTATTGAGGGTCGCATGCGCTCTACGACGGCTCC y Extremo C 5'-AGAGGAGAGTTAGAGCCTCAGCCGCGCAGCACCTTGG (SEQ ID Nos.: 27 y 28).

Cebadores *aspC* de *Escherichia coli*: Extremo N: 5'-GGTATTGAGGGTCGCATGTTTGAGAACATTACCGC-3' y Extremo C 5'-AGAGGAGAGTTAGAGCCTTACAGCACTGCCACAATCG-3' (SEQ ID Nos.: 29 y 30).

Cebadores *tyrB* de *Escherichia coli*: Extremo N: 5'-GGTATTGAGGGTCGCGTGTTTCAAAAAGTTGACGC y Extremo C 5'-AGAGGAGAGTTAGAGCCTTACATCACCGAGCAAACG-3' (SEQ ID Nos.: 33 y 34).

El gen derivado de *S. meliloti* (*tatA*) fue amplificado utilizando el siguiente protocolo de PCR. En una reacción de 50 µl, se utilizaron 0.1-0.5 µg de plantilla, 1.5 µM de cada cebador, 0.4 mM de cada dNTP, 3.5 U de Polimerasa de Alta Fidelidad de Expansión (Roche, Indianápolis, IN), y 1X de amortiguador Expand^{MR} con magnesio. El programa termociclador utilizado incluyó un inicio en caliente a 96°C por 5 minutos, seguido por 29 repeticiones de los siguientes pasos: 94°C por 30 segundos, 55°C por 2 minutos, y 72°C por 2.5 minutos. Después de las 29 repeticiones la muestra fue mantenida a 72°C por 10 minutos y luego almacenada a 4°C. Este protocolo de PCR produjo un producto de 1199 pares de bases.

193

Las secuencias de los genes derivados de *R. sphaeroides* (*msa* y *tatA*), *araT* de *L. amylovorus*, y *araT* de *Bacillus* fueron amplificados utilizando el siguiente protocolo de PCR. En una reacción de 50 µl, se agregaron 0.1-0.5 µg de plantilla, 1.5 µM de cada cebador, 0.4 mM de dNTP, 3.5 U de Polimerasa Expand High Fidelity^{MR} y 1X de amortiguador Expand^{MR} con magnesio. El programa termociclador utilizado incluyó un inicio en caliente a 96°C por 5 minutos, seguido por 29 repeticiones de los siguientes pasos. 94°C por 30 segundos, 40-60°C por 1 minuto, 45 segundos (termociclador en gradiente) y 72°C por 2 minutos, 15 segundos. Después de las 29 repeticiones, la muestra fue mantenida a 72°C por 10 minutos, y luego se almacenó a 4°C.

Para cada gen *msa* de *R. Sphaeroides*, las temperaturas de recocido de 42°C y 48°C produjeron múltiples productos, pero una banda distinta a aproximadamente a 1464 pares de bases. Para *araT* de *L. amiylovorus*, las temperaturas de recocido de 42°C, 48°C y 56°C produjeron productos simples con bandas intensas a 1173 pares de bases. Para *araT* de *B. subtilis*, las temperaturas de recocido de 40°C, 45°C, 50°C, 55°C generaron productos intensos simples (1173 pares de bases) a partir del ADN genómico y de las colonias. Para *bsat* de *L. major*, la temperatura de recocido de 55°C dio el producto más limpio (1239 pares de bases). Para los genes *tatA* de *Rhodobacter*, las temperaturas de recocido de 50°C,

55°C dieron productos limpios al tamaño correcto (1260 pares de bases). Para los genes de *E. coli* y el gen *dat* de *B. subtilis*, se utilizó una temperatura de recocido de 55-60°C, y el tiempo de recocido fue acortado a 45 segundos. Fueron obtenidos productos limpios de los tamaños correctos (aproximadamente 1.3 kb para los genes de *E. coli*, 850 pares de bases para el gen *dat*).

Clonación

Los productos de PCR fueron purificados en gel a partir de geles de TAE-agarosa a 0.8 o 1% utilizando el equipo de extracción en gel Qiagen (Valencia, CA). Los productos de PCR fueron cuantificados mediante comparación a los estándares sobre un gel de agarosa, y luego tratados con la ADN polimerasa T4 siguiendo los protocolos recomendados por el fabricante para clonación independiente de la ligadura (Novagen, Madison, WI).

En resumen, aproximadamente 0.2 pmol del producto de PCR purificado fueron tratados con una u de ADN polimerasa-T4 en presencia de dGTP por 30 minutos a 22°C. La polimerasa elimina las bases sucesivas de los extremos 3' del producto de PCR. Cuando la polimerasa encuentra un residuo de guanina, la actividad de la polimerasa 5' a 3' de la enzima, contraataca la actividad de exonucleasa para prevenir efectivamente la extirpación posterior. Esto crea proyecciones de una sola hebra que son compatibles con el

vector pET Xa/LIC. La polimerasa es inactivada por incubación a 75°C por 20 minutos.

El vector y el inserto tratados fueron recocidos como es recomendado por Novagen. Aproximadamente 0.02 pmol del inserto tratado y 0.01 pmol del vector se incubaron por 5 minutos a 22°C, se agregó EDTA 6.25 mM (concentración final), y la incubación a 22°C fue repetida. 1 µl de la reacción de recocido fue agregado a las células componentes simples NovaBlue^{MR} (Novagen, Madison, WI) y se incubó sobre hielo por 5 minutos. Después del mezclado, las células fueron transformadas mediante choque térmico por 30 segundos a 42°C. Las células fueron colocadas sobre hielo por 2 minutos, y se dejaron recuperar en 250 µL de SOC a temperatura ambiente por 30 minutos a 37°C con agitación a 225 rpm. Las células fueron sembradas en placas sobre placas de LV que contenían kanamicina (25-50 µg/mL).

El ADN plasmídico fue purificado utilizando el equipo de minipreparación por centrifugación Qiagen y seleccionado para los insertos correctos mediante digestión por restricción con XhoI y XbaI. Las secuencias de los plásmidos que parecieron tener el inserto correcto fueron verificadas mediante secuenciamiento de ADN de terminación de cadena de didesoxi.

Las SEQ ID Nos.: 1-14 y 31-32 muestran secuencias de nucleótidos y de los aminoácidos correspondientes de las

196

aminotransferasa recombinantes, cualesquiera cambios a partir de las secuencias de Genbank ya sea sustituciones conservadoras silenciosas o generadas, en la secuencia de proteína. Las SEQ ID Nos.: 11 y 12 son secuencias novedosas.

Expresión del Gen y Ensayos

El ADN plasmídico, verificado por análisis secuencial, fue subclonado dentro de hospederos de expresión de *E. coli* BLR(DE3) ó BL21(DE3) (Novagen, Madison, WI). Los cultivos fueron desarrollados y los plásmidos fueron aislados utilizando el equipo miniprep de Qiagen, y analizados por digestión por restricción para confirmar la identidad.

La inducción fue inicialmente realizada con araT de *L. amylovorus*, araT de *B. Subtilis*, y tatA de *S. meliloti* en células BLR(DE3) y BL21(DE3). Un estudio en el curso de tiempo fue utilizado con los cultivos desarrollados en LB que contenía kanamicina (30 mg/L) a una DO_{600} de 0.5-0.8 e inducida con IPTG 1 mM (tiogalactósido de isopropilo) y muestreadas a 0, 1, 2 y 4 horas después de la inducción. Las células provenientes de 2.0 mL fueron resuspendidas en 0.10 mL de Tris-HCl 120 mM, pH 6.8, que contenía dodecilsulfato de sodio al 10%, 10% de 2-mercaptoetanol, y 20% de glicerol, calentado a 95°C por 10 minutos, y enfriado y diluido con 0.10 ml de agua. Las alícuotas de estas muestras de proteína celular total fueron analizadas mediante SDS-PAGE, utilizando un gel en gradiente de 4 a 15%. No existieron diferencias

197

significativas en la cantidad de proteína expresada entre la inducción a 2 horas y a 4 horas, ni entre las células BLR(DE3) y BL21(DE3).

Los extractos celulares fueron también preparados a partir de muestras de 4 horas al suspender los botones celulares a partir de 2 ml de cultivo en 0.25 ml de reactivo Novagen BugBuster^{MR} que contenía 0.25 µl de nucleasa benzonasa, incubando a temperatura ambiente por 20 minutos con agitación suave, y centrifugando a 16,000 x g para eliminar el desecho celular. Los sobrenadantes (extractos celulares) fueron cargados sobre geles en gradiente de 4 a 15% para el análisis de las proteínas solubles celulares.

Los tres clones (*L. amylovorus araT* (SEQ ID NOS.: 11 y 12), *B. Subtilis araT* (SEQ ID NOS.: 9 y 10), y *S. meliloti tatA* (SEQ ID NOS.: 1 y 2) mostraron proteína soluble que correspondió al tamaño correcto (aproximadamente 45 kDa). El producto del gen *araT de B. Subtilis* fue sobreexpresado al más alto nivel y/o fue más soluble que los otros dos productos génicos.

En métodos de expresión subsecuentes, el ADN plasmídico proveniente de los clones positivos fue subclonado en BL21(DE3) debido a las mejores características de desarrollo de este hospedero. La inducción fue repetida utilizando un IPTG 1 mM con cultivos desarrollados en LB que contenían kanamicina a 50 mg/litro, induciendo cuando la OD_{600}

alcanzó aproximadamente 0.8. Las células fueron cosechadas después de 4 horas de desarrollo a 37°C, centrifugadas a 3000 rpm por 10 minutos (4°C), lavadas con amortiguador TEGGP (Tris-HCl 50 mM (pH 7.0), EDTA 0.5 mM, 100 mg/litro de glutatión, 5% de glicerol, con coctel inhibidor de proteasa completo de Roche), y se congeló instantáneamente en etanol a -80°C.

Las muestras fueron resuspendidas en 5 ml/g de peso de células húmedas del reactivo de Bugbuster^{MR} (Novagen) que contenía 5 µl/ml del equipo #3 de coctel inhibidor de proteasa (Calbiochem-Novabiochem Corp., San Diego, CA) y 1 µl/ml de la nucleasa benzonasa. Las muestras fueron incubadas a temperatura ambiente por 20 minutos en un agitador orbital. Los desechos celulares insolubles fueron removidos mediante centrifugación a 16,000 x g por 20 minutos a 4°C.

Los extractos celulares fueron analizados mediante SDS-PAGE, y evaluados para la actividad de triptofano-aminotransferasa al seguir la producción del ácido indol-pirúvico utilizando el siguiente protocolo. Se llevaron a cabo reacciones de un ml en tetraborato de sodio 50 mM (pH 8.5), EDTA 0.5 mM, arseniato de sodio 0.5 mM, fosfato de piridoxal 50 µM, α-cetoglutarato 5 mM, y L-triptofano 5 mM. Las reacciones fueron iniciadas por la adición de extractos libres de células o la enzima purificada, y se incubaron 30 minutos a 30°C. Se agregaron 200 µl de TCA al 20% para

detener la reacción, y la proteína precipitada fue removida mediante centrifugación. La absorbancia a 327 nm fue medida y comparada a una curva estándar de indol-3-piruvato recién preparado, en el amortiguador de ensayo. Las reacciones control sin el sustrato de triptofano o utilizando extractos libres de células, provenientes de clones transformados con pET30a solo, fueron también realizadas.

Debido al antecedente de las aminotransferasas nativas de *E. coli* en los extractos celulares, se purificaron las proteínas recombinantes utilizando cartuchos His-Bind siguiendo los protocolos del fabricante (Novagen, Madison, WI). Las fracciones eluyentes fueron desaladas sobre columnas PD-10 (Amersham Biosciences, Piscataway, NJ) y eluidas en Tris 50 mM, pH 7.0. Las proteínas purificadas fueron analizadas mediante SDS-PAGE y evaluadas para la actividad de aminotransferasa.

Los resultados provenientes de la inducción a 37°C con IPTG 1 mM (4 horas) demuestran que *L. major bsat, S. meliloti tatA, E. coli aspC* y ambos clones de *tatA de R. sphaeroides* tienen niveles significativos de actividad de triptofano-aminotransferasa. La proteína *araT de B. subtilis* fue sobre-expresada y fue soluble, pero mostró poca actividad enzimática. El producto del gen *araT de L. amylovorus* pareció ser soluble en el extracto celular, pero la purificación utilizando un cartucho His-Bind dio como resultado únicamente

pequeñas cantidades de proteína con el peso molecular correcto. Los productos del gen *msa* fueron insolubles y además fueron realizados experimentos de expresión a 24°C para reducir al mínimo la formación de cuerpos de inclusión. Se utilizaron varias concentraciones de IPTG entre 10 µm y 1 mM para elevar al máximo la cantidad de proteína soluble.

La Tabla 1 lista las actividades específicas medidas en los microorganismos de indol-3-piruvato (I3P) formados por miligramo de proteína por minuto. En algunos casos, cantidades muy pequeñas de proteína recombinante mostraron altos niveles de actividad por arriba del intervalo lineal efectivo del ensayo. En estos casos un signo '>' precede el número de actividad específica.

Tabla 1: Actividades Específicas de los Clones en Extractos Celulares (CE) y Proteínas Purificadas (P) y Enzimas Comerciales

Enzima	Actividad Específica (µg I3P/mg proteína/ minuto)	Nota
L. major bsat CE	>49.3	
L. major bsat P	>4280	
S. meliloti tatA CE	>28.6	
S. meliloti tatA P	>931	
2.4.1 *tatA* CE	>41.2	
2.4.1 *tatA* P	1086	

Enzima	Actividad Específica (µg I3P/mg proteína/ minuto)	Nota
35053 *tatA* CE	>62.3	
35053 *tatA* P	>486	
L. amylovorus araT CE	1.26	
L. amylovorus araT P	0	poca proteína después del cartucho His-Bind
B. subtilis araT CE	0	no detectable
B. subtilis araT P	1.5-4.5	
2.4.1 *msa* CE	2.05	muy poca proteína soluble
2.4.1 *msa* P	0	no hay proteína después del cartucho His-Bind
35053 *msa* CE	3.97	muy poca proteína soluble
35053 *msa* P	0	no hay proteína después del cartucho His-Bind
E. coli aspC (P)	800	
E. coli tyrB (P)	1	no muy soluble
B. subtilis D-aminotransf. (P)	2.7	usando D-triptofano como sustrato
Transaminasa de intervalo amplio	22	Sigma Cat # T 7684
Porcina tipo II-A	1.5	Sigma G7005
Porcina tipo I	1	Sigma G2751

Un alineamiento que compara todas las proteínas recombinantes clonadas, ilustra que no existen muchas áreas altamente conservadas entre las secuencias de *araT*, *tatA*, *bsat*, y *msa*. Un alineamiento de las proteínas recombinantes de más alta actividad. Los homólogos del producto del gen

tatA de Rhodobacter, la aminotransferasa de sustrato amplio de *L. major*, y la tirosina-aminotransferasa de *Sinorhizobium meliloti* mostraron varias regiones conservadas, no obstante éstas son únicamente aproximadamente 30 a 46% idénticas al nivel de proteína. La disponibilidad de la aminotransferasa específica de D (D-alanina) de amplio intervalo, puede ser útil en la producción de otros estereoisómeros de monatina.

Ejemplo 2

Conversión del indol-3-lactato al indol-3-piruvato

Como se muestra en las Figuras 1 y 3, el ácido indol-3-láctico puede ser utilizado para producir el indol-3-piruvato. La conversión entre el ácido láctico y el piruvato es una reacción reversible, como lo es la conversión entre el indol-3-piruvato y el indol-3-lactato. La oxidación del indol-lactato fue típicamente seguida debido a la alta cantidad de antecedente a 340 nm proveniente del indol-3-piruvato.

La mezcla de ensayo estándar contenía fosfato de potasio 100 mM, pH 8.0, NAD^+ 0.3 mM, 7 unidades de lactato-deshidrogenasa (LDH) (Sigma-L2395, Saint Louis, MO), y sustrato 2 mM en 0.1 ml. El ensayo fue realizado por duplicado en una placa de microtitulación transparente a UV, utilizando un lector de placas Molecular Devices SpectraMax Plus. El polipéptido y el amortiguador fueron mezclados y llevados con pipeta hacia pozos que contenían el ácido indol-

3-láctico y NAD^+, y la absorbancia a 340 nm de cada pozo fue leída a intervalos de 9 segundos después de un breve mezclado. La reacción fue mantenida a 25°C por 5 minutos. El incremento en la absorbancia a 340 nm sigue la producción de NADH a partir de NAD^+. Se realizaron controles negativos separados sin NAD^+ y sin sustrato. D-LDH proveniente de *Leuconostoc mesenteroides* (Sigma número de catálogo L2395) pareció mostrar más actividad con los sustratos derivados de indol que L-LDH proveniente de *Bacillus stearothermophilus* (Sigma número de catálogo L5275).

Se utilizaron métodos similares con el ácido D-láctico y NAD^+ o NADH y piruvato, los sustratos naturales de los polipéptidos de D-LDH. La V_{max} para la reducción de piruvato fue 100 a 1000 veces mayor que la V_{max} para la oxidación del lactato. La V_{max} para la reacción de oxidación del indol-3-láctico con D-LDH fue aproximadamente un quinto de aquella con el ácido láctico. La presencia del indol-3-piruvato fue también medida siguiendo el cambio en la absorbancia a 327 nm (el derivado de enol-borato) utilizando amortiguador de borato de sodio 50 mM que contenía EDTA 0.5 mM y arseniato de sodio 0.5 mM. Se observaron cambios en absorbancia, pequeños pero repetibles, en comparación a los controles negativos para los polipéptidos L- y D-LDH.

Adicionalmente, las lactato-deshidrogenasas de amplia especificidad (enzimas con actividad asociada con EC

1.1.1.27, EC 1.1.1.28 y/o EC 1.1.2.3), pueden ser clonadas y utilizadas para elaborar el indol-3-piruvato a partir del ácido indol-3-láctico. Las fuentes de deshidrogenasas de amplia especificidad incluyen *E. coli*, *Neisseria gonorrheae* y *Lactobacillus plantarum*.

Alternativamente, el indol-3-piruvato puede ser producido al poner en contacto el indol-3-lactato con extractos celulares de *Clostridium sporogenes* que contienen una deshidrogenasa del indol-lactato (EC 1.1.1.110); o extractos celulares de *epimastigotes de Tripanosoma cruzi* que contienen la p-hidroxifenilacetato-deshidrogenasa (EC 1.1.1.222) que se sabe tiene también actividad sobre el indol-3-piruvato; o extractos celulares de *Pseudomonas, acidovorans* o *E. coli*, que contienen una deshidrogenasa del lactato de imidazol-5-ilo (EC 1.1.1.111); o *Coleus blumei*, que contiene una hidroxifenilpiruvato-reductasa (EC 1.1.1.222); o *Candida maltosa* que contiene una deshidrogenasa de lactato-D-aromático (EC 1.1.1.222). Las referencias que describen tales actividades incluyen, Nowicki et al. (*FEMS Microbiol Lett* 71: 119-24, 1992), Jean y DeMoos (*Canadian J. Microbiol*. 14 1968, Coote y Hassall (*Biochem. J* 111: 237-9, 1969), Cortese et al. (*C.R. Seances Soc. Biol. Fil*. 162 390-5, 1968), Petersen y Alfermann (*Z. Naturforsch. C: Biosci*. 43 501-4, 1988), y Bhatnagar et al. (*J. Gen Microbiol* 135: 353-60, 1989). Además, una lactato-oxidasa tal como aquella

proveniente de *Pseudomonas sp* (Gu et al. J. Mol. Catalysis B: Enzimatic: 18: 299-305, 2002), puede ser utilizada para la oxidación del indol-3-lactato a indol-3-piruvato.

Ejemplo 3

Conversión de L-triptofano a indol-3-piruvato utilizando la oxidasa de L-aminoácido

Este ejemplo describe los métodos utilizados para convertir el triptofano a indol-3-piruvato vía una oxidasa (EC 1.4.3.2), como una alternativa al uso de una triptofano-aminotransferasa como se describe en el Ejemplo 1. La oxidasa de L-aminoácido o L-aminoácido-oxidasa fue purificada a partir de *Crotalus durissus* (Sigma, Saint Louis, MO, número de catálogo A-2805). Los números de acceso de las oxidasas de L-aminoácido para la clonación molecular incluye: CAD21325.1, AAL14831, NP_490275, BAB78253, A38314, CAB71136, JE0266, T08202, S48644, CAC00499, P56742, P81383, O93364, P81382, P81375, S62692, P23623, AAD45200, AAC32267, CAA88452, AP003600 y Z48565.

Las reacciones fueron realizadas en tubos para microcentrífuga en un volumen total de 1 ml, incubados por 10 minutos mientras que se agitaba a 37°C. La mezcla de reacción contenía L-triptofano 5 mM, amortiguador de fosfato de sodio 100 mM pH 6.6, arseniato de sodio 5 mM, EDTA 0.5 mM, tetraborato de sodio 25 mM, catalasa 0.016 mg (83 U, Sigma C.3515), 0.008 mg de FAD (Sigma), y 0.005-0.125 Unidades de

oxidasa de L-aminoácido. Los controles negativos contenían todos los componentes excepto el triptofano, y los blancos contenían todos los componentes excepto la oxidasa. Se utilizó la catalasa para remover el peróxido de hidrógeno formado durante la desaminación oxidativa. El arseniato y el tetraborato de sodio se utilizaron para estabilizar la forma de enol-borato del indol-3-piruvato, que muestra una absorbancia máxima a 327 nm. Se prepararon estándares de indol-3-piruvato a concentraciones de 0.1-1 mM en la mezcla de reacción.

La oxidasa de L-aminoácido adquirida tuvo una actividad específica de 540 µg de indol-3-piruvato formado por minuto por mg de proteína. Este es el mismo orden de magnitud que la actividad específica de las enzimas triptofano-aminotransferasa.

Ejemplo 4

Conversión de indol-3-piruvato a ácido 2-hidroxi-2-(indol-3-ilmetil)-4-ceto-glutárico con una aldolasa

Este ejemplo describe los métodos que pueden ser utilizados para convertir el indol-3-piruvato al precursor de monatina (MP) ácido 2-hidroxi-2-(indol-3-ilmetil)-4-ceto-glutárico utilizando una aldolasa (liasa) (Figura 2). Las condensaciones de aldol son reacciones que forman enlaces carbono-carbono entre el carbono beta de un aldehído o cetona

y el carbono carbonílico de otro aldehído o cetona. Se forma un carbanión sobre el carbono adyacente al grupo carbonilo de un sustrato, y sirve como un nucleófilo que ataca el carbono carbonílico del segundo sustrato (el carbono electrofílico). Más comúnmente, el sustrato electofílico es un aldehído, de modo que la mayoría de las aldolasas caen dentro de la categoría EC 4.1.2. Muy frecuentemente, el sustrato nucleofílico es el piruvato. Es menos común que las aldolasas catalicen la condensación entre dos ceto-ácidos o dos aldehídos.

No obstante, las aldolasas que catalizan la condensación de dos ácidos carboxílicos han sido identificadas. Por ejemplo, la Patente Europea EP-1045-029 describe la producción del ácido L-4-hidroxi-2-cetoglutárico a partir del ácido glioxílico y piruvato utilizando un cultivo de *Pseudomonas* (EC 4.1.3.16). Además, la 4-hidroxi-4-metil-2-oxoglutarato-aldolasa (4-hidroxi-4-metil-oxoglutarato-piruvato-liasa, EC 4.1.3.17) puede catalizar la condensación de dos cetoácidos. Por lo tanto, se utilizaron polipéptidos de aldolasa similares para catalizar la condensación del indol-3-piruvato con piruvato.

Clonación

Las 4-hidroxi-4-metil-2-oxoglutarato-piruvato-liasas (aldolasa ProA, EC 4.1.3.17) y la 4-hidroxi-2-oxoglutarato-glioxilato-liasa (aldolasa KHG, EC 4.1.3.16) catalizan

reacciones muy similares a la reacción de la aldolasa de la Figura 2. Fueron diseñados cebadores con grupos sobresalientes compatibles para el vector pET30 Xa/LIC (Novagen, Madison, WI). El diseño de estos cebadores es descrito anteriormente en el Ejemplo 1.

Los siguientes cebadores fueron diseñados para la clonación de pET30 Xa/LIC:

1. El gen *proA de Pseudomonas straminea* (Genbank Acceso No. 12964663 Versión: 12964663) y el gen *proA de Comamonas testosteroni* (SEQ ID NOs: 65-66, secuencia de ácidos nucleicos y secuencia de aminoácidos, respectivamente) delantera 5'-GGTATTGAGGGTCGCATGTACGAACTGGGAGTTGT-3' e inversa 5'-AGAGGAGAGTTAGAGCCTTAGTCAATATATTTCAGGC-3' (SEQ ID NOs: 55 y 56).
2. gen SMc00502 de *Sinorhizobium meliloti* 1021 (homólogo a *proA*, Genbank Acceso Nos. 15074579 y CAC46344, secuencia de ácidos nucleicos y secuencia de aminoácidos, respectivamente) delantera 5'-GGTATTGAGGGTCGCATGAGCGTGGTTCACCGGAA-3' e inversa 5'-AGAGGAGAGTTAGAGCCTCAATCGATATATTTCAGTC-3' (SEQ ID NOs: 61 y 62).
3. gen fldZ de *Sphingomonas sp* LB126 (Genbank Acceso No. 7573247 Versión: 7573247, códigos para una acil-transferasa putativa) delantera 5'-GGTATTGAGGGTCGCATGTCCGGCATCGTTGTCCA-3' e inversa 5'-

AGAGGAGAGTTAGAGCCTCAGACATATTTCAGTCCCA-3' (SEQ ID NOs: 57 y 58).

4. gen pcmE de *Arthrobacter keyseri* (Genbank Acceso No. AF331043 Versión: AF331043.1, codifica para una oxalocitramalato-aldolasa) delantera 5'-GGTATTGAGGGTCGCATGCGACTGAACAACCTCGG-3' e inversa 5'-AGAGGAGAGTTAGAGCCTCAGTTCTCCACGTATTCCA-3' (SEQ ID NOs: 59 y 60).

5. gen YPO0082 de *Yersinia pestis* cepa CO92 (Genbank Acceso No. 15978115 Versión: 15978115, codifica para una posible transferasa) delantera 5'-GGTATTGAGGGTCGCATGAGCCTGGTTAATATGAA-3' e inversa 5'-AGAGGAGAGTTAGAGCCTTATGACTTTAACGCGTTGA-3' (SEQ ID NOs: 63 y 64).

6. gen *khg* de *Bacillus subtilis* (Genbank Acceso Nos. Z99115.1 GI:2634478, 126711-127301 y CAB14127.1, secuencia de ácidos nucleicos y secuencia de aminoácidos, respectivamente) delantera 5'-GGTATTGAGGGTCGCATGGAGTCCAAAGTCGTTGA-3' e inversa 5'-AGAGGAGAGTTAGAGCCTTACACTTGGAAAACAGCCT-3' (SEQ ID NOs: 35 y 36).

7. gen *khg de E. coli* (Genbank Acceso Nos. AE000279.1 1331-1972 y AAC74920.1, secuencia de ácidos nucleicos y secuencia de aminoácidos, respectivamente) delantera 5'-GGTATTGAGGGTCGCATGAAAAACTGGAAAACAAG-3' e inversa 5'-

210

AGAGGAGAGTTAGAGCCTTACAGCTTAGCGCCTTCTA-3' (SEQ ID NOs: 37 y 38).

8. gen *khg de S. meliloti* (Genbank Acceso Nos. AL591792.1 GI:15075850, 65353-64673 y CAC47463.1, secuencia de ácidos nucleicos y secuencia de aminoácidos, respectivamente) delantera 5'-GGTATTGAGGGTCGCATGCGAGGGGCATTATTCAA-3' e inversa 5'-AGAGGAGAGTTAGAGCCTCAGCCCTTGAGCGCGAAG-3' (SEQ ID NOs: 39 y 40).

El ADN genómico proveniente de los organismos descritos en 1-2 y 6-8 anteriores, fue purificado utilizando el protocolo de punta genómica Qiagen. Utilizando técnicas similares el ADN genómico de los organismos descritos en 3-5 puede ser purificado.

Las *Pseudomonas straminea* (ATCC 33636) fue desarrollada a 30°C en Caldo Nutritivo y en medio de hidroxibenzoato. Se desarrolló *Comamonas testosteroni* (ATCC 49249) a 26°C en Caldo Nutritivo y medio de hidroxibenzoato. *Sphingomonas sp.* LB126 (Flemish Institute for Technological Research, VITO, B-2400 Mol, Bélgica) se desarrolla de acuerdo al método descrito por Wattiau et al. (*Research in Microbiol.* 152: 861-72, 2001). *Arthrobacter keyseri* (Gulf Ecology Division, National Health and Environmental Effects Research Laboratory, U.S. Environmental Protection Agency, Gulf Breeze, FL 32561, EUA) se desarrolla de acuerdo al protocolo

descrito por Eaton (*J. Bacteriol.* 183: 3689-3703, 2001) *Sinorhizobium meliloti* 1021 (ATCC 51124) se desarrolló a 26°C en el medio TY de la ATCC y medio de hidroxibenzoato. *Yersinia pestis* cepa CO92 (ATCC) se desarrolla a 26°C en el medio 739 de agar de sangre de Caballo de la ATCC. *Bacillus subtilis* 6051 (ATCC) se desarrolla a 30°C en Caldo Nutritivo Bereto (Difco; Detroit, MI). El ADN genómico de *E. coli* fue aislado de la cepa DH10B (Invitrogen) como se describe en el Ejemplo 1.

Los protocolos de PCR, de clonación y de selección descritos en el Ejemplo 1 fueron utilizados para clonar las secuencias *proA de C. testosteroni* y *de S. meliloti*, así como las secuencias *khg* de *E. coli, B. Subtilis* y *S. meliloti*. Pueden ser utilizados los mismos métodos para clonar las otras secuencias descritas anteriormente.

Los clones positivos fueron secuenciados utilizando el secuenciamiento por terminación de cadena didesoxi (Seqwright, Houston, Texas) con los cebadores S-tag y terminador T7 (Novagen), y los cebadores internos provenientes de Integrated DNA Technologies, Inc. (Coralville, IA).

Ensayos de Expresión y de Actividad

El ADN plasmídico (verificado por análisis secuencial) fue subclonado dentro de un hospedero de expresión BL21(DE) (Novagen). Los cultivos fueron desarrollados en medio LB con

212

50 mg/litro de kanamicina, los plásmidos aislados utilizando un equipo de minipreparación de plásmido giratorio Qiagen y subsecuentemente se analizaron mediante digestión por restricción para confirmar la identidad. Se realizaron experimentos de inducción con las construcciones BL21(DE3) desarrolladas en medio LB que contenía 50 mg/litro de kanamicina a 37°C. La expresión de la proteína fue inducida utilizando IPTG 0.1 mM después de que la OD_{600} alcanzó aproximadamente 0.6. Las células fueron desarrolladas por 4 horas a 30°C y cosechadas mediante centrifugación. Las células fueron luego lisadas utilizando el reactivo Bugbuster^{MR} (Novagen) y las proteínas recombinantes de marcador de His (His-tag) fueron purificadas utilizando los cartuchos His-Bind como se describió anteriormente (Ejemplo 1). Las proteínas purificadas fueron desaladas sobre columnas desechables PD-10 y eluidas en amortiguador de Tris-HCl 50 mM, pH 7.3 con cloruro de magnesio 2 mM.

Las proteínas fueron analizadas mediante SDS-PAGE sobre geles en gradiente de 4 a 15% para detectar los niveles de proteína soluble al peso molecular predicho de la proteína de fusión recombinante.

Las proteínas fueron evaluadas para la actividad utilizando el indol-3-piruvato y el piruvato de sodio como sustrato. La mezcla de ensayo contenía Tris-HCl 100 mM (pH 7-pH 8.9), cloruro de magnesio 0-8 mM, fosfato de potasio 3 mM

(pH 8), y 6 mM de cada sustrato en 1 ml. La reacción se inició mediante la adición de diversas cantidades del polipéptido (por ejemplo de 10 a 100 µg), y se incubó a 25°C-37°C por 30 minutos, se filtró y luego se congeló a -80°C.

Resultados de Actividad con los Productos del Gen *proA*

Las construcciones génicas *proA* de *C. testosteroni* y *SMc00502 de S. meliloti* tuvieron altos niveles de expresión cuando fueron inducidas con IPTG. Las proteínas recombinantes fueron altamente solubles, como se determinó por el análisis de SDS-PAGE de la proteína total y las muestras de extracto celular. El producto génico de *C. testosteroni* fue purificado a una pureza mayor de 95%. Debido a que el rendimiento del producto génico de *S. meliloti* fue muy bajo después de la purificación por afinidad utilizando el cartucho His-Bind, se utilizó el extracto celular para los ensayos enzimáticos.

Ambas aldolasas recombinantes catalizaron la formación de MP a partir del indol-3-piruvato y del piruvato. La presencia de fosfato de potasio y de magnesio divalente fue requerida para la actividad enzimática. Ningún producto fue aparente cuando estuvo ausente el indol-3-piruvato, el piruvato o el fosfato de potasio. Una pequeña cantidad del producto fue también formada en ausencia de la enzima (típicamente un orden de magnitud menor que cuando la enzima estuvo presente).

El pico de producto eluyó a partir de la columna C18 de fase inversa, ligeramente más tarde que el estándar de indol-3-piruvato, el espectro de masa de este pico mostró un ion progenitor inducido por colisión ([M+H]+) de 292.1, el ion progenitor esperado para el producto MP. Los fragmentos hijos mayores presentes en el espectro de masa incluyeron aquellos con m/z=158 (el ion carbanión de 1H-indol-3-carbaldehído), 168 (ion carbonio de 3-buta-1,3-dienil-1H-indol), 274 (292-H_2O), 256 (292-2 H_2O), 238 (292-3 H_2O), 228 (292-CH_4O_3), y 204 (pérdida de piruvato). El producto también mostró un espectro de UV característico de otros compuestos que contienen indol tales como triptofano, con la λ_{max} de 279-280 y un hombro pequeño aproximadamente a 290 nm.

La cantidad de MP producida por la aldolasa de *C. testosteroni* se incrementó con un incremento en la temperatura de reacción a partir de la temperatura ambiente hasta 37°C, la cantidad del sustrato, y la cantidad del magnesio. La actividad sintética de la enzima disminuyó conforme se incrementaba el pH, el producto máximo observado fue a pH 7. Con base en los estándares de triptofano, la cantidad de MP producido bajo un ensayo estándar utilizando 20 µg de la proteína purificada, fue aproximadamente 10-40 µg por un ml de reacción.

Debido al alto grado de homología de las secuencias

que codifican para la aldolasa *proA de S. meliloti* y *C. testosteroni* con los otros genes descritos anteriormente, se espera que todos los productos génicos recombinantes puedan catalizar esta reacción. Además, se espera que las aldolasas que tienen treonina (T) en las posiciones 49 y 87, arginina (R) en 119, aspartato (D) en 120, e histidina (H) en 31 y 71, (con base en el sistema de numeración de *C. testosteroni*) tendrán actividad similar.

Resultados de Actividad con los productos del gen *khg*

Las construcciones del gen *khg de B. Subtilis de E. coli* tuvieron altos niveles de expresión de la proteína cuando se indujo con IPTG, mientras que *khg de S. meliloti* tuvo un nivel menor de expresión. Las proteínas recombinantes fueron altamente solubles, como se juzgó por el análisis de SDS-PAGE de las proteínas totales y de los extractos celulares. Los productos del gen *khg de B. Subtilis y de E. coli* fueron purificados a una pureza mayor de 95%; el rendimiento del producto del gen de *S. meliloti* no fue tan alto después de la purificación por afinidad después de un cartucho His-Bind.

No existe evidencia de que el magnesio y el fosfato sean requeridos para la actividad para esta enzima. No obstante, la literatura reporta la realización de los ensayos en amortiguador de fosfato de sodio, y la enzima al parecer es bifuncional y tiene actividad sobre los sustratos

fosforilados tales como el 2-ceto-3-desoxi-6-fosfogluconato (KDPG). Los ensayos enzimáticos fueron realizados como se describió anteriormente, y en algunos casos el fosfato fue omitido. Los resultados indican que las aldolasas de KHG recombinantes produjeron MP, pero no fueron tan activas como las aldolasas de ProA. En algunos casos, el nivel de MP producido por KHG fue casi idéntico a la cantidad producida por el magnesio y el fosfato solos. El fosfato no pareció incrementar las actividades de KHG. La enzima de *Bacillus* tuvo la más alta actividad, aproximadamente 20-25% mayor actividad que el magnesio y el fosfato solos, como se determinó por SRM (ver Ejemplo 10). La enzima de *Sinorhizobium* tuvo la menor cantidad de actividad, la cual puede estar asociada con problemas de plegamiento y de solubilidad anotados en la expresión. Estas tres enzimas tienen el glutamato en el sitio activo (posición 43 en el sistema de numeración de *B. subtilis*) así como la lisina requerida para la formación de la base de Shiff con piruvato (posición 130); no obstante, la enzima de *B. Subtilis* contiene una treonina en la posición 47, un residuo de sitio activo, en vez de arginina. La KHG de *B. subtilis* es más pequeña y parece estar en un grupo distinto de las enzimas de *S. meliloti* y *E. coli*, con otras enzimas que tienen la treonina del sitio activo. Las diferencias en el sitio activo pueden ser la razón para la actividad incrementada de la

enzima de *B. subtilis*.

Mejoramiento de la Actividad de Aldolasa

Anticuerpos catalíticos pueden ser tan eficientes como las aldolasas naturales, aceptar una amplia gama de sustratos, y pueden ser utilizados para catalizar la reacción mostrada en la Figura 2.

Las aldolasas pueden también ser mejoradas por la evolución dirigida, por ejemplo como se describió previamente, para una aldolasa de KDPG (altamente homóloga a KHG descrita anteriormente) evolucionada por el entremezclado de ADN y la PCR propensa a errores, para eliminar el requerimiento del fosfato y para invertir la enantioselectividad. Los polipéptidos de KDPG de aldolasa son útiles en las reacciones bioquímicas, ya que éstos son altamente específicos para el sustrato ganador (en la presente, piruvato), pero son relativamente flexibles con respecto al sustrato aceptor (por ejemplo indol-3-piruvato) (Koeller & Wong, *Nature* 409: 232-9, 2001). La aldolasa de KHG tiene la actividad para la condensación del piruvato con un número de ácidos carboxílicos. Versiones de mamífero de la aldolasa de KHG se piensa que tienen más amplia especificidad que las versiones bacterianas, incluyendo su actividad sobre la 4-hidroxi-4-metil-2-oxoglutarato y la aceptación de ambos estereoisómeros del 4-hidroxi-2-cetoglutarato. Las fuentes bacterianas parecen tener una preferencia 10 veces mayor por

el isómero R. Existen casi 100 homólogos de KHG disponibles en las bases de datos genómicas, y la actividad ha sido demostrada en *Pseudomonas*, *Paracoccus*, *Providencia*, *Sinorhizobium*, *Morganella*, *E. coli* y en tejidos de mamífero. Estas enzimas pueden ser utilizadas como un punto inicial para diseñar a la medida la enantioespecificidad que es deseada para la producción de monatina.

Las aldolasas que utilizan el piruvato y otro sustrato que es ya sea un cetoácido y/o tiene un grupo hidrofóbico voluminoso como el indol, puede ser "evolucionado" para diseñar a la medida la especificidad del polipéptido, la velocidad y la selectividad del mismo. Además de las aldolasas KHG y ProA demostradas en la presente, los ejemplos de estas enzimas incluyen, pero no están limitados a: aldolasas de KDPG y polipéptidos relacionados (KDPH); transcarboxibenzalpiruvato-hidratasa-aldolasa proveniente de *Nocardioides* st; 4-(2-carboxifenil)-2-oxobut-3-enoato-aldolasa (2'-carboxibenzalpiruvato-aldolasa) que condensa el piruvato y el 2-carboxibenzaldehído (un sustrato que contiene anillo aromático); la trans-O-hidroxibencilidenpiruvato-hidratasa-aldolasa proveniente de *Pseudomonas putida* y *Sphingomonas aromaticivorans*, que también utiliza piruvato y un aldehído que contiene aromático, como sustratos; la 3-hidroxiaspartato-aldolasa (eritro-3-hidroxi-L-aspartato-glioxilato-liasa), que utiliza los 2-oxoácidos como los

sustratos y se piensa que está en el microorganismo *Micrococcus denitrificans*; benzoina-aldolasa (benzaldehído-liasa), que utiliza los sustratos que contienen grupos bencilo; dihidroneopterina-aldolasa; L-treo-3-fenilserina-benzaldehído-liasa (fenilserina-aldolasa) que condensa la glicina con el benzaldehído; 4-hidroxi-2-oxovalerato-aldolasa; 1,2-dihidroxibencilpiruvato-aldolasa; y 2-hidroxibenzalpiruvato-aldolasa.

Un polipéptido que tiene la actividad deseada puede ser seleccionado mediante la selección de los clones de interés utilizando los siguientes métodos. Auxótrofos de triptofano son transformados con vectores que poseen los vectores de interés sobre un casete de expresión y son desarrollados sobre un medio que contiene pequeñas cantidades de monatina o MP. Ya que las reacciones de aminotransferasas y de aldolasas son reversibles, las células son capaces de producir triptofano a partir de una mezcla racémica de monatina. Similarmente, pueden ser seleccionados los organismos (recombinantes y de tipo silvestre) por la habilidad para utilizar MP o la monatina como fuente de carbono y de energía. Una fuente de aldolasas objetivo es las bibliotecas de expresión de diversas *Pseudomonas* y cepas rhizobacterianas. Las *Pseudomonas* tienen muchas vías catabólicas inusuales para la degradación de las moléculas aromáticas y éstas también contienen muchas aldolasas;

220

mientras que las rhizobacterias contienen aldolasas, se sabe que se desarrollan en la rhizosfera vegetal, y tienen muchos de los genes descritos para la construcción de una vía biosintética para la monatina.

Ejemplo 5

Síntesis Química del Precursor de Monatina

El Ejemplo 4 describe un método para utilizar una aldolasa para convertir el indol-3-piruvato al precursor de monatina (MP) ácido 2-hidroxi-2-(indol-3-ilmetil)-4-ceto-glutárico. Este ejemplo describe un método alternativo para sintetizar químicamente MP.

El MP es formado mediante el uso de una condensación típica tipo aldol (Figura 4). En resumen, una reacción típica tipo aldol involucra la generación de un carbanión del éster de piruvato utilizando una base fuerte, tal como LDA (diisopropilamida de litio), hexametildisilazano de litio o butil-litio. El carbanión que se genera reacciona con el indol-piruvato para formar el producto acoplado.

Los grupos protectores que pueden ser utilizados para proteger el nitrógeno del indol incluyen, pero no están limitados a: t-butiloxicarbonilo (Boc), y benciloxicarbonilo (Cbz). Los grupos bloqueadores para los ácidos carboxílicos incluyen, pero no están limitados a, ésteres de alquilo (por ejemplo, ésteres de metilo, etilo, bencilo). Cuando se utilizan tales grupos protectores, no es posible controlar la

estereoquímica del producto que se forma. Sin embargo, si R2 y/o R3 son grupos protectores quirales (Figura 4), tal como el (S)-2-butanol, mentol, o una amina quiral, esto puede favorecer la formación de un enantiómero de MP sobre el otro.

Ejemplo 6

Conversión del Triptofano o del Indol-3-piruvato a Monatina

Un proceso *in vitro* que utiliza dos enzimas, una aminotrasferasa y una aldolasa, produjo la monatina a partir del triptofano y el piruvato. En el primer paso, el alfa-cetoglutarato fue el aceptor del grupo amino proveniente del triptofano en una reacción de transaminación que genera el indol-3-piruvato y glutamato. Una aldolasa catalizó la segunda reacción en la cual el piruvato se hizo reaccionar con indol-3-piruvato, en presencia de Mg^{2+} y fosfato, generando el derivado de alfa-ceto de la monatina (MP), ácido 2-hidroxi-2-(indol-3-ilmetil)-4-cetoglutárico. La transferencia del grupo amino a partir del glutamato formado en la primera reacción, produjo el producto deseado, monatina. La purificación y la caracterización del producto estableció que el isómero formado era S,S-monatina. Los sustratos, enzimas y condiciones alternativas se describen, así como los mejoramientos que fueron realizados a este proceso.

Enzimas

La aldolasa, 4-hidroxi-4-metil-2-oxoglutarato-

piruvato-liasa (aldolasa de ProA, gen *proA*) (EC 4.1.3.17) de *Comamonas testosteroni* fue clonada, expresada y purificada como se describe en el Ejemplo 4). Las 4-hidroxi-2-oxoglutarato-glioxilato-liasas (aldolasas de KHG) (EC 4.1.3.16) proveniente de *B. subtilis, E. coli y S. meliloti* fueron clonadas, expresadas y purificadas como se describe en el Ejemplo 4.

Las aminotransferasas utilizadas en conjunto con las aldolasas para producir la monatina fueron la L-aspartato-aminotransferasa codificada por el gen *aspC de E. coli*, la tirosina-aminotransferasa codificada por el gen *tyrB de E. coli*, la enzima TatA de *S. meliloti*, la aminotrasferasa de sustratos amplios codificada por el gen *bsat de L. major*, o la transaminasa glutámica-oxaloacética de corazón de cerdo (Tipo IIa). La clonación, expresión y purificación de las proteínas no de mamífero se describen en el Ejemplo 1. La transaminasa glutámica-oxaloacética proveniente de corazón de cerdo (tipo IIa) fue obtenida de Sigma (#G7005).

Método que utiliza la aldolasa de ProA y la L-aspartato-aminotransferasa

La mezcla de reacción contenía acetato de amonio 50 mM, pH 8.0, cloruro de magnesio 4 mM, fosfato de potasio 3 mM, fosfato de piridoxal 0.05 mM, piruvato de amonio 100 mM, triptofano 50 mM, alfa-cetoglutarato 10 mM, 160 mg de aldolasa ProA de *C. testosteroni* recombinante (extracto

celular no purificado, aproximadamente 30% de aldolasa), 233 mg de L-aspartato-aminotrasferasa de *E. coli* recombinante (extracto celular no purificado, aproximadamente 40% de aminotransferasa) en un litro. Todos los componentes excepto las enzimas se mezclaron conjuntamente y se incubaron a 30°C hasta que el triptofano se disolvió. Las enzimas fueron luego agregadas y la solución de reacción se incubaron a 30°C con agitación suave (100 rpm) por 3.5 horas. A 0.5 y 1 hora después de la adición de las enzimas, a la reacción se agregaron alícuotas de triptofano sólido (50 mmoles cada una). Todo el triptofano agregado no se disolvió, pero la concentración fue mantenida a 50 mM o mayor. Después de 3.5 horas, el triptofano sólido se filtró. El análisis de la mezcla de reacción por LC/MS utilizando una cantidad definida de triptofano como un estándar, mostró que la concentración de triptofano en la solución fue de 60.5 mM y la concentración de la monatina fue de 5.81 mM (1.05 g).

Se utilizaron los siguientes métodos para purificar el producto final. Noventa por ciento de la solución clara se aplicó a una columna de resina AG50W-X8 de BioRad (225 ml; capacidad de enlace de 1.7 milieq/ml). La columna se lavó con agua, recolectando fracciones de 300 ml, hasta que la absorbancia a 280 nm fue menor de 5% del primer flujo a través de la fracción. La columna fue luego eluida con acetato de amonio 1 M, pH 8.4, recolectando 4 fracciones de

300 ml. Las 4 fracciones contenían monatina y se evaporaron hasta 105 ml utilizando un evaporador rotatorio con un baño de agua tibia. Se formó un precipitado conforme el volumen se reducía y se filtró sobre el curso del proceso de evaporación.

El análisis de las fracciones de la columna por LC/MS mostró que el 99% del triptofano y de la monatina se enlazaron a la columna. El precipitado que se formó durante el proceso de evaporación contenía más de 97% de triptofano y menos de 2% de monatina. La proporción de triptofano al producto en el sobrenadante fue de aproximadamente 2:1.

Se aplicaron 7 ml de sobrenadante a una columna de 100 ml de DEAE Sepharose de Flujo Rápido (Amersham Biosciences) previamente convertida a la forma de acetato por lavado con 0.5 litros de hidróxido de sodio 1 M, 0.2 litros de agua, 1.0 litro de acetato de amonio 1.0 M, pH 8.4 y 0.5 litro de agua. El sobrenadante se cargó a menos de 2 ml/minuto y la columna se lavó con agua a 3-4 ml/minuto hasta que la absorbancia a 280 nm fue de aproximadamente 0. La monatina se eluyó con acetato de amonio 100 mM, pH 8.4, recolectando 4 fracciones de 100 ml.

El análisis de las fracciones mostró que la proporción de triptofano en la monatina en el flujo a través de las fracciones fue de 85:15, y la proporción en las fracciones del eluyente fue de 7:93. Asumiendo el coeficiente de

extinción a 280 nm de la monatina que es el mismo que triptofano, las fracciones del eluyente contenían 0.146 mmol del producto. La extrapolación a la reacción total de 1 litro produciría aproximadamente 2.4 mmoles (aproximadamente 710 mg) de monatina, para una recuperación de 68%.

Las fracciones del eluyente provenientes de la columna de DEAE Sepharose se evaporaron a menos de 20 ml. Una alícuota del producto se purificó adicionalmente por la aplicación a una columna preparativa de fase inversa C_8 utilizando las mismas condiciones cromatográficas que aquellas descritas en el Ejemplo 10 para la caracterización de la monatina a escala analítica. El software Waters Fractionlynx^{MR} fue empleado para disparar la recolección automatizada de las fracciones, de la monatina, con base en la detección del ion m/z=293. Las fracciones provenientes de la columna C_8 con el ion molecular protonado, correspondiente para la monatina, fue recolectada, evaporada hasta sequedad, y luego disuelta en un pequeño volumen de agua. Esta fracción fue utilizada para la caracterización del producto.

El producto resultante fue caracterizado utilizando los siguientes métodos.

Espectroscopia de UV/Visible. Las mediciones espectroscópicas de UV/visible de la monatina producida enzimáticamente fueron llevadas a cabo utilizando un espectrofotómetro Cary 100 Bio de UV/visible. El producto

purificado, disuelto en agua, mostró un máximo de absorción de 280 nm con un hombro a 288 nm, características típicas de los compuestos que contienen indol.

Análisis de LC/MS. Los análisis de las mezclas para la monatina derivada de las reacciones bioquímicas *in vitro* fueron llevados a cabo como se describe en el Ejemplo 10. Un análisis típico de LC/MS de la monatina en una mezcla sintética enzimática *in vitro*, se ilustra en la Figura 5. El panel inferior de la Figura 5 ilustra un cromatograma de iones selectos para el ion molecular protonado de la monatina a m/z=293. Esta identificación de la monatina en la mezcla fue corroborada por el espectro de masa ilustrado en la Figura 6. El análisis del producto purificado por LC/MS mostró un pico simple con un ion molecular de 293 y absorbancia a 280 nm. El espectro de masa fue idéntico a aquel mostrado en la Figura 6.

Análisis de MS/MS. Los experimentos de iones hijos LC/MS/MS, como se describe en el Ejemplo 10, fueron también realizados sobre la monatina. Un espectro de masa de iones hijos de monatina se ilustra en la Figura 7. Asignaciones estructurales tentativas de todos los iones fragmentarios marcados en la Figura 7, fueron realizadas. Estas incluyen los iones fragmentarios de m/z=275 (293-H_2O), 257 (293-(2 x H_2O)), 230 (275-COOH), 212 (257-COOH), 168 (ion indolcarbonio de 3-buta-1,3-dienil-1H), 158 (ion 1H-indol-3-carbaldehído),

144 (ion carbonio de 3-etil-1H-indol), 130 (ion carbonio de 3-metilen-1H-indol), y 118 (ion carbonio del indol). Muchos de éstos son los mismos que aquellos obtenidos para MP (Ejemplo 4), como se esperaba si se deriva de la porción indol de la molécula. Algunos son 1 unidad de masa mayor que aquellos observados por MP, debido a la presencia de un grupo amino en vez de una cetona.

Análisis de MS de Alta Resolución. La Figura 8 ilustra el espectro de masa obtenido para la monatina purificada, empleando un espectrómetro de masa híbrido cuadrupolar/tiempo de vuelo Applied Biosystems-Perkin Elmer Q-Star. La masa medida para la monatina protonada utilizando el triptofano como un estándar de calibración de masa interna fue de 293.1144. La masa calculada de la monatina protonada, con base en la composición elemental, $C_{14}H_{17}N_2O_5$ es 293.1137. Este es un error de medición de masa menor de 2 partes por millón (ppm), proporcionando evidencia concluyente de la composición elemental de la monatina producida enzimáticamente.

Espectroscopía de RMN. Los experimentos de RMN fueron realizados sobre un instrumento Varian Inova de 500 MHz. La muestra de monatina (aproximadamente ~ 3 mg) fue disuelta en 0.5 ml de D_2O. Inicialmente, el solvente (D_2O) fue utilizado como la referencia interna a 4.78 ppm. Ya que el pico para el agua fue grande, la RMN-^{1}H fue corrida con supresión del pico para el agua. Subsecuentemente, debido a la amplitud del pico

de agua, el protón C-2 de la monatina se utilizó como el pico de referencia, y se ajustó al valor publicado de 7.192 ppm.

Para RMN-^{13}C, una corrida inicial de varios cientos de exploraciones indicó que la muestra estaba demasiado diluida para obtener un espectro de ^{13}C adecuado en el tiempo asignado. Por lo tanto, se realizó un experimento de coherencia cuántica múltiple heteronuclear (MC), el cual hizo posible la correlación de los hidrógenos y los carbonos a los cuales éstos están enlazados, y proporcionando también información sobre los desplazamientos químicos de los carbonos.

Un resumen de los datos de ^{1}H y MC se muestra en las Tablas 2 y 3. Por comparación a los valores publicados, los datos de RMN indicaron que la monatina enzimáticamente producida fue ya sea (S,S), (R,R), o una mezcla de ambas.

Análisis Quiral de LC/MS. Para establecer que la monatina producida *in vitro* era un isómero, y no una mezcla de los enantiómeros (R,R) y (S,S), se llevaron a cabo análisis quirales de LC/MS utilizando la instrumentación descrita en el Ejemplo 10.

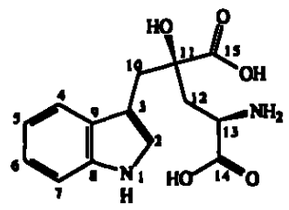
Las separaciones de LC quiral fueron realizadas utilizando una columna de cromatografía quiral Chirobiotic T (Advances Separations Technology) a temperatura ambiente. La separación y detección, basada en los protocolos publicados del vendedor, fueron optimizados para los isómeros R-(D) y S-

229

(L) del triptofano. La fase móvil LC consistió de A) agua que contenía 0.05% (v/v) de ácido trifluoroacético; B) metanol que contenía 0.05% (v/v) de ácido trifluoroacético. La elución fue isocrática a 70% de A y 30% de B. La velocidad de flujo fue de 1.0 ml/minuto, y la absorbancia de PDA fue monitorizada desde los 200 nm hasta 400 nm. Los parámetros instrumentales utilizados para el análisis quiral de LC/MS del triptofano y la monatina son idénticos a aquellos descritos en el Ejemplo 10 para el análisis de LC/MS. Fue utilizada La recolección de espectros de masa para la región m/z 150-400. Los cromatogramas de iones seleccionados para los iones moleculares protonados ($[M+H]^+$=205 para el R- y el S-triptofano y $[M+H]^+$=293 para la monatina) permitieron la identificación directa de estos analitos en las mezclas.

Los cromatogramas del R- y el S-triptofano y la monatina, separados por cromatografía quiral y monitorizados por MS, se muestran en la Figura 9. El pico simple en el cromatograma de la monatina indica que el compuesto es un isómero, con un tiempo de retención casi idéntico al S-triptofano.

Tabla 2: datos de RMN ^1H



230
2nd

	Cargill		Vleggaar et al.[1]		Takeshi et al.[2]	
Atomo	δ_H	J(HH) Hz	δ_H	J(HH) Hz	δ_H	J(HH) Hz
2	7.192 (1H, s)		7.192 (s)		7.18 (s)	
4	7.671 (d)	7.99	7.686 (d)	7.9	7.67 (d)	8.0
5	7.104 (dd)	7.99	7.102 (dd)	8.0, 8.0	7.11 (dd)	7.5, 7.5
6	7.178 (dd)	*	7.176 (dd)	8.0, 8.0	7.17 (dd)	7.5, 7.5
7	7.439 (d)	7.99	7.439 (d)	8.1	7.43 (d)	8.0
10a	3.242 (d)	14.5	3.243 (d)	14.3	3.24 (d)	14.5
10b	3.033 (d)	14.5	3.051 (d)	14.3	3.05 (d)	14.5
12	2.626 (dd)	15.5, 1.5	2.651 (dd)	15.3, 1.7	2.62 (dd)	15.5, 1.8
	2.015 (dd)	15.0, 12.0	2.006 (dd)	15.3, 11.7	2.01 (dd)	15.5, 12.0
13	3.571 (dd)	10.75*, 1.5	3.168 (dd)	11.6, 1.8	3.57 (dd)	12.0, 1.8

[1] Vleggaar et al. (*J.C.S. Perkin Trans.* 1: 3095-8, 1992).

[2] Takeshi y Shusuke (JP2002060382, 2002-02-26).

Tabla 3: datos de RMN-^{13}C (del espectro de HMQC)

	Cargill	Vleggaar et al.[1]
Atomo	δ_C	δ_C
2	126.1	126.03
3	*	110.31
R	120.4	120.46
5	120.2	120.25
6	122.8	122.74
7	112.8	112.79
8	*	137.06
9	*	129.23
10a	36.4	36.56
12	39.5	39.31
13	54.9	54.89
14	*	175.30
15	*	181.18

[1] Vleggaar et al. (*J.C.S. Perkin Trnas.* 1: 3095-8, 1992).

Polarimetría. La rotación óptica fue medida sobre un polarímetro Rudolph Autopol III. La monatina fue preparada como una solución de 14.6 mg/ml en agua. La rotación específica esperada ($[\alpha]_D^{20}$) para la S,S-monatina (forma salina) es de -49.6 para una solución de 1 g/ml en agua (Vleggaar et al.). La $[\alpha]_D^{20}$ observada fue de -28.1 para la monatina enzimáticamente producida, purificada, indicando que éste fue el isómero S,S.

Mejoramientos

Las condiciones de reacción, incluyendo las concentraciones del reactivo y la enzima, fueron optimizadas y fueron producidos rendimientos de 10 mg/ml utilizando la siguiente mezcla de reactivo:

Acetato de amonio 50 mM pH 8.3, cloruro de magnesio 2 mM, piruvato 200 mM (sal de sodio o de amonio), alfa-cetoglutarato 5 mM (sal de sodio), fosfato de piridoxal 0.05 mM, agua desaireada para alcanzar un volumen final de 1 ml después de la adición de las enzimas, fosfato de potasio 3 mM, 50 µg/ml de aldolasa ProA recombinante (extracto celular; concentración de proteína total de 167 µg/ml), 1000 µg/ml de L-aspartato-aminotransferasa modificada por el gen *aspC de E. coli* (extracto celular; concentración de proteína total de 2500 µg/ml), y triptofano sólido para proporcionar una

concentración mayor de 60 mM (saturada; cierta cantidad no disuelta a todo lo largo de la reacción). La mezcla se incubó a 30°C por 4 horas con agitación o mezclado suaves.

Sustituciones

La concentración del alfa-cetoglutarato puede ser reducida a 1 mM y suplementada con aspartato 9 mM con un rendimiento equivalente de monatina. Los aceptores de aminoácidos alternativos pueden ser utilizados en el primer paso, tal como el oxaloacetato.

Cuando se utilizó la aminotransferasa de sustrato amplio de *L. major*, recombinante, en lugar de la L-aspartato-aminotransferasa de *E. coli*, se lograron rendimientos similares de la monatina. No obstante, un segundo producto no identificado (3-10% del producto mayor) con una masa molecular de 292, fue también detectado por el análisis de LC-MS. Las concentraciones de monatina de 0.1-0.5 mg/ml fueron producidas cuando la enzima codificada por *tyrB de E. coli*, la enzima codificada por *tatA de S. meliloti* o la transaminasa glutámica-oxaloacética proveniente de corazón de cerdo (tipo IIa) se agregó como la aminotransferasa. Cuando se comienza la reacción a partir del indol-3-piruvato, se puede realizar una aminación reductiva para el último paso con la glutamato-deshidrogenasa y NADH (como en el Ejemplo 7).

Las aldolasas de KHG de *B. Subtilis, E. coli y s.*

meliloti fueron también utilizadas con la L-aspartato-aminotransferasa de *E. coli* para producir monatina enzimáticamente. Se utilizaron las siguientes condiciones de reacción: acetato de amonio 50 mM, pH 8.3, cloruro de magnesio 2 mM, piruvato 200 mM, glutamato 5 mM, fosfato de piridoxal 0.05 mM, agua desaireada para alcanzar un volumen final de 0.5 ml después de la adición de las enzimas, fosfato de potasio 3 mM, 20 µg/ml de aldolasa de KHG de *B. Subtilis* recombinante (purificada), aproximadamente 400 µg/ml de L-aspartato-aminotransferasa (AspC) de *E. coli* no purificada proveniente del extracto celular, e indol-3-piruvato 12 mM. Las reacciones fueron incubadas a 30°C por 30 minutos con agitación. La cantidad de monatina producida utilizando la enzima de *B. Subtilis* fue de 80 ng/ml, y se incrementó conforme se incrementaron las cantidades de aldolasa. Si el indol-3-piruvato y el glutamato fueron reemplazados por cantidades de saturación del triptofano y alfa-cetoglutarato 5 mM, la producción de la monatina fue incrementada a 360 ng/ml. Las reacciones fueron repetidas con 30 µg/ml de cada una de las tres enzimas KHG en Tris 50 mM pH 8.3, con cantidades saturantes de triptofano, y se dejaron proceder por una hora con el fin de incrementar la detección. La enzima de *Bacillus* tuvo la más alta actividad como en el Ejemplo 4, produciendo aproximadamente 4000 ng/ml de monatina. La KHG de *E. coli* produjo 3000 ng/ml de monatina, y

la enzima de *S. meliloti* produjo 2300 ng/ml.

Ejemplo 7

Interconversión entre MP y Monatina

La aminación de MP para formar la monatina puede ser catalizada por las aminotransferasas tales como aquellas identificadas en los Ejemplos 1 y 6, o por las deshidrogenasas que requieren un cofactor tal como NADH o NADPH. Estas reacciones son reversibles y pueden ser medidas en cualquier dirección. La direccionalidad, cuando se utiliza una enzima deshidrogenasa, puede ser controlada en gran medida por la concentración de las sales de amonio.

Actividad de deshidrogenasa. La desaminación oxidativa de la monatina fue monitorizada siguiendo el incremento en la absorbancia a 340 nm conforme el NAD(P)+ se convirtió al NAD(P)H cromofórico. La monatina fue enzimáticamente producida y purificada como se describe en el Ejemplo 6.

Una mezcla de ensayo típica contenía Tris-HCl 50 mM, pH 8.0 a 8.90, NAD^+ o $NADP^+$ 0.33 mM, 2 a 22 unidades de glutamato-deshidrogenasa (Sigma) y sustrato 10-15 mM en 0.2 ml. El ensayo se realizó por duplicado en una placa de microtitulación transparente a UV, sobre un lector de placas Molecular Devices SpectraMax Plus. Una mezcla de la enzima, amortiguador y $NAD(P)^+$ se vació con pipeta dentro de los pozos que contenían el sustrato y se monitorizó el incremento en la absorbancia a 340 nm a intervalos de 10 segundos después de

un breve mezclado. La reacción se incubó a 25°C por 10 minutos. Los controles negativos fueron llevados a cabo sin la adición del sustrato, y el glutamato fue utilizado como un control positivo. La glutamato-deshidrogenasa del tipo III proveniente de hígado bovino (Sigma # G-7882) catalizó la conversión de la monatina al precursor de monatina a una velocidad de conversión aproximadamente un centésimo de la velocidad de la conversión del glutamato a alfa-cetoglutarato.

Actividad de transaminación. Fueron conducidos ensayos de monatina-aminotransferasa con la aspartato-aminotransferasa (AspC) de *E. coli*, la tirosina-aminotransferasa (TyrB) de *E. coli*, la aminotransferasa de sustrato amplio (BSAT) de *L. major*, y las dos glutamato-oxaloacetato-aminotransferasas porcinas comercialmente disponibles, descritas en el Ejemplo 1. El oxaloacetato y el alfa-cetoglutarato fueron probados como el aceptor de amino. La mezcla de ensayo contenía (en 0.5 ml) Tris-HCl 50 mM, pH 8.0, PLP 0.05 mM, aceptor de amino 5 mM, monatina 5 mM, y 25 µg de aminotransferasa. Los ensayos fueron incubados a 30°C por 30 minutos, y las reacciones fueron detenidas por la adición de 0.5 ml de alcohol isopropílico. La pérdida de monatina fue monitorizada por LC/MS (Ejemplo 10). La más alta cantidad de actividad fue notada con BSAT de *L. major* con oxaloacetato como el aceptor de amino, seguido por la misma

236

enzima con alfa-cetoglutarato como el aceptor de amino. La actividad relativa con el oxaloacetato fue: BSAT>AspC>porcina tipo IIa>porcina tipo I = TyrB. La actividad relativa con el alfa-cetoglutarato fue: BSAT>AspC>porcina tipo I>porcina tipo IIa>TyrB.

Ejemplo 8

Producción de Monatina a partir de Triptofano y Fuentes de C3 Diferentes del Piruvato

Como se describió anteriormente en el Ejemplo 6, el indol-3-piruvato o el triptofano pueden ser convertidos a monatina utilizando piruvato como la molécula de 3 átomos de carbono. No obstante, en algunas circunstancias, el piruvato puede no ser una materia prima deseable. Por ejemplo, el piruvato puede ser más caro que otras fuentes de 3 átomos de carbono, o puede tener efectos adversos sobre las fermentaciones, si se agrega al medio. La alanina puede ser transaminada por muchas enzimas de PLP para producir el piruvato.

Enzimas similares a la triptofanasa realizan las reacciones de beta-eliminación a velocidades más rápidas que otras enzimas PLP tales como las aminotransferasas. Las enzimas provenientes de esta clase (4.1.99.-) pueden producir amoniaco y piruvato a partir de aminoácidos tales como L-serina, L-cisteína y derivados de serina y cisteína con buenos grupos salientes tales como O-metil-L-serina, O-

bencil-L-serina, S-metilcisteína, S-bencilcisteína, S-alquil-L-cisteína, O-acil-L-serina, O-bencil-L-serina, S-metilcisteína, S-bencilcisteína, S-alquil-L-cisteína, O-acil-L-serina, 3-cloro-L-alanina.

Los procesos para producir monatina utilizando los polipéptidos de EC 4.1.99.- pueden ser mejorados mediante la mutación de la β-tirosinasa (TPL) o la triptofanasa de acuerdo al método de la Mouratou et al. (*J. Biol. Chem* 274: 1320-5, 1999). Mouratou et al. describe la habilidad para convertir la β-tirosinasa a una β-liasa de aminoácido dicarboxílico, la cual no se ha reportado que ocurra en la naturaleza. El cambio en la especificidad fue logrado mediante la conversión de la valina (V) 283 a arginina (R) y la arginina (R) 100 a treonina (T). Estos cambios de aminoácidos permiten que la liasa acepte un aminoácido dicarboxílico para la reacción de desaminación hidrolítica (tal como el aspartato). El aspartato, por lo tanto, puede ser utilizado también como una fuente de piruvato para reacciones subsecuentes de condensación de aldol.

Adicionalmente, las células o los reactores enzimáticos pueden ser suministrados con lactato y una enzima que convierte el lactato a piruvato. Los ejemplos de enzimas capaces de catalizar esta reacción incluyen la lactato-deshidrogenasa y la lactato-oxidasa.

238

Aislamiento del ADN Genómico

Los polipéptidos de triptofanasa han sido previamente reportados en, por ejemplo, Mouratou et al. (*JBC* 274: 1320-5, 1999). Para aislar los genes que codifican para los polipéptidos de triptofanasa, se utilizó el ADN genómico proveniente de *E. coli* DH10B como una plantilla para PCR como se describe en el Ejemplo 1.

El gen para la tirosina-fenol-liasa fue aislado de *C. freundii* (ATCC catálogo numero 8090, Designación ATCC 13316; NCTC 9750) y desarrollado sobre agar Nutritivo (Difco 0001) y caldo nutritivo (Difco 0003) a 37°C a una densidad óptica (DO) de 2.0. El ADN genómico fue purificado utilizando un equipo Qiagen Genomic-tip^{MR} 100/G.

Amplificación por PCR de las Secuencias de Codificación

Los cebadores fueron diseñados con grupos sobresalientes compatibles para el vector Xa/LIC de pET 30 (Novagen, Madison, WI) como se describe anteriormente en el Ejemplo 1.

t*na de E. coli* (SEQ ID NO: 41). Cebador N-terminal para la clonación de Xa/LIC de pET 30: 5'-GGT ATT GAG GGT CGC ATG GAA AAC TTT AAA CAT CT-3' (SEQ ID NO: 43). Cebador C-terminal para la clonación de Xa/LIC de pET30: 5'-AGA GGA GAG TTA GAG CCT TAA ACT TCT TTA AGT TTT G-3' (SEQ ID NO: 44).

t*pl de C. freundii* (SEQ ID NO: 42). Cebador N-terminal para la clonación de Xa/LIC de pET30: 5'-GGT ATT GAG GGT CGC

239

ATGAATTATCCGGCAGAACC-3' (SEQ ID NO: 45). Cebador C-terminal para la clonación de Xa/LIC de pET30: 5'-AGA GGA GAG TTA GAG CCTTAGATGTAATCAAAGCGTG-3' (SEQ ID NO: 46).

El Ciclador Térmico Eppendorf Mastercycler[MR] Gradient 5331 fue utilizado para todas las reacciones de PCR. En 50 µl se agregaron 0.5 µg de plantilla (ADN genómico), 1.0 µM de cada cebador, 0.4 mM de cada dNTP, 3.5 U de Polimerasa de Alta Fidelidad de Expansión (Expand High Fidelity Polymerase (Roche), 1X de Amortiguador Expand con magnesio, y 5% de DMSO (concentración final). El programa utilizado de PCR por termociclador fue como sigue: inicio de calentamiento a 96°C (5 minutos), 94°C - 30 segundos, 40-60°C - 1 minuto 45 segundos, 72°C - 2 minutos 15 segundos; 30 repeticiones. El paso de polimerización final fue por 7 minutos, y las muestras fueron luego almacenadas a 4°C.

Clonación

Se utilizaron procedimientos de clonación y de identificación de clones positivos, detallados anteriormente en el Ejemplo 1, para identificar los clones apropiados.

Expresión del Gen y Ensayos de Actividad

El ADN plasmídico (verificado por análisis secuencial) fue subclonado dentro del hospedero de expresión BL21(DE3) (Novagen). Los cultivos fueron desarrollados en medio LB con 30 mg/litro de kanamicina, los plásmidos fueron aislados utilizando un equipo de minipreparación Qiagen, y analizados

mediante digestión de restricción para confirmar la identidad.

Se realizaron experimentos de inducción con el hospedero de expresión BL21(DE3), las construcciones fueron desarrolladas en medio LB que contenía 50 mg/litro de kanamicina a 37°C. La expresión de la proteína fue inducida utilizando IPTG 0.1 mM después de que la DO_{600} del cultivo alcanzó aproximadamente 0.6. Las células fueron desarrolladas por 4 horas a 30°C y cosechadas mediante centrifugación. Las células fueron luego lisadas en 5 ml/g de peso de células húmedas del reactivo BugBuster™ (Novagen) que contenía 5 µl/ml de equipo de coctel de inhibidor de proteasa #III (Calbiochem) y 1 µl/ml de benzonasa-nucleasa (Novagen), y las proteínas recombinantes marcadas con His fueron purificadas utilizando los cartuchos His-Bind como se describe anteriormente en el Ejemplo 1. Las proteínas purificadas fueron desaladas sobre una columna PD-10 (G25 Sephadex, Amersham Biosciences) y eluidas en amortiguador Tris-HCl 100 mM, pH 8.0. Las proteínas fueron analizadas mediante SDS-PAGE sobre geles en gradiente de 4 a 15% para verificar los niveles de proteína soluble al peso molecular predicho de la proteína de fusión recombinante.

Mutagénesis

Algunos miembros del polipéptido clase 4.1.99.- (triptofanasa y β-tirosinasa) realizarán la reacción de beta-

liasa con aspartato o aminoácidos similares sin ninguna modificación. No obstante, algunos miembros de la clase pueden necesitar ser mutagenizados para permitir el uso de sustratos y/o la creación de productos. Además, en algunos casos los polipéptidos que pueden realizar la conversión pueden ser además optimizados mediante mutagénesis.

La mutagénesis dirigida al sitio fue realizada con base en el análisis de la estructura tridimensional de los polipéptidos que se enlazan a PLP. Dos ejemplos para cambiar la especificidad del sustrato de los polipéptidos se muestran enseguida.

Mutagénesis de Triptofanasa Ejemplo 1

El protocolo de mutagénesis proporcionado enseguida introdujo dos mutaciones puntuales en la secuencia de aminoácidos. La primera mutación puntual cambió a arginina (R) en la posición 103 a treonina (T) y la segunda mutación puntual cambió valina (V) en la posición 299 a arginina (R) (sistema de numeración para la proteína madura de *E. coli*). Fueron realizados experimentos de mutagénesis por ATG Laboratories (Eden Prairie, MN). Las mutaciones fueron introducidas secuencialmente por PCR de los fragmentos de genes y el reensamble de los fragmentos fue logrado por PCR también. Cebadores para convertir arginina (R) 103 a treonina (T): 5'-CCAGGGCACCGGCGCAGAGCAAATCTATATT-3' (SEQ ID NO: 47) y 5'-TGCGCCGGTGCCCTGGTGAGTCGGAATGGT-3' (SEQ ID NO: 48).

Cebadores para convertir valina (V)299 a arginina (R): 5'-TCCTGCACGCGGCAAAGGGTTCTGCACTCGGT-3' (SEQ ID NO: 49) y 5'-CTTTGCCGCGTGCAGGAAGGCTTCCCGACA-3' (SEQ ID NO: 50).

Los mutantes fueron seleccionados mediante digestión de restricción con XbaI/HindIII y SphI, y verificados por secuenciamiento.

Mutagénesis de Tirosina-Fenol-Liasa (β-tirosinasa) Ejemplo 2

Se realizaron dos mutaciones puntuales a la secuencia de aminoácidos de la tirosina-fenol-liasa. Estas mutaciones convirtieron la arginina (R) en la posición 100 a treonina (T), y valina (V) en la posición 283 a arginina (R) (en la secuencia de la proteína madura de *C. freundii*).

Los cebadores para conversión de R100T fueron: 5'-AGGGGACCGGCGCAGAAAACCTGTTATCG-3' (SEQ ID NO: 51) y 5'-AGGGGACCGGCGCAGAAAACCTGTTATCG-3' (SEQ ID NO: 52). Los cebadores para la conversión de V282R fueron: 5'-GTTAGTCCGCGTCTACGAAGGGATGCCAT-3' (SEQ ID NO: 53) y 5'-GTAGACGCGGACTAACTCTTTGGCAGAAG-3' (SEQ ID NO: 54).

Los métodos descritos anteriormente fueron utilizados, y los clones fueron seleccionados mediante digestión con KpnI/SacI, y digestión con BstXI. Las secuencias fueron verificadas mediante el secuenciamiento de terminación de cadena didesoxi. La proteína recombinante fue producida como se describió anteriormente para las enzimas de tipo silvestre.

La mezcla de reacción consistió de Tris-HCl 50 mM pH

8.3, cloruro de magnesio 2 mM, fuente 3 átomos de carbono 200 mM, alfa-cetoglutarato 5 mM, sal de sodio, fosfato de piridoxal 0.05 mM, agua desaireada para alcanzar un volumen final de 0.5 ml después de la adición de las enzimas, fosfato de potasio 3 mM pH 7.5, 25 µg de aldolasa ProA de *C. testosteroni* recombinante, cruda, como se describe en el Ejemplo 4, 500 µg de L-aspartato-aminotransferasa cruda (AspC) como se prepara en el Ejemplo 1, y triptofano sólido para proporcionar una concentración mayor de 60 mM (saturada; cierta cantidad no disuelta a todo lo largo de la reacción). La mezcla de reacción fue incubada a 30°C por 30 minutos con mezclado. La serina, la alanina y el aspartato se suministraron como fuente de 3 átomos de carbono. Los ensayos fueron realizados con y sin enzimas PLP secundarias (purificadas) capaces de realizar las reacciones de beta-eliminación y beta-liasa (triptofanasa (TNA), triptofanasa doble mutante, β-tirosinasa (TPL)). Los resultados se muestran en la Tabla 4:

Tabla 4: Producción de monatina utilizando fuentes alternativas de 3 átomos de carbono

Fuente de 3 átomos de carbono	Enzima PLP adicional	Actividad Relativa
ninguna	ninguna	0%
piruvato	ninguna	100%
serina	ninguna	3%
serina	11 µg de TNA de tipo silvestre (1 U)	5.1%

Fuente de 3 átomos de carbono	Enzima PLP adicional	Actividad Relativa
serina	80 µg de TNA doble mutante	4.6%
alanina	ninguna	32%
alanina	11 µg de TNA tipo silvestre	41.7%
alanina	80 µg de TNA mutante	43.9%
aspartato	110 µg de TNA tipo silvestre (10 U)	7.7%
aspartato	5 U de TPL tipo silvestre (crudo)	5.1%
aspartato	80 µg de TNA mutante	3.3%

La monatina producida a partir de la alanina y la serina como fuentes de 3 átomos de carbono fue verificada por el análisis de exploración hija LC/MS/MS, y fue idéntica a la monatina caracterizada producida en el Ejemplo 6. La alanina fue la mejor alternativa probada, y fue transaminada por la enzima AspC. La cantidad de monatina producida fue incrementada por la adición de la triptofanasa, la cual es capaz de realizar la transaminación como una actividad secundaria. La cantidad de monatina producida con la serina como una fuente de carbono casi se duplicó con la adición de las enzimas triptofanasas, aún cuando únicamente un quinto de la cantidad de triptofanasa fue agregado en comparación a la aminotransferasa. AspC es capaz de realizar cierta cantidad de actividad de beta-eliminación sola. Los resultados con el aspartato indican que la actividad de triptofanasa sobre el aspartato no se incrementa con las mismas mutagénesis

245
24

dirigidas al sitio que las que se sugirieron previamente para la β-tirosinasa. Se espera que la β-tirosinasa mutante tenga mayor actividad para la producción de monatina.

Ejemplo 9

Síntesis Química de Monatina

La adición de la alanina al ácido indol-3-pirúvico produce monatina, y esta reacción puede ser realizada sintéticamente con un reactivo de Grignard o de organolitio.

Por ejemplo, a la 3-cloro- o -3-bromo-alanina que ha sido apropiadamente bloqueada en los grupos carboxilo y amino, se agrega magnesio bajo condiciones anhidras. El indol-3-piruvato (apropiadamente bloqueado) es luego agregado para formar el producto acoplado, seguido por la eliminación de los grupos protectores para formar la monatina. Grupos protectores que son particularmente útiles incluyen THP (éter tetrahidropiranílico) que es fácilmente acoplado y eliminado.

Ejemplo 10

Detección de la Monatina y MP

Este ejemplo describe los métodos utilizados para detectar la presencia de monatina, o su precursor el ácido 2-hidroxi-2-(indol-3-ilmetil)-4-cetoglutárico.

Análisis de LC/MS

Los análisis de las mezclas para la forma del alfa-cetoácido de la monatina (precursor de la monatina, MP) y la

monatina derivada de las reacciones bioquímicas *in vitro* o *in vivo* se realizaron utilizando un instrumento de espectroscopía de masa en tandem con cromatografía líquida Waters/Micromass (LC/MS/MS) incluyendo un cromatógrafo de líquidos Waters 2690 con un monitor de absorbancia de Arreglo de Foto-Diodo (PDA) Waters 996, colocado en serie entre el cromatógrafo y un espectrómetro de masa cuadrupolar triple Micromass Quattro Ultima. Se realizaron separaciones de LC utilizando una columna de cromatografía de fase inversa Supelco Discovery C_{18}, de 2.1 mm x 150 mm, o una columna de cromatografía de fase inversa Xterra MS C_8, de 2.1 mm x 250 mm, a temperatura ambiente. La fase móvil de LC consistió de A) agua que contenía 0.05% (v/v) de ácido trifluoroacético y B) metanol que contenía 0.05% (v/v) de ácido trifluoroacético.

La elución en gradiente fue lineal desde 5% de B a 35% de B, 0-9 minutos, lineal desde 35% de B a 90% de B, 9-16 minutos, isocrático a 90% de B, 16-20 minutos, lineal desde 90% de B hasta 5% de B, 20-22 minutos, con un periodo de re-equilibrio de 10 minutos entre las corridas. La velocidad de flujo fue de 0.25 ml/minutos, y la absorbancia de PDA fue monitorizada desde 200 nm hasta 400 nm. Todos los parámetros del ESI-MS fueron optimizados y seleccionados con base en la generación de los iones moleculares protonados ($[M+H^+]$) de los analitos de interés, y producción de los iones de los

247

fragmentos característicos.

Los parámetros instrumentales siguientes fueron utilizados para el análisis de LC/MS de la monatina: Capilar: 3.5 kV; Cono: 40 v; Hex 1: 20 V; Abertura: 0 V; Hex 2: 0 V; Temperatura de la fuente: 100°C; temperatura de desolvatación: 350°C; gas de desolvatación: 500 litro/hora; gas del cono: 50 litro/hora; baja resolución de masa (Q1): 15.0; alta resolución de masa (Q1): 15.0; energía iónica: 0.2; entrada: 50 V; energía de colisión: 2; salida: 50 V; baja resolución de masa (Q2): 15; alta resolución de masa (Q2): 15; energía iónica (Q2): 3.5; Multiplicador: 650. Las incertidumbres para las proporciones reportadas de masa/carga (m/z) y las masas moleculares son ± 0.01%. La detección inicial de la forma de alfa-cetoácido de la monatina (MP) y la monatina en las mezclas fue lograda por monitoreo de LC/MS con recolección de espectros de masa para la región m/z 150-400. Los cromatogramas de iones seleccionados para los iones moleculares protonados ([M+H]=292 para MP, $[M+H]^+$=293 para monatina) permitieron la identificación directa de estos analitos en las mezclas.

Análisis de MS/MS

Se realizaron experimentos de iones hijos LC/MS/MS sobre la monatina, como sigue. Un análisis de iones hijos involucra la transmisión del ion progenitor (por ejemplo m/z=293) para la monatina) de interés a partir del primer

analizador de masa (Q1) hacia la celda de colisión del espectrómetro de masa, donde se introduce argón y disocia químicamente el progenitor en iones fragmentarios (hijos). Estos iones fragmentarios son luego detectados con el segundo analizador de masa (Q2), y pueden ser utilizados para corroborar la asignación estructural del progenitor.

Fueron utilizados los siguientes parámetros instrumentales para el análisis de LC/MS/MS de la monatina: Capilar: 3.5 kV; Cono: 40 V; Hex 1: 20 V; Abertura: 0 V; Hex 2: 0 V; Temperatura de la Fuente: 100°C; Temperatura de desolvatación: 350°C; Gas de desolvatación: 500 litros/hora; gas del cono: 50 litros/hora; Baja resolución de masa (Q1): 13.0; Alta resolución de masa (Q1): 13.0; energía iónica: 0.2; Entrada: -5 V; Energía de colisión: 14; Salida: 1 V; Baja resolución de masa (Q2): 15; Alta resolución de masa (Q2): 15; Energía iónica (Q2): 3.5; Multiplicador: 650.

Determinación de Alto Rendimiento de la Monatina

Los análisis de alto rendimiento (<5 minutos/muestra) de las mezclas para la monatina derivada de las reacciones *in vitro* o *in vivo*, se llevaron a cabo utilizando la instrumentación descrita anteriormente, y los mismos parámetros que los que se describen para LC/MS/MS. Se realizaron separaciones de LC utilizando la cromatografía Waters Xterra MS C_8 (2.1 mm x 50 mm) a temperatura ambiente con elución isocrática en metanol acuoso al 15%, ácido

249

acético al 0.25% a una velocidad de flujo de 0.3 ml/minuto. La detección de la monatina en las mezclas se logró utilizando la espectrometría de masa en tandem (SRM) monitoreando la reacción seleccionada. Esto involucró el monitoreo específico del ion progenitor inducido por colisiones ($[M+H]^+$=293.1) a las transiciones del ion hijo (p. ej., el ion fragmentario a m/z = 168.1, tentativamente asignado como un ion carbonio de 3-buta-1,3-dienil-1H-indol) para elevar al máximo la sensibilidad, la selectividad y el rendimiento para la detección de la monatina. Los datos de absorbancia de PDA fueron recolectados en paralelo para la verificación posterior de la identidad de la monatina.

Ejemplo 11

Producción de Monatina en Bacterias

Este ejemplo describe los métodos utilizados para producir la monatina en las células de *E. coli*. Una persona de experiencia en la técnica comprenderá que pueden ser utilizados métodos similares para producir la monatina en otras células bacterianas. Además, pueden ser utilizados los vectores que contienen otros genes en la vía de la síntesis de la monatina (Figura 2).

El medio de Trp-1 + glucosa, un medio mínimo que ha sido utilizado para la producción incrementada de triptofano en células de *E. coli* (Zeman et al. *Folia Microbiol.* 35: 200-4, 1990), fue preparado como sigue. A 700 ml de agua nanopura

se agregaron los siguientes reactivos: 2 g de sulfato de amonio, 13.6 g de fosfato diácido de potasio, 0.2 g de sulfato de magnesio heptahidratado, 0.01 g de cloruro de calcio bihidratado, y 0.5 mg de sulfato ferroso heptahidratado. El pH fue ajustado a 7.0, el volumen fue incrementado a 850 ml, y el medio fue calentado en autoclave. Se preparó una solución de glucosa al 50% separadamente, y se esterilizó por filtración. Se agregaron cuarenta ml al medio base (850 ml) para un volumen final de 1 litro.

Una solución de 10g/litro de L-triptofano fue preparada en fosfato de sodio 0.1 M pH 7, y se esterilizó por filtración. Un décimo del volumen fue típicamente agregado a los cultivos como se especificó más adelante. Se preparó también una solución de piruvato de sodio al 10% y se esterilizó por filtración. Se utilizó típicamente una alícuota de 10 ml por litro de cultivo. Se prepararon reservas de ampicilina (100 mg/ml), kanamicina (25 mg/ml) e IPTG (840 mM), se esterilizaron por filtración, y se almacenaron a -20°C antes del uso. Se utilizó Tween 20 (polioxietileno 20-monolaurato de sorbitán) a una concentración final de 0.2% (volumen/volumen). Se utilizó ampicilina a concentraciones no letales, típicamente una concentración final de 1-10 µg/ml.

Placas frescas de BL21(DE3) de *E. coli::proA de C. testosteroni*/pET30 Xa/LIC (descrita en el Ejemplo 4) se

prepararon sobre medio LB que contenía 50 µg/ml de kanamicina. Los cultivos de toda la noche (5 ml) fueron inoculados a partir de una colonia simple y desarrollados a 30°C en medio LB con kanamicina. Típicamente, se utilizó de 1 a 50 inóculos para la inducción en el medio de trp-1 + glucosa. El antibiótico fresco fue agregado a una concentración final de 50 mg/litro. Matraces en agitación se desarrollaron a 37°C antes de la inducción.

Las células fueron muestreadas cada hora hasta que se obtuvo una DO_{600} de 0.35-0.8. Las células fueron luego inducidas con IPTG 0.1 mM, y la temperatura reducida a 34°C. Se recolectaron muestras de 1 ml antes de la inducción (punto de tiempo cero) y se centrifugaron a 5000 x g. El sobrenadante se congeló a -20°C para el análisis de LC/MS. Cuatro horas después de la inducción, se recolectó otra muestra de 1 ml, y se centrífugo para separar el caldo del botón celular. Se agregaron como se describió anteriormente, triptofano, piruvato de sodio, ampicilina y Tween.

Las células fueron desarrolladas por 48 horas después de la inducción, y se tomó otra muestra de 1 ml y se preparó como se describe anteriormente. A las 48 horas, se agregó otra alícuota más de triptofano y de piruvato. El volumen de cultivo completo fue centrifugado después de aproximadamente 70 horas de crecimiento (post-inducción), por 20 minutos a 4°C y 3500 rpm. El sobrenadante fue decantado y el caldo y

las células fueron congeladas a -80°C. Las fracciones de caldo fueron filtradas y analizadas por LC/MS. Las alturas y las áreas de los picos de $[M+H]^+$=293, fueron monitorizadas como se describe en el Ejemplo 10. El nivel antecedente del medio fue sustraído. Los datos fueron también normalizados para el desarrollo celular al trazar gráficamente la altura del pico $[M+H]^+$=293 dividido entre la densidad óptica del cultivo a 600 nm.

Se produjeron niveles más altos de monatina cuando se agregaron piruvato, ampicilina y Tween, 4 horas después de la inducción, en vez de en la inducción. Otros aditivos tales como PLP, fosfato adicional o cloruro de magnesio adicional no incrementaron la producción de monatina. Se obtuvieron títulos más altos de monatina cuando se utilizó triptofano en vez del indol-3-piruvato, y cuando el triptofano fue agregado después de la inducción en vez de en la inoculación o en la inducción. Antes de la inducción, y 4 horas después de la inducción (al tiempo de la adición del sustrato), no existió típicamente nivel detectable de la monatina en el caldo de fermentación o en los extractos celulares. Se realizaron controles negativos utilizando células con el vector pET30a únicamente, así como cultivos donde no se agregaron triptofano y piruvato. Una exploración MS progenitora demostró que el compuesto con (m+1)/z=293 no fue derivado de moléculas más grandes, y las exploraciones hijas (realizadas

como en el Ejemplo 10, fueron similares a la monatina elaborada *in vitro*.

El efecto de Tween fue estudiado mediante la utilización de concentraciones finales de 0, 0.2% (volumen/volumen) y 0.6% de Tween 20. La más alta cantidad de monatina producida por los matraces en agitación fue a 0.2% de Tween. La concentración de ampicilina fue variada entre 0 y 10 µg/ml. La cantidad de monatina en el caldo celular se incrementó rápidamente (2.5 X) entre 0 y 1 µg/ml, y se incrementó 1.3 X cuando la concentración de ampicilina fue incrementada de 1 a 10 µg/ml.

Un experimento de curso de tiempo que muestra los resultados típicos, se muestra en la Figura 10. La cantidad de monatina secretada hacia el caldo celular se incrementó, aun cuando los valores son normalizados para el desarrollo celular. Mediante el uso del coeficiente de extinción molar del triptofano, se estimó que la cantidad de monatina en el caldo era menor de 10 µg/ml. Se repitió el mismo experimento con las células que contenían el vector sin el inserto *proA*. Muchos de los números fueron negativos, indicando que a altura máxima a m/z=293 fue menor en estos cultivos que en el medio solo (Figura 10). Los números fueron consistentemente menores cuando estuvieron ausentes el triptofano y el piruvato, demostrando que la producción de monatina es un resultado de una reacción enzimática catalizada por la enzima

aldolasa.

La producción *in vivo* de la monatina en células bacterianas fue repetida en experimentos con matraces de agitación de 800 ml y en fermentadores. Una muestra de 250 ml de monatina (en el caldo libre de células) fue purificada mediante cromatografía de intercambio aniónico y cromatografía líquida preparativa de fase inversa. Esta muestra fue evaporada, y enviada para el análisis de masa de alta resolución (descrito en el Ejemplo 6). La MS de alta resolución indicó que el metabolito que es producido es la monatina.

Los ensayos *in vitro* indican que la aminotransferasa necesita estar presente a niveles más altos que la aldolasa (ver Ejemplo 6), por lo tanto, la aspartato-aminotransferasa proveniente de *E. coli* fue sobreexpresada en combinación con el gen de la aldolasa para incrementar la cantidad de monatina producida. Fueron diseñados cebadores para introducir *proA de C. testosteroni* dentro de un operón con *aspC*/pET30 Xa/LIC, como sigue: cebador 5': ACTCGGATCCGAAGGAGATATACATATGTACGAACTGGGACT (SEQ ID NO: 67) y cebador 3': CGGCTGTCGACCGTTAGTCAATATATTTCAGGC (SEQ ID NO: 68). El cebador 5' contiene un sitio BamHI, el cebador 3' contiene un sitio SalI para la clonación. Se realizó la PCR como se describe en el Ejemplo 4, y se purificó el gel. La construcción *aspC*/pET30 Xa/LIC fue digerida con BamHI y SalI,

y fue el producto de PCR. Las digestiones fueron purificadas utilizando una columna giratoria Qiagen. El producto de PCR de *proA* fue ligado al vector utilizando el equipo de Ligadura de ADN Roche Rapid (Indianápolis, IN) de acuerdo a las instrucciones del fabricante. Fueron realizadas transformaciones químicas utilizando Novablues Singles (Novagen) como se describe en el Ejemplo 1. Las colonias fueron desarrolladas en medio LB que contenía 50 mg/litro de kanamicina, y el ADN plasmídico fue purificado utilizando el equipo de minipreparación giratoria Qiagen. Los clones fueron seleccionados mediante análisis de digestión por restricción y la secuencia se confirmó mediante Seqwright (Houston, Texas). Las construcciones se subclonaron dentro de BLR(DE3), BLR(DE3)pLysS, BL21(DE3) y BL21(DE3)pLysS (Novagen). La construcción *proA*/pET30 Xa/LIC fue también transformada en BL21(DE3)pLysS.

Las comparaciones iniciales de las muestras de BLR(DE3) en matraces con agitación, bajo las condiciones estándares descritas anteriormente, demostraron que la adición del segundo gen (*aspC*) mejoró la cantidad de monatina producida, en siete veces. Para acelerar el desarrollo, se utilizaron cepas hospederas derivadas de BL21(DE3). Los clones de *proA* y los dos clones del operón del gen fueron inducidos en el medio Trp-1 como se describió anteriormente, los hospederos de pLysS tuvieron también cloranfenicol (34

mg/litro) agregado al medio. Los experimentos con matraces de agitación fueron realizados con y sin la adición de Tween 20 al 0.2% y 1 mg/litro de ampicilina. La cantidad de monatina en el caldo fue calculada utilizando monatina purificada, producida *in vitro* como un estándar. Los análisis de SRM fueron realizados como se describe en el Ejemplo 10. Las células fueron muestreadas a cero, 4 horas, 24 horas, 48 horas, 72 horas y 96 horas de crecimiento.

Los resultados se muestran en la Tabla 5 para las cantidades máximas producidas en los caldos de cultivo. En la mayoría de los casos, las dos construcciones génicas tuvieron valores más altos que la construcción de *proA* sola. Las cepas pLysS, las cuales deberían tener envolturas celulares con más fugas, tuvieron más altos niveles de monatina secretada, aun cuando estas cepas se desarrollan típicamente a una velocidad más lenta. Las adiciones de Tween y ampicilina fueron benéficas.

Tabla 5: Cantidad de Monatina Producida por Bacterias *E. coli*

Construcción	Hospedero	Tween + Amp	µg/ml de monatina	Tiempo
proA	BL21(DE3)	-	0.41	72 horas
proA	BL21(DE3)	+	1.58	48 horas
proA	BL21(DE3)pLysS	-	1.04	48 horas
proA	BL21(DE3)pLysS	+	1.60	48 horas
aspC:proA	BL21(DE3)	-	0.09	48 horas
aspC:proA	BL21(DE3)	+	0.58	48 horas
aspC:proA	BL21(DE3)pLysS	-	1.39	48 horas
aspC:proA	BL21(DE3)pLysS	+	6.68	48 horas

257

Ejemplo 12

Producción de Monatina en Levadura

Este ejemplo describe los métodos utilizados para producir la monatina en células eucarióticas. Una persona de experiencia en la técnica comprenderá que pueden ser utilizados métodos similares para producir la monatina en cualquier célula de interés. Además, pueden ser utilizados otros genes (por ejemplo, aquellos listados en la Figura 2) además de, o alternativamente a aquellos descritos en este ejemplo.

Se utilizó el Sistema de Vector de Marcación de Epitopos de Levadura pESC (Stratagene, La Jolla, Ca) para clonar y expresar los genes *aspC de E. coli* y *proA de C. testosteroni*, dentro de *Saccharomyces cerevisiae*. Los vectores pESC contienen los promotores GAL1 y GAL10 sobre hebras opuestas, con dos sitios de clonación múltiples distintos, permitiendo la expresión de dos genes al mismo tiempo. El vector pESC-His también contiene el gen *His3* para la complementación de la auxotrofía de la histidina en el hospedero (YPH500). Los promotores de GAL1 y GAL10 son reprimidos por la glucosa e inducidos por la galactosa; se utiliza una secuencia Kozak para la expresión óptima en levadura. Los plásmidos pESC son vectores lanzadera, permitiendo que la construcción inicial sea realizada en *E. coli* (con el gen *bla* para la selección); no obstante, no

están presentes sitios de enlace al ribosoma bacteriano, en los sitios de clonación múltiple.

Los siguientes cebadores fueron diseñados para la clonación de pESC-His (los sitios de restricción están subrayados, la secuencia Kozak está en negritas): *aspC* (*Bam*HI/*Sal*I), GAL1: 5'-CGCGGATCC**ATAATGG**TTGAGAACATTACCG-3' (SEQ ID NO: 69) y 5'-ACGCGTCGACTTACAGCACTGCCACAATCG-3' (SEQ ID NO: 70). *proA* (*Eco*RI/*Not*I), GAL10: 5'-CCGGAATTC**ATAATGG**TCGAACTGGGAGTTGT-3' (SEQ ID NO: 7) y 5'-GAATGCGGCCGCTTAGTCAATATATTTCAGGCC-3' (SEQ ID NO: 72).

El segundo codón para ambas proteínas maduras fue cambiado de un aminoácido aromático a valina debido a la introducción de la secuencia Kozak. Los genes de interés fueron amplificados utilizando el ADN · miniprep de pET30 Xa/LIC a partir de los clones descritos en el Ejemplo 1 y Ejemplo 4 como plantilla. Se realizó la PCR utilizando un termociclador de gradiente ciclador Eppendorf Master y el siguiente protocolo para una reacción de 50 µl. 1.0 µl de plantilla, 1.0 µM de cada cebador, 0.4 mM de cada dNTP, 3.5 U de Polimerasa de Alta Fidelidad de Expansión (Roche, Indianápolis, IN), y el amortiguador 1X Expand^{MR} con magnesio. El programa termociclador utilizado consistió de un inicio en caliente a 94°C por 5 minutos, seguido por 29 repeticiones de los siguientes pasos: 94°C por 30 segundos, 50°C por 1 minuto 45 segundos, y 72°C por 2 minutos 15 segundos. Después de las

29 repeticiones la muestra se mantuvo a 72°C por 10 minutos y luego se almacenó a 4°C. Los productos de PCR fueron purificados mediante separación sobre un gel de TAE-agarosa al 1%, seguido por la recuperación utilizando el Equipo de Extracción en Gel QIAquick (Qiagen, Valencia, CA).

2.7 µg del ADN vector de pESC-His se digirieron con *BamHI/SalI* y se purificaron en gel como se describió anteriormente. El producto de PCR de *aspC* fue digerido con *BamHI/SalI* y se purificó con una Columna de Purificación de PCR QIAquick. Las ligaduras fueron realizadas con el Equipo de Ligadura de ADN Roche Rapid siguiendo los protocolos del fabricante. Las ligaduras desaladas fueron sometidas a electoporesis dentro de 40 µl de células competentes Electromax DH10B (Invitrogen) en una cubeta desechable Biorad de 0.2 cm, utilizando un Pulsador de Genes II Biorad con controlador de pulso plus, de acuerdo a las instrucciones del fabricante. Después de 1 hora de recuperación en 1 ml de medio SOC, los transformantes fueron sembrados en placa sobre medio LB que contenía 100 µg/ml de ampicilina. Las preparaciones de ADN plasmídico para los clones fueron realizadas utilizando los Equipos Giratorios Miniprep de QIAprep. El ADN plasmídico fue seleccionado mediante digestión por restricción, y secuenciado (Seqwright) para la verificación, utilizando los cebadores diseñados para el vector.

El clon *aspC*/pESC-His fue digerido con *Eco*RI y *Not*I, como lo fue el producto de PCR *proA*. El ADN fue purificado como se describió anteriormente, y se ligó como se describió anteriormente. La construcción de dos genes fue transformada dentro de células DH10B y seleccionada mediante digestión por restricción y secuenciamiento de ADN.

La construcción fue transformada dentro de *S. cerevisiae* cepa YPH500 utilizando el Equipo de Transformación S.c. EasyComp^{MR} (Invitrogen). Las reacciones de transformación fueron sembradas en placa sobre medio mínimo SC-His (Invitrogen pYES2 manual) que contenía 2% de glucosa. Las colonias de levadura individuales fueron seleccionadas para la presencia de los genes *proA y aspC* mediante PCR de colonias utilizando los cebadores de PCR anteriores. Las células concentradas (2 µl) fueron suspendidas en 20 µl de Amortiguador de Y-Lisis (Zymo Research) que contenía 1 µl de zimolasa y se calentó a 37°C por 10 minutos. Cuatro µl de esta suspensión fueron luego utilizados en una reacción de PCR de 50 µl utilizando la mezcla de reacción de PCR y el programa descrito anteriormente.

Se desarrollaron cultivos de cinco ml toda la noche sobre SC-His + glucosa a 30°C y 225 rpm. Las células fueron gradualmente ajustadas para el desarrollo sobre rafinosa, con el fin de reducir al mínimo el periodo de espera antes de la inducción con galactosa. Después de aproximadamente 12 horas

de crecimiento, se tomaron las mediciones de absorbancia a 600 nm, y un volumen apropiado de las células fue centrifugado y resuspendido para dar una DO de 0.4 en el medio SC-His fresco. Se utilizaron las siguientes fuentes de carbono secuencialmente: 1% de rafinosa + 1% de glucosa, 0.5% de glucosa + 1.5% de rafinosa, 2% de rafinosa y finalmente 1% de rafinosa + 2% de galactosa para la inducción.

Después de aproximadamente 16 horas de crecimiento en el medio de inducción, cultivos de 50 ml se dividieron en cultivos de 25 ml duplicados, y se agregaron los siguientes únicamente a uno de los duplicados: (concentraciones finales) 1 g/litro de L-triptofano, fosfato de sodio 5 mM pH 7.1, 1 g/litro de piruvato de sodio, cloruro de magnesio 1 mM. Las muestras de los caldos y los botones celulares provenientes del medio de no inducción, y de los cultivos de 16 horas antes de la adición de los sustratos para la vía de la monatina, se guardaron como controles negativos. Además, las construcciones que contenían únicamente el gen *aspC* funcional (y un gen *proA* truncado) se utilizaron como otro control negativo. Las células se dejaron desarrollar por un total de 69 horas después de la inducción. Ocasionalmente, las células de levadura fueron inducidas a una DO menor, y únicamente se desarrollaron por 4 horas antes de la adición del triptofano y el piruvato. No obstante, estos sustratos de monatina parecen inhibir el desarrollo, y la adición a DO más alta fue

más efectiva.

Los botones celulares provenientes de los cultivos fueron lisados con 5 ml de Yeastbuster^{MR} + 50 µl de THP (Novagen) por gramo (peso húmedo) de las células siguiendo los protocolos del fabricante, con la adición de inhibidores de la proteasa y benzonasa-nucleasa como se describió en los ejemplos previos. El caldo de cultivo y los extractos celulares fueron filtrados y analizados mediante SRM como se describe en el Ejemplo 10. Utilizando este método, no se detectó la monatina en las muestras de caldo, indicando que las células no pudieron secretar la monatina bajo estas condiciones. La fuerza motriz de protones puede ser suficiente bajo estas condiciones o los transportadores generales de aminoácidos pueden ser saturados con triptofano. La expresión de proteína no fue a un nivel que permitiera la detección de los cambios utilizando SDS-PAGE.

La monatina fue detectable (aproximadamente 60 µg/ml) transitoriamente en extractos celulares del cultivo con dos genes funcionales, cuando el triptofano y el piruvato fueron agregados al medio. La monatina no fue detectada en ninguno de los extractos de células control negativas. Los ensayos *in vitro* para la monatina fueron realizados por duplicado con 4.4 mg/ml de proteína total (aproximadamente el doble de lo que se utiliza típicamente para extractos celulares de *E. coli*) utilizando el ensayo optimizado descrito en el Ejemplo

263

6. Se realizaron otros ensayos con la adición ya sea de 32 µg/ml de aldolasa ProA de *C. testosteroni* o 400 µg/ml de aminotransferasa AspC, para determinar cuál enzima era limitante en el extracto celular. Los controles negativos fueron realizados sin la adición de enzima, o la adición únicamente de la aminotransferasa AspC (la condensación de aldol puede ocurrir en cierto grado sin enzima). Los controles positivos fueron realizados con enzimas parcialmente puras (30-40%), utilizando 16 µg/ml de aldolasa y 400 µg/ml de aminotransferasa.

Los resultados *in vitro* fueron analizados mediante SRM. El análisis de los extractos celulares mostró que el triptofano fue efectivamente transportado hacia las células cuando éste fue agregado al medio después de la inducción, dando como resultado niveles de triptofano dos órdenes de magnitud mayores que aquellos en los cuales no se agregó triptofano adicional. Los resultados para el análisis de monatina *in vitro* son mostrados en la Tabla 6 (los números indican ng/ml).

Tabla 6: Producción de monatina con extractos de células de levadura

	Construcción de *aspC*	+ aldolasa	+ aspC	Construcción de dos genes	+ aldolasa	+AspC
reprimido (medio de glucosa)	0	888.3	173.5	0	485.2	829
inducido 24 horas	0	2832.8	642.4	0	1375.6	9146.8

	Construcción de *aspC*	+ aldolasa	+ aspC	Construcción de dos genes	+ aldolasa	+AspC
Inducido 69 horas	0	4937.3	340.3	71.9	1652.8	23693.5
69 horas + subs.	0	556.9	659.1	21.9	755.6	16688.2
+ control (enzimas purificadas)	21853			21853		
- control (sin enzimas)	0		254.3	0		254.3

Se obtuvieron resultados positivos con los extractos celulares de la construcción completa de dos genes, con y sin el sustrato agregado al medio de crecimiento. Estos resultados, en comparación a los controles positivos, indican que las enzimas fueron expresadas a niveles cercanos al 1% de la proteína local en la levadura. La cantidad de monatina producida cuando el extracto celular de la construcción *aspC* (con *proA* truncada) fue evaluada con la aldolasa, fue significativamente mayor que cuando los extractos celulares fueron evaluados solos, e indica que la aminotransferasa AspC recombinante comprende aproximadamente 1-2% de la proteína total de levadura. Los extractos celulares de los cultivos no inducidos tuvieron una pequeña cantidad de actividad cuando se evaluaron con la aldolasa debido a la presencia de aminotransferasas nativas en las células. Cuando se evaluó con la aminotransferasa AspC, la actividad de los extractos provenientes de la células no inducidas, se incrementó a la cantidad de monatina producida por el control negativo con AspC (aproximadamente 200 ng/ml). En contraste, la actividad

observada cuando se evalúa el extracto celular de construcción de dos genes, se incrementa más cuando la aminotransferasa es suplementada que cuando la aldolasa es agregada. Ya que ambos genes deben ser expresados al mismo nivel, esto indica que la cantidad de monatina producida es elevada al máximo cuando el nivel de aminotransferasa es más alto que aquel de la aldolasa, en concordancia con los resultados mostrados en el Ejemplo 6.

La adición de triptofano y piruvato no solamente inhibe el desarrollo celular, sino que aparentemente inhibe también la expresión de la proteína. La adición del plásmido pESC-Trp puede ser utilizada para corregir la auxotrofía de triptofano de las células hospederas YPH500, para proporcionar un medio de suministrar el triptofano con menos efectos sobre el desarrollo, la expresión y la secreción.

Ejemplo 13

Mejoramiento de los Procesos Enzimáticos Utilizando Reacciones Acopladas

En teoría, si no ocurren las reacciones colaterales o la degradación de los sustratos o los intermediarios, la cantidad máxima de producto formado a partir de la reacción enzimática ilustrada en la Figura 1, es directamente proporcional a las constantes en el equilibrio de cada reacción, en las concentraciones de triptofano y piruvato. El triptofano no es un sustrato altamente soluble, y las

766

concentraciones de piruvato mayores de 200 mM parecen tener un efecto negativo sobre el rendimiento (ver Ejemplo 6).

Idealmente, la concentración de monatina es elevada al máximo con respecto a los sustratos, con el fin de disminuir el costo de la separación. Las separaciones físicas pueden ser realizadas tal que la monatina es removida de la mezcla de reacción, previniendo que ocurran las reacciones inversas. Las materias primas y los catalizadores pueden ser luego generados. Debido a la similitud de la monatina en tamaño, carga e hidrofobicidad a varios de los reactivos e intermediarios, las separaciones físicas serán difíciles a no ser que exista una alta cantidad de afinidad por la monatina (tal como una técnica de cromatografía de afinidad). No obstante, las reacciones de monatina pueden ser acopladas a otras reacciones tal que el equilibrio del sistema es desplazado hacia la producción de monatina. Los siguientes son ejemplos de procesos para mejorar el rendimiento de la monatina obtenida a partir de triptofano o indol-3-piruvato.

Reacciones acopladas utilizando oxaloacetato-descarboxilasa (EC 4.1.1.3)

La Figura 11 es una ilustración de la reacción. La oxidasa y la catalasa de triptofano son utilizadas para impulsar la reacción en la dirección de la producción de indol-3-piruvato. La catalasa es utilizada en exceso, tal que el peróxido de hidrógeno no está disponible para reaccionar

en la dirección inversa, o para dañar las enzimas o los intermediarios. El oxígeno es regenerado durante la reacción de catalasa. Alternativamente, el indol-3-piruvato puede ser utilizado como el sustrato.

El aspartato se utiliza como el donador de amino para la aminación de MP, y se utiliza una aspartato-aminotransferasa. Idealmente, una aminotransferasa que tiene una baja especificidad por la reacción de triptofano/indol-3-piruvato en comparación a la reacción de MP a monatina, se utiliza de modo que el aspartato no es utilizado para reaminar el indol-3-piruvato. La oxaloacetato-descarboxilasa (de *Pseudomonas sp.*) puede ser agregada para convertir el oxaloacetato a piruvato y dióxido de carbono. Ya que el CO_2 es volátil, éste no está disponible para la reacción con las enzimas, disminuyendo o incluso previniendo las reacciones inversas. El piruvato producido en este paso puede también ser utilizado en la reacción de condensación de aldol. Pueden ser utilizadas otras enzimas descarboxilasas, y son conocidos los homólogos que existen en *Actinobacillus actinomycetemcomitans, Aquifex aeolicus, Archaeoglobus fulgidus, Azotobacter vinelandii, Bacteroides fragilis,* varias especies de *Bordetella, Campylobacter jejuni, Chlorobium tepidum, Chloroflexus aurantiacus, Enterococcus faecalis, Fusobacterium nucleatum, Klebsiella pneumoniae, Legionella pneumophila, Magnetococcus MC-1, Mannheimia*

haemolytica, *Methylobacillus flagellatus KT*, *Pasteurella multocida* Pm70, *Petrotoga miotherma*, *Porphyromonas gingivalis*, varias especies de *Pseudomonas*, varias especies de *Pyrococcus*, *Rhodococcus*, varias especies de *Salmonella*, varias especies de *Streptococcus*, *Thermochromatium tepidum*, *Thermotoga maritima*, *Treponema pallidum* y varias especies de *Vibrio*.

Fueron realizados ensayos de triptofano-aminotransferasa con la aspartato-aminotransferasa (AspC) proveniente de *E. coli*, la tirosina-aminotransferasa (TyrB) de *E. coli*, la aminotransferasa de sustrato amplio (BSAT) de *L. major*, y las dos glutamato-oxaloacetato-aminotransferasas porcinas comercialmente disponibles como se describe en el Ejemplo 1. El oxaloacetato y el alfa-cetoglutarato fueron probados como el aceptor de amino. La proporción de la actividad utilizando monatina (Ejemplo 7) versus la actividad utilizando triptofano fue comparada, para determinar cuál enzima tuvo la mayor especificidad para la reacción de aminotransferasa de monatina. Estos resultados indicaron que la enzima con la más alta especificidad para la reacción de la monatina versus la reacción de triptofano es la glutamato-oxaloacetato-aminotransferasa porcina tipo IIa, GOAT (Sigma G7005). Esta especificidad fue independiente de qué aceptor de amino fuera utilizado. Por lo tanto, esta enzima fue utilizada en las reacciones acopladas con oxaloacetato-

descarboxilasa.

Una reacción típica que comienza a partir del indol-3-piruvato incluyó (concentraciones finales) Tris-HCl 50 mM pH 7.3, indol-3-piruvato 6 mM, piruvato de sodio 6 mM, aspartato 6 mM, PLP 0.05 mM, fosfato de potasio 3 mM, cloruro de magnesio 3 mM, 25 µg/ml de aminotransferasa, 50 µg/ml de aldolasa ProA de *C. testosteroni*, y 3 Unidades/ml de descarboxilasa (Sigma 04878). Las reacciones se dejaron proceder por 1 hora a 26°C. En algunos casos, la descarboxilasa fue omitida o el aspartato fue sustituido con alfa-cetoglutarato (como controles negativos). Las enzimas aminotransferasas descritas anteriormente fueron también probadas en lugar de GOAT para confirmar los experimentos de especificidad previos. Las muestras fueron filtradas y analizadas por LC/MS como se describe en el Ejemplo 10. Los resultados muestran que la enzima GOAT produjo la más alta cantidad de monatina por mg de proteína, con la menor cantidad de triptofano producido como un subproducto. Además, existió un beneficio doble a triple al haber agregado la enzima descarboxilasa. La enzima AspC de *E. coli* también produjo grandes cantidades de monatina en comparación a las otras aminotransferasas.

La producción de monatina fue incrementada por: 1) la adición periódica de concentraciones 2 mM de indol-piruvato,

piruvato y aspartato (cada media hora a una hora) y 2) realizando las reacciones en una ambiente anaeróbico o con amortiguadores desgasificados, 3) permitiendo que las reacciones procedieran toda la noche, y 4) utilizando descarboxilasa recién preparada que no ha sido congelada-descongelada múltiples veces. La descarboxilasa fue inhibida por concentraciones de piruvato mayores de 12 mM. A concentraciones de indol-3-piruvato mayores de 4 mM, se aceleraron las reacciones colaterales con indol-3-piruvato. La cantidad de indol-3-piruvato utilizada en la reacción pudo ser incrementada si la cantidad de aldolasa fue también incrementada. Se encontró que altos niveles de fosfato (50 mM) y aspartato (50 mM) son inhibitorios para la enzima descarboxilasa. La cantidad de enzima descarboxilasa agregada pudo ser reducida a 0.5 U/ml sin disminución en la producción de la monatina en una reacción de 1 hora. La cantidad de monatina producida se incrementó cuando la temperatura fue incrementada de 26°C a 30°C y de 30°C a 37°C; no obstante, a 37°C las reacciones colaterales del indol-3-piruvato fueron también aceleradas. Las cantidades de monatina producida se incrementaron conforme se incrementaba el pH de 7 a 7.3, y fue relativamente estable de pH 7.3 a 8.3.

Una reacción típica que comienza con el triptofano incluyó (concentraciones finales) Tris-HCl 50 mM pH 7.3, triptofano 20 mM, aspartato 6 mM, piruvato de sodio 6 mM, PLP

0.05 mM, fosfato de potasio 3 mM, cloruro de magnesio 3 mM, 25 µg/ml de aminotransferasa, 50 µg/ml de aldolasa ProA de *C. testosteroni*, 4 Unidades/ml de descarboxilasa, 5-200 mU/ml de L-aminoácido-oxidasa (sigma A-2805), 168 U/ml de catalasa (Sigma C-3515) y 0.008 mg de FAD. Las reacciones fueron llevadas a cabo por 30 minutos a 30°C. Se observó mejoramiento con la adición de descarboxilasa. La mayor cantidad de monatina fue producida cuando se utilizaron 50 mU/ml de oxidasa. Los mejoramientos fueron similares a aquellos observados cuando se utilizó la indol-3-piruvato como el sustrato. Además, la cantidad de monatina producida se incrementó cuando 1) el nivel de triptofano fue bajo (por ejemplo, por debajo de la K_m de la enzima aminotransferasa, y por lo tanto incapaz de competir con MP en el sitio activo), y 2) la proporción de oxidasa a aldolasa y aminotransferasa fue mantenida a un nivel tal que el indol-3-piruvato no pudo acumularse.

Ya sea que se inicie con el indol-3-piruvato o con el triptofano, la cantidad de monatina producida en los ensayos con tiempos de incubación de 1 a 2 horas, se incrementó cuando se utilizaron 2 a 4 veces las cantidades de todas las enzimas, mientras que se mantenía la misma proporción de enzima. Utilizando cualquier sustrato, se lograron concentraciones de aproximadamente 1 mg/ml de monatina. La

cantidad de triptofano producido si se comienza a partir de indol-piruvato fue típicamente menor de 20% de la cantidad del producto, lo cual muestra el beneficio de utilizar las reacciones acopladas. Con optimización adicional y control de las concentraciones de los intermediarios y las reacciones colaterales, se puede mejorar la productividad y el rendimiento en gran medida.

Reacciones acopladas utilizando lisina-epsilon-aminotransferasa (EC 2.6.1.36)

La lisina-epsilon-aminotransferasa (L-Lisina-6-transaminasa) se encuentra en varios organismos, incluyendo *Rhodococcus, Mycobacterium, Streptomyces, Nocardia, Flavobacterium, Candida utilis,* y *Streptomyces*. Esta es utilizada por los organismos como el primer paso en la producción de algunos antibióticos de beta-lactama (Rius y Demain, *J. Microbiol. Biotech.* 7: 95-100, 1997). Esta enzima convierte la lisina a L-2-aminoadipato-6-semialdehído (alisina), por una transaminación mediada por PLP del carbono 6 de la lisina, utilizando alfa-cetoglutarato como el aceptor del grupo amino. La alisina es inestable y espontáneamente sufre una deshidratación intramolecular para formar 6-carboxilato de 1-piperideina, una molécula cíclica. Esto inhibe efectivamente que ocurra cualquier reacción inversa. El esquema de reacción se describe en la Figura 12. Se puede utilizar también una enzima alternativa, la lisina-piruvato-

6-transaminasa (EC 2.5.1.71).

Una reacción típica contenía en 1 ml: Tris-HCl 50 mM pH 7.3, indol-3-piruvato 20 mM, PLP 0.05 mM, fosfato de potasio 6 mM pH 8, piruvato de sodio 2-50 mM, cloruro de magnesio 1.5 mM, lisina 50 mM, 100 µg de aminotransferasa (lisina-epsilon-aminotransferasa LAT-101, BioCatalytics Pasadena, California), y 200 µg de aldolasa ProA de *C. testosteroni*. La actividad de monatina producida se incrementó conforme se incrementaban las concentraciones de piruvato. La cantidad máxima utilizando estas condiciones de reacción (a 50 mM de piruvato) fue 10 veces menor que la que se observó con las reacciones acopladas utilizando oxaloacetato-descarboxilasa (aproximadamente 0.1 mg/ml).

Un pico con $[M+H]^+$=293 eluyó al tiempo esperado para la monatina, y el espectro de masa contenía varios de los mismos fragmentos observados con otros procesos enzimáticos. Un segundo pico con la proporción correcta de masa a carga (293) eluyó ligeramente más temprano que lo que se observa típicamente para la S,S-monatina producida en el Ejemplo 6 y puede indicar la presencia de otro isómero de monatina. Muy poco triptofano fue producido por esta enzima. No obstante, existe probablemente alguna actividad sobre el piruvato (produciendo alanina como un subproducto). También, se sabe que la enzima es inestable. Pueden ser realizados

mejoramientos al llevar a cabo los experimentos de evolución directa para incrementar la estabilidad, reducir la actividad con el piruvato, e incrementar la actividad con MP. Estas reacciones pueden también ser acopladas a la L-aminoácido-oxidasa/catalasa como se describió anteriormente.

Otras reacciones acopladas

Otra reacción de acoplamiento que puede mejorar el rendimiento de la monatina a partir del triptofano o el indol-piruvato, se muestra en la Figura 13. La formiato-deshidrogenasa (EC 1.2.1.2 ó 1.2.1.43) es una enzima común. Algunas formiato-deshidrogenasas requieren NADH mientras que otras pueden utilizar NADPH. La glutamato-deshidrogenasa catalizó la interconversión entre el precursor de monatina y la monatina en ejemplos previos, utilizando amortiguadores basados en amonio. La presencia del formiato de amonio y la formiato-deshidrogenasa es un sistema eficiente para la regeneración de cofactores, y la producción de dióxido de carbono es una manera eficiente para disminuir la velocidad de las reacciones inversas (Bommarius et al., *Biocatalysis* 10: 37, 1994 y Galkin et al. *Appl. Environ. Microbiol.* 63: 4651-6, 1997). Además, grandes cantidades de formiato de amonio pueden ser disueltas en el amortiguador de reacción. El rendimiento de la monatina producido por las reacciones de glutamato-deshidrogenasa (o aminaciones reductivas similares) pueden ser mejoradas por la adición de la formiato-

deshidrogenasa y el formiato de amonio.

Pueden ser utilizados otros procesos para impulsar el equilibrio hacia la producción de monatina. Por ejemplo, si el aminopropano es utilizado como el donador de aminoácido en la conversión de MP a monatina con una omega-aminoácido-aminotransferasa (EC 2.6.1.18) tal como aquellas descritas en las Patentes de los Estados Unidos Nos. 5,360,724 y 5,300,437, uno de los productos resultantes podría ser la acetona, un producto más volátil que el sustrato, aminopropano. La temperatura puede ser elevada periódicamente por periodos cortos para encender la acetona, con lo cual se alivia el equilibrio. La acetona tiene un punto de ebullición de 47°C, una temperatura que probablemente no degrada los intermediarios si se utiliza por periodos cortos de tiempo. La mayoría de las aminotransferasas que tienen actividad sobre la alfa-cetoglutarato tienen también actividad sobre el precursor de monatina. Similarmente, si se utiliza una aminotransferasa de glioxilato/ácido aromático (EC 2.6.1.60) con la glicina como el donador de amino, se produce glioxilato que es relativamente inestable y tiene un punto de ebullición altamente reducido en comparación a la glicina.

Ejemplo 14

Expresión Recombinante

Con las secuencias de ADNc y de aminoácidos de la enzima, públicamente disponibles, y las enzimas y las

secuencias descritas aquí, tales como las SEQ ID NOs: 11 y 12, así como las variantes, polimorfismos, mutantes, fragmentos y fusiones de las mismas, se hace posible la expresión y purificación de cualquier proteína, tal como una enzima, mediante técnicas estándares de laboratorio. Una persona de experiencia en la técnica comprenderá que las enzimas y fragmentos de las mismas pueden ser producidas recombinantemente en cualquier célula u organismo de interés, y purificadas antes del uso, por ejemplo antes de la producción de la SEQ ID NO: 12 y derivados de la misma.

Los métodos para producir proteínas recombinantes son bien conocidos en la técnica. Por lo tanto, el alcance de esta invención incluye la expresión recombinante de cualquier proteína o fragmento de la misma, tal como una enzima. Por ejemplo, ver la Patente de los Estados Unidos No. 5,342,764 a Jonson et al.; Patente de los Estados Unidos No. 5,846,819 a Pausch et al.; Patente de los Estados Unidos No. 5,876,969 a Fleer et al. y Sambrook et al. (*Molecular Cloning; A Laboratory Manual*, Cold Spring Harbor, New York, 1989, Capítulo 17).

En resumen, las secuencias de ADNc parciales, de longitud completa o variantes, que codifican para una proteína o péptido, pueden ser ligadas dentro de un vector de expresión, tal como un vector de expresión bacteriano o eucariótico. Las proteínas y/o los péptidos pueden ser

producidos mediante la colocación de un promotor corriente arriba (5') de la secuencia de ADNc. Los ejemplos de promotores incluyen, pero no están limitados a *lac*, *trp*, *tac*, *trc*, operador mayor y regiones promotoras del fago lambda, la región de control de la proteína de recubrimiento fd, los promotores temprano y tardío del SV40, los promotores derivados del polioma, adenovirus, retrovirus, baculovirus y virus de simio, el promotor para la 3-fosfoglicerato-cinasa, los promotores de la fosfatasa ácida de levadura, el promotor de los factores de acoplamiento alfa de levadura y combinaciones de los mismos.

Los vectores adecuados para la producción de proteínas nativas intactas incluyen pKC30 (Shimatake y Rosenberg, 1981, *Nature* 292: 128), pKK177-3 (Amann y Brosius, 1985, *Gene* 40: 183) y pET-3 (Studier y Moffatt, 1986, *J. Mol. Biol.* 189: 113). Una secuencia de ADN puede ser transferida a otros vehículos de clonación, tales como otros plásmidos, bacteriófagos, cósmidos, virus animales y cromosomas artificiales de levadura (YACs) (Burke et al., 1987, *Science* 236: 806-12). Estos vectores pueden ser introducidos dentro de una variedad de hospederos incluyendo células somáticas, y organismos simples o complejos, tales como bacterias, hongos (Timberlake y Marshall, 1989, *Science* 244: 1313-7), invertebrados, plantas (Gasser y Fraley, 1989, *Science* 244: 1293), y mamíferos (Pursel et al., 1989, *Science* 244: 1281-

278

8), que se hacen transgénicas por la introducción del ADNc heterólogo.

Para la expresión en células de mamífero, una secuencia de ADNc puede ser ligada a los promotores heterólogos, tales como el virus de simio SV40, el promotor en el vector pSV2 (Mulligan y Berg, 1981, *Proc. Natl. Acad. Sci. EUA* 78: 2972-6), e introducidos dentro de las células, tales como las células COS-1 de mono (Gluzman, 1981, *Cell* 23: 175-82), para lograr la expresión transitoria o a largo plazo. La integración estable de la construcción del gen quimérico puede ser mantenida en células de mamífero mediante selección bioquímica, tal como neomicina (Southern y Berg, 1982, *J. Mol. Appl. Genet.* 1: 327-41) y ácido micofenólico (Mulligan y Berg, 1981, *Proc. Natl. Acad. Sci. EUA* 78: 2072-6).

La transferencia del ADN dentro de células eucarióticas, tales como células humanas o de otro mamífero, es una técnica convencional. Los vectores son introducidos dentro de las células recipientes como ADN puro (transfección) por ejemplo, mediante precipitación con fosfato de calcio (Graham y Vander Eb, 1973, *Virology* 52: 466), fosfato de estroncio (Brash et al., 1987, *Mol. Cell Biol.* 7: 2013), electroporación (Neumann et al., 1982, *EMBO J.* 1: 841), lipofección (Felgner et al., 1987, *Proc. Natl. Acad. Sci. EUA* 84: 7413), DEAE-dextrano (McCuthan et al.,

1968, *J. Natl. Cancer Inst.* 41: 351), microinyección (Mueller et al., 1978, *Cell* 15: 579), fusión de protoplastos (Schafner, 1980, *Proc. Natl. Acad. Sci. EUA* 77: 2163-7), o pistolas de pellas (Klein et al., 1987, *Nature* 327: 70). Alternativamente, el ADNc puede ser introducido mediante infección con vectores virales, por ejemplo retrovirus (Bernstein et al., 1985, *Gen. Engrg.* 7: 235) tales como adenovirus (Ahmad et al., 1986, *J. Virol.* 57: 267) o Herpes (Spaete et al., 1982, *Cell* 30: 295).

En vista de las muchas posibles modalidades a las cuales pueden ser aplicados los principios de la presente descripción, se debe reconocer que las modalidades ilustradas son únicamente ejemplos particulares de la descripción, y no deben ser tomadas como una limitación sobre el alcance de la descripción. Más bien, el alcance de la descripción está en concordancia con las siguientes reivindicaciones. Por lo tanto, se reclama como invención todo aquello que entre dentro del alcance y espíritu de estas reivindicaciones.

280

REIVINDICACIONES

1. Un método para producir monatina, caracterizado porque comprende la conversión de glucosa a monatina.

2. Un método para producir monatina, que comprende la conversión de triptofano y/o ácido indol-3-láctico a monatina, caracterizado el método porque comprende:

poner en contacto el triptofano o el ácido indol-3-láctico con un primer polipéptido, en donde el primer polipéptido convierte el triptofano o el ácido indol-3-láctico a indol-3-piruvato;

poner en contacto el indol-3-piruvato con un segundo polipéptido y una fuente de 3 carbonos, en donde el segundo polipéptido convierte el indol-3-piruvato y la fuente de 3 átomos de carbono a ácido 2-hidroxi-2-(indol-3-ilmetil)-4-cetoglutárico; y

poner en contacto el ácido 2-hidroxi-2-(indol-3-ilmetil)-4-cetoglutárico con un tercer polipéptido, en donde el tercer polipéptido convierte el ácido 2-hidroxi-2-(indol-3-ilmetil)-4-cetoglutárico a monatina.

3. El método de conformidad con la reivindicación 2, caracterizado porque el primer polipéptido se selecciona del grupo que consiste de triptofano-aminotransferasa (EC 2.6.1.27), triptofano-deshidrogenasa (EC 1.4.1.19), tirosina(aromático)-aminotransferasa (EC 2.6.1.5),

triptofano-fenilpiruvato-transaminasa (EC 2.6.1.28), aminotransferasa de sustratos múltiples (EC 2.6.1.-), aspartato-aminotransferasa (EC 2.6.1.1), triptofano-oxidasa, L-aminoácido-oxidasa (EC 1.4.3.2), D-aminoácido-deshidrogenasa (EC 1.4.99.1), D-aminoácido-oxidasa (EC 1.4.3.3), D-triptofano-aminotransferasa, D-alanina-aminotransferasa (EC 2.6.1.21), glutamato-deshidrogenasa (EC 1.4.1.2-4), fenilalanina-deshidrogenasa (EC 1.4.1.20), y una combinación de los mismos, y en donde el primer polipéptido convierte el triptofano a indol-3-piruvato.

4. El método de conformidad con la reivindicación 2, caracterizado porque el primer polipéptido se selecciona del grupo que consiste de indol-lactato-deshidrogenasa (EC 1.1.1.1110), R-4-hidroxifenil-lactato-deshidrogenasa (EC 1.1.1.222), 3-(4)-hidroxifenilpiruvato-reductasa (EC 1.1.1.237), lactato-deshidrogenasa (EC 1.1.1.27, 1.1.1.28, 1.1.2.3), (3-imidazol-5-il)lactato-deshidrogenasa (EC 1.1.1.111), lactato oxidasa (EC 1.1.3-), o una combinación de los mismos, y en donde el primer polipéptido convierte el ácido indol-3-láctico a indol-3-piruvato.

5. El método de conformidad con la reivindicación 2 ó 3, caracterizado porque el primero y/o tercer polipéptido comprende:

una secuencia de aminoácidos mostrada en la SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 32, Genbank Acceso No. NO_388848.1,

282

Genbank Acceso No. ZP00005082.1, o Genbank Acceso No. AAC74014.1;

una secuencia de aminoácidos que comprende al menos 90% de identidad secuencial con la secuencia mostrada en la SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 32, Genbank Acceso No. NP_388848.1, Genbank Acceso No. ZP00005082.1 o Genbank Acceso No. AAC74014.1; o

las secuencias de aminoácidos que difieren de las SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 32, Genbank Acceso No. NP_388848.1, Genbank Acceso No. ZP00005982.1, o Genbank Acceso No. AAC74014.1 por menos de 50 sustituciones conservadoras de aminoácidos, en donde la secuencia de aminoácidos convierte el triptofano al indol-3-piruvato.

6. El método de conformidad con la reivindicación 2, caracterizado porque el segundo polipéptido es una enzima EC 4.1.2.- o EC 4.1.3.-.

7. El método de conformidad con la reivindicación 6, caracterizado porque la enzima EC 4.1.3.- es la 4-hidroxi-2-oxoglutarato-glioxilato-liasa (EC 4.1.3.16), la 4-hidroxi-4-metil-2-oxoglutarato-piruvato-liasa (EC 4.1.3.17) o una combinación de las mismas.

8. El método de conformidad con la reivindicación 2, caracterizado porque el tercer polipéptido se selecciona del grupo que consiste de tirosina(aromático)aminotransferasa (EC 2.6.1.5), triptofano-deshidrogenasa (EC 1.4.1.19),

triptofano-fenilpiruvato-transaminasa (EC 2.6.1.28), triptofano-aminotransferasa (EC 2.6.1.27), aspartato-aminotransferasa (2.6.1.1), D-alanina-aminotransferasa (EC 2.6.1.21), D-triptofano-aminotransferasa, glutamato-deshidrogenasa (EC 1.4.1.2-4), fenilalanina-deshidrogenasa (EC 1.4.1.20), aminotransferasa de sustratos múltiples (EC 2.6.1.-), D-aminoácido-deshidrogenasa (EC 1.4.99.1), y una combinación de las mismas.

9. El método de conformidad con la reivindicación 2, caracterizado porque el primer polipéptido comprende una aminoácido-oxidasa (EC 1.4.3.-) y una catalasa (EC 1.11.1.6), y en donde el primer polipéptido convierte el triptofano a indol-3-piruvato.

10. El método de conformidad con la reivindicación 9, caracterizado porque la aminoácido-oxidasa comprende una triptofano-oxidasa.

11. El método de conformidad con la reivindicación 2, caracterizado porque la fuente de 3 átomos de carbono se selecciona del grupo que consiste de piruvato, fosfoenolpiruvato, alanina, serina, cisteina o una combinación de los mismos.

12. El método de conformidad con la reivindicación 11, caracterizado porque el piruvato es producido por la alanina y un polipéptido capaz de realizar la transaminación de la alanina; la serina y un polipéptido capaz de realizar

la beta-eliminación de la serina; la cisteína y un polipéptido capaz de la beta-eliminación de la cisteina; aspartato y un polipéptido capaz de realizar reacciones de beta-liasa con un aminoácido dicarboxílico; lactato y lactato-oxidasa (EC 1.1.3.-) o lactato-deshidrogenasa (EC 1.1.1.27, 1.1.1.28, 1.1.2.3); o una combinación de las mismas.

13. El método de conformidad con la reivindicación 2, caracterizado porque el método comprende además la reducción de una cantidad de peróxido de hidrógeno producido cuando el triptofano es convertido a indol-3-piruvato, al poner en contacto el peróxido de hidrógeno con una catalasa.

14. El método de conformidad con la reivindicación 2, caracterizado porque el segundo polipéptido comprende una secuencia de aminoácidos seleccionada del grupo que consiste de:

SEQ ID NO: 66, Genbank Acceso No. CAC46344; Genbank Acceso No. CAB14127.1, Genbank Acceso No. AAC74920.1 o Genbank Acceso No. CAC47463.1;

una secuencia de aminoácidos que comprende al menos 90% de identidad secuencial a la SEQ ID NO: 66, Genbank Acceso No. CAC46344, Genbank Acceso No. CAB14127.1, Genbank Acceso No. AAC74920.1, o Genbank Acceso No. CAC47463.1; o

las secuencias de aminoácidos que difieren de la SEQ ID NO: 66, Genbank Acceso No. CAC46344, Genbank Acceso No.

CAB14127.1, Genbank Acceso No. AAC74920.1, o Genbank Acceso No. CAC47463.1 por menos de 50 sustituciones conservadoras de aminoácidos, en donde la secuencia de aminoácidos convierte el indol-3-piruvato y el piruvato a MP.

15. El método de conformidad con la reivindicación 8, caracterizado porque el tercer polipéptido comprende una aspartato-aminotransferasa, en donde el ácido 2-hidroxi-2-(indol-3-ilmetil)-4-cetoglutárico es puesto en contacto con el aspartato para producir oxaloacetato y monatina.

16. El método de conformidad con la reivindicación 15, caracterizado porque comprende además poner en contacto el oxaloacetato con una descarboxilasa.

17. El método de conformidad con la reivindicación 2, caracterizado porque el tercer polipéptido comprende una lisina-epsilon-aminotransferasa, y en donde además el ácido 2-hidroxi-2-(indol-3-ilmetil)-4-cetoglutárico es puesto en contacto con la lisina.

18. El método de conformidad con la reivindicación 2, caracterizado porque el tercer polipéptido comprende una enzima que tiene actividad de aminación reductora y un polipéptido que es capaz de reciclar NAD(P)H.

19. El método de conformidad con la reivindicación 18, caracterizado porque la enzima que tiene actividad de aminación reductora se selecciona del grupo que consiste de glutamato-deshidrogenasa, fenilalanina-deshidrogenasa,

triptofano-deshidrogenasa, o una combinación de las mismas.

20. El método de conformidad con la reivindicación 18, caracterizado porque el polipéptido capaz de reciclar el NAD(P)H produce además un producto volátil.

21. El método de conformidad con la reivindicación 20, caracterizado porque el polipéptido que es capaz de reciclar NAD(P)H y producir un producto volátil comprende formiato-deshidrogenasa, y en donde el producto volátil producido es dióxido de carbono.

22. El método de conformidad con la reivindicación 2, caracterizado porque el método comprende la conversión del ácido indol-3-láctico a monatina.

23. El método de conformidad con la reivindicación 2, caracterizado porque se producen al menos 10 g/litro de monatina.

24. Un método para producir la monatina, caracterizado porque comprende la conversión del indol-3-piruvato a la monatina.

25. Un método para producir la monatina, caracterizado porque comprende la conversión del triptofano a monatina.

26. Una célula capaz de producir monatina, caracterizada porque la célula comprende al menos una secuencia exógena de ácido nucleico que codifica para al menos un polipéptido, en donde al menos un polipéptido

convierte un primer intermediario en una vía para elaborar monatina en un segundo intermediario o en monatina.

27. La célula de conformidad con la reivindicación 26, caracterizada porque al menos un polipéptido se selecciona del grupo que consiste de triptofano-aminotransferasa (EC 2.6.1.27), triptofano-deshidrogenasa (EC 1.4.1.19), tirosina(aromático)-aminotransferasa (EC 2.6.1.5), triptofano-fenilpiruvato-transaminasa (EC 2.6.1.28), aminotransferasa de sustratos múltiples (EC 2.6.1.-), aspartato-aminotransferasa (EC 2.6.1.1), triptofano-oxidasa, L-aminoácido-oxidasa (EC 1.4.3.2), D-aminoácido-deshidrogenasa (EC 1.4.99.1), D-aminoácido-oxidasa (EC 1.4.3.3), D-alanina-aminotransferasa (EC 2.5.1.21), D-triptofano-aminotransferasa, indol-lactato-deshidrogenasa (EC 1.1.1.110), R-4-hidroxifenil-lactato-deshidrogenasa (EC 1.1.1.222), 3-(4)-hidroxifenilpiruvato-reductasa (EC 1.1.1.237), lactato-deshidrogenasa (EC 1.1.1.27, 1.1.1.28, 1.1.2.3), (3-imidazol-5-il)-lactato-deshidrogenasa (EC 1.1.1.111), lactato-oxidasa (EC 1.1.3.-), sintasa/liasa (4.1.2.-), sintasa/liasa (4.1.3.-), fenilalanina-deshidrogenasa (EC 1.4.1.20), glutamato-deshidrogenasa (EC 1.4.1.2, 1.4.1.3, 1.4.1.4), y una combinación de las mismas.

28. La célula de conformidad con la reivindicación 27, caracterizada porque al menos un polipéptido se selecciona del grupo que consiste de triptofano-

aminotransferasa (EC 2.6.1.27), triptofano-deshidrogenasa (EC 1.4.1.19), tirosina(aromático)-aminotransferasa (EC 2.6.1.5), triptofano-fenilpiruvato-transaminasa (EC 2.6.1.28), aminotransferasa de sustratos múltiples (EC 2.6.1.-), aspartato-aminotransferasa (EC 2.6.1.1), triptofano-oxidasa, L-aminoácido-oxidasa (EC 1.4.3.2), D-amincácido-deshidrogenasa (EC 1.4.99.1), D-aminoácido-oxidasa (EC 1.4.3.3), D-alanina-aminotransferasa (EC 2.6.1.21), D-triptofano-aminotransferasa, sintasa/liasa (EC 4.1.3.-), sintasa/liasa (4.1.2.-), fenilalanina-deshidrogenasa (EC 1.4.1.20), glutamato-deshidrogenasa (EC 1.4.1.2, 1.4.1.3, 1.4.1.4) y una combinación de las mismas.

29. La célula de conformidad con la reivindicación 27, caracterizada porque al menos un polipéptido comprende una actividad seleccionada del grupo que consiste de indol-lactato-deshidrogenasa (EC 1.1.1.110), R-4-hidroxifenil-lactato-deshidrogenasa (EC 1.1.1.222), 3-(4)-hidroxifenilpiruvato-reductasa (EC 1.1.1.237), lactato-deshidrogenasa (EC 1.1.1.27, 1.1.1.28, 1.1.2.3), (3-imidazol-5-il)-lactato-deshidrogenasa (EC 1.1.1.111), lactato-oxidasa (EC 1.1.3.-), sintasa/liasa (4.1.2.-), sintasa/liasa (4.1.3.-), triptofano-deshidrogenasa (EC 1.4.1.19), triptofano-fenilpiruvato-transaminasa (EC 2.6.1.28), triptofano-aminotransferasa (EC 2.6.1.27), tirosina(aromático)-aminotransferasa (EC 2.6.1.5), aminotransferasa de sustrato

múltiples (EC 2.6.1.-), aspartato-aminotransferasa (EC 2.6.1.1), glutamato-deshidrogenasa (EC 1.4.1.2, 1.4.1.3, 1.4.1.4), fenilalanina-deshidrogenasa (EC 1.4.1.20), D-triptofano-aminotransferasa, D-aminoácido-deshidrogenasa (EC 1.4.99.1), D-alanina-aminotransferasa (EC 2.6.1.21) y una combinación de las mismas.

30. La célula de conformidad con la reivindicación 27, caracterizada porque la sintasa/liasa (EC 4.1.3.-) comprende 4-hidroxi-2-oxoglutarato-aldolasa (EC 4.1.3.16) o 4-hidroxi-4-metil-2-oxoglutarato-aldolasa (EC 4.1.3.17).

31. La célula recombinante de conformidad con la reivindicación 26, caracterizada porque la célula es una célula bacteriana.

32. La célula recombinante de conformidad con la reivindicación 26, caracterizada porque la célula es una célula de levadura.

33. La célula de conformidad con la reivindicación 27, caracterizada porque la célula comprende además uno o más polipéptidos que son codificados sobre al menos una secuencia exógena de ácido nucleico, en donde los polipéptidos comprenden actividades seleccionadas del grupo que consiste de: catalasa, oxaloacetato-descarboxilasa, lisina-epsilon-aminotransferasa, y formiato-deshidrogenasa.

34. Un polipéptido, caracterizado porque comprende una secuencia de aminoácidos seleccionada del grupo que

290

consiste de: Genbank Acceso No. P00913), Genbank Acceso No. NP_290344, Genbank Acceso No.CAA34096, Genbank Acceso No. NP_246359, Genbank Acceso No. AAB96579, Genbank Acceso No. NP_232561, Genbank Acceso No. Q59342, Genbank Acceso No. P28796, Genbank Acceso No. 1AX4_A, Genbank Acceso No. BAA34638, Genbank Acceso No. P31014, Genbank Acceso No. P31015, Genbank Acceso No. NP_280989, Genbank Acceso No. AAC33284, Genbank Acceso No. NP_147838, Genbank Acceso No. AAF63441, o Genbank Acceso No. AAA24679, donde los residuos de aminoácidos en las posiciones 103 y 299 con respecto a Genbank Acceso No. NP_418164 han sido cambiados para ampliar una especificidad del polipéptido para aceptar sustratos de aminoácido dicarboxílico.

35. El polipéptido de conformidad con la reivindicación 34, caracterizado porque el cambio en la secuencia de aminoácidos se selecciona del grupo que consiste de: arginina 103 cambiada a treonina, arginina 103 cambiada a treonina y valina 299 a arginina; y valina 299 a arginina.

36. Un polipéptido caracterizado porque comprende una secuencia de aminoácidos seleccionada del grupo que consiste de Genbank Acceso No. 2TPL_A, Genbank Acceso No. P31012, Genbank Acceso No. P31013, A49493, Genbank Acceso No. P31011, Genbank Acceso No. AAB24234, Genbank Acceso No. 1TPL_A, Genbank Acceso No. 1TPL_B, Genbank Acceso No. NP_245748, Genbank Acceso No. AAB41499, Genbank Acceso No.

291

Q08897, Genbank Acceso No. T45297, y Genbank Acceso No. AAA71928, donde los residuos de aminoácidos en las posiciones 100 y 283 con respecto a Genbank Acceso No. X66978, han sido cambiados para ampliar una especificidad del polipéptido para aceptar sustratos de aminoácido de ácido dicarboxílico.

37. El polipéptido de conformidad con la reivindicación 36, caracterizado porque el cambio en la secuencia de aminoácidos se selecciona del grupo que consiste de: arginina 100 cambiada a treonina, arginina 100 cambiada a treonina y valina 283 a arginina; y valina 283 a arginina.

38. El polipéptido de conformidad con la reivindicación 34 ó 36, caracterizado porque el sustrato es aspartato.

39. Un método para elaborar la monatina, caracterizado porque comprende la reacción de al menos un derivado de alanina bloqueado en los grupos carboxilo y amino, y al menos un derivado de indol-3-piruvato bloqueado en un grupo carboxilo.

40. El método para elaborar la monatina de conformidad con la reivindicación 39, caracterizado porque el derivado de alanina es 3-cloro-alanina o 3-bromo-alanina, y en donde el indol-3-piruvato se hace reaccionar con el derivado de alanina en una reacción de Grignard o de organolitio.

41. Un método para elaborar la monatina,

caracterizado porque comprende la conversión de un sustrato seleccionado del grupo que consiste de triptofano, ácido indol-3-láctico, indol-3-piruvato, y ácido 2-hidroxi-2-(indol-3-ilmetil)-4-cetoglutárico, a uno o más intermediarios, en donde es elaborada la monatina.

42. El método de conformidad con la reivindicación 41, caracterizado porque al menos uno de los productos intermediarios es convertido vía una reacción química.

43. El método de conformidad con la reivindicación 41, caracterizado porque al menos uno de los productos intermediarios es convertido vía un polipéptido.

44. Un ácido nucleico aislado, caracterizado porque comprende al menos 90% de identidad secuencial a la SEQ ID NO: 11.

45. El ácido nucleico aislado de conformidad con la reivindicación 44, caracterizado porque el ácido nucleico comprende la SEQ ID NO: 11.

46. El ácido nucleico aislado de conformidad con la reivindicación 44, caracterizado porque el ácido nucleico comprende al menos 20 nucleótidos consecutivos de la SEQ ID NO: 11.

47. Un vector, caracterizado porque comprende el ácido nucleico aislado de conformidad con la reivindicación 44.

48. Una célula recombinante, caracterizada porque

293

comprende el vector de conformidad con la reivindicación 47.

49. Una proteína aislada, caracterizada porque es codificada por el ácido nucleico aislado de conformidad con la reivindicación 44.

50. La proteína aislada de conformidad con la reivindicación 49, caracterizada porque la proteína comprende una secuencia que tiene al menos 90% de identidad secuencial a la SEQ ID NO: 12.

51. La proteína aislada de conformidad con la reivindicación 50, caracterizada porque la proteína comprende la SEQ ID NO: 12.

52. Un método para producir la monatina, caracterizado porque comprende la conversión del ácido 2-hidroxi-2-(indol-3-ilmetil)-4-cetoglutárico a monatina.

53. El método de conformidad con la reivindicación 52, caracterizado porque el método comprende además la detección del ácido 2-hidroxi-2-(indol-3-ilmetil)-4-cetoglutárico por espectroscopia de masa.

54. Una composición, caracterizada porque comprende el ácido 2-hidroxi-2-(indol-3-ilmetil)-4-cetoglutárico.

55. El método de conformidad con la reivindicación 2, caracterizado porque el triptofano es producido a partir de indol utilizando un polipéptido de EC 4.1.99.1.

56. Un método para producir monatina, caracterizado porque comprende la reacción del triptofano con uno o más

294

polipéptidos.

57. El método de conformidad con la reivindicación 56, caracterizado porque uno o más polipéptidos se seleccionan del grupo que consiste de: triptofano-aminotransferasa (EC 2.6.1.27), triptofano-deshidrogenasa (EC 1.4.1.19), tirosina(aromático)-aminotransferasa (EC 2.6.1.5), triptofano-fenilpiruvato-transaminasa (EC 2.6.1.28), aminotransferasa de sustratos múltiples (EC 2.6.1.-), aspartato-aminotransferasa (EC 2.6.1.1), triptofano-oxidasa, L-aminoácido-oxidasa (EC 1.4.3.2), D-aminoácido-deshidrogenasa (EC 1.4.99.1), D-aminoácido-oxidasa (EC 1.4.3.3), D-alanina-aminotransferasa (EC 2.6.1.21), D-triptofano-aminotransferasa, sintasa/liasa (EC 4.1.3.-), sintasa/liasa (4.1.2.-), fenilalanina-deshidrogenasa (EC 1.4.1.20), glutamato-deshidrogenasa (EC 1.4.1.2, 1.4.1.3, 1.4.1.4) y una combinación de las mismas.

295

RESUMEN DE LA INVENCION

Se describen los métodos y las composiciones que pueden ser utilizados para elaborar la monatina a partir de glucosa, triptofano, ácido indol-3-láctico, indol-3-piruvato y ácido 2-hidroxi-2-(indol-3-ilmetil)-4-cetoglutárico. Se describen también los métodos para producir los intermediarios del indol-3-piruvato y el ácido 2-hidroxi-2-(indol-3-ilmetil)-4-cetoglutárico. Las composiciones proporcionadas incluyen moléculas de ácido nucleico, polipéptidos, estructuras químicas y células. Los métodos incluyen los procesos *in vitro* e *in vivo*, y los métodos *in vitro* incluyen las reacciones químicas.

296
247

Sheet No. 1

PCT

REQUEST

The undersigned requests that the present International application be processed according to the Patent Cooperation Treaty.

For receiving Office use only

International Application No.

International Filing Date

Name of receiving Office and "PCT International Application"

Applicant's or agent's file reference
(If desired) (12 characters maximum): 6682-64349

Box No. I TITLE OF INVENTION
POLYPEPTIDES AND BIOSYNTHETIC PATHWAYS

Box No. II APPLICANT ☐ This person is also inventor.

Name and address: *(Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country. The country of the address indicated in this Box is the applicant's State (that is, country) of residence if no State of residence is indicated below)*

CARGILL, INCORPORATED
Law Department
P.O. Box 5624
Minneapolis, Minnesota 55440-5624
United States of America

Telephone No.
(952) 742-5402
Facsimile No.
(952) 742-6349
Teleprinter No.
Applicant's registration No. with the Office

State *(that is, country)* of nationality: US

State *(that is, country)* of residence: US

This person is applicant for the purposes of: ☒ all designated States ☐ all designated States except the United States of America ☐ the United States of America only ☐ the States indicated in the Supplemental Box

BOX No. III FURTHER APPLICANT(S) AND/OR (FURTHER) INVENTOR(S)

Name and address: *(Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country. The country of the address indicated in this Box is the applicant's State (that is, country) of residence if no State of residence is indicated below.)*

ABRAHAM, Timothy W.
18507 Apple Tree Court
Minnetonka, Minnesota 55345
United States of America

This person is:
☐ applicant only
☐ applicant and inventor
☒ inventor only *(If this check-box is marked, do not fill in below.)*

Applicant's registration No. with the Office

State *(that is, country)* of nationality:

State *(that is, country)* of residence:

This person is applicant for the purposes of: ☐ all designated States ☐ all designated States except the United States of America ☐ the United States of America only ☐ the States indicated in the supplemental box

☒ Further applicants and/or (further) inventors are indicated on a continuation sheet.

BOX No. IV AGENT OR COMMON REPRESENTATIVE; OR ADDRESS FOR CORRESPONDENCE

The person identified below is hereby/has been appointed to act on behalf of the applicant(s) before the competent International Authorities as: ☒ agent ☐ common representative

Name and address: *(Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country)*

SKELTON, Jeffrey J.
Cargill, Incorporated
Law Department
P.O. Box 5624
Minneapolis, Minnesota 55440-5624
United States of America

Telephone No.
(952) 742-5402
Facsimile No.
(952) 742-6349
Teleprinter No.
Agent's registration No. with the office
42,152

☐ *Address for correspondence:* Mark this check-box where no agent or common representative is/has been appointed and the space above is used instead to indicate a special address to which correspondence should be sent.

Sheet No. 2

Continuation of Box No. III FURTHER APPLICANT(S) AND/OR (FURTHER) INVENTOR(S)

If none of the following sub-boxes is used, this sheet should not be included in the request.

Name and address: *(Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country. The country of the address indicated in this Box is the applicant's State (that is, country) of residence if no State of residence is indicated below.)*

CAMERON, Douglas C.
3590 Rustler Lane, N.
Plymouth, Minnesota 55447
United States of America

This person is:
☐ applicant only
☐ applicant and inventor
☒ inventor only *(If this check-box is marked, do not fill in below.)*

Applicant's registration No. with the Office

State *(that is, country)* of nationality:

State *(that is, country)* of residence:

This person is applicant for the purposes of: ☐ all designated States ☐ all designated States except the United States of America ☐ the United States of America only ☐ the States indicated in the Supplemental Box

Name and address: *(Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country. The country of the address indicated in this Box is the applicant's State (that is, country) of residence if no State of residence is indicated below.)*

DALLUGE, Joseph
3516 County Road 30 S.W.
Waverly, Minnesota 55390
United States of America

This person is:
☐ applicant only
☐ applicant and inventor
☒ inventor only *(If this check-box is marked, do not fill in below.)*

Applicant's registration No. with the Office

State *(that is, country)* of nationality:

State *(that is, country)* of residence:

This person is applicant for the purposes of: ☐ all designated States ☐ all designated States except the United States of America ☐ the United States of America only ☐ the States indicated in the Supplemental Box

Name and address: *(Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country. The country of the address indicated in this Box is the applicant's State (that is, country) of residence if no State of residence is indicated below.)*

HICKS, Paula M.
11601 Tanglewood Drive
Eden Prairie, Minnesota 55347
United States of America

This person is:
☐ applicant only
☐ applicant and inventor
☒ inventor only *(If this check-box is marked, do not fill in below.)*

Applicant's registration No. with the Office

State *(that is, country)* of nationality:

State *(that is, country)* of residence:

This person is applicant for the purposes of: ☐ all designated States ☐ all designated States except the United States of America ☐ the United States of America only ☐ the States indicated in the Supplemental Box

Name and address: *(Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country. The country of the address indicated in this Box is the applicant's State (that is, country) of residence if no State of residence is indicated below.)*

HOBSON, Russell J.
316 Vinewood Lane North
Plymouth, Minnesota 55441
United States of America

This person is:
☐ applicant only
☐ applicant and inventor
☒ inventor only *(If this check-box is marked, do not fill in below.)*

Applicant's registration No. with the Office

State *(that is, country)* of nationality:

State *(that is, country)* of residence:

This person is applicant for the purposes of: ☐ all designated States ☐ all designated States except the United States of America ☐ the United States of America only ☐ the States indicated in the Supplemental Box

☒ Further applicants and/or (further) inventors are indicated on another continuation sheet.

298

Sheet No. 3

Continuation of Box No. III FURTHER APPLICANT(S) AND/OR (FURTHER) INVENTOR(S)

If none of the following sub-boxes is used, this sheet should not be included in the request.

Name and address: *(Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country. The country of the address indicated in this Box is the applicant's State (that is, country) of residence if no State of residence is indicated below.)*

MCFARLAN, Sara C.
2151 Bayard Avenue
St. Paul, Minnesota 55116
United States of America

This person is:
- ☐ applicant only
- ☐ applicant and inventor
- ☒ inventor only *(If this check-box is marked, do not fill in below)*

Applicant's registration No. with the Office

State *(that is, country)* of nationality:

State *(that is, country)* of residence:

This person is applicant for the purposes of: ☐ all designated States ☐ all designated States except the United States of America ☐ the United States of America only ☐ the States indicated in the Supplemental Box

Name and address: *(Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country. The country of the address indicated in this Box is the applicant's State (that is, country) of residence if no State of residence is indicated below.)*

MILLIS, Jim
2360 Yuma Lane North
Plymouth, Minnesota 55447
United States of America

This person is:
- ☐ applicant only
- ☐ applicant and inventor
- ☒ inventor only *(If this check-box is marked, do not fill in below)*

Applicant's registration No. with the Office

State *(that is, country)* of nationality:

State *(that is, country)* of residence:

This person is applicant for the purposes of: ☐ all designated States ☐ all designated States except the United States of America ☐ the United States of America only ☐ the States indicated in the Supplemental Box

Name and address: *(Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country. The country of the address indicated in this Box is the applicant's State (that is, country) of residence if no State of residence is indicated below.)*

ROSAZZA, John
453 Linder Road, N.E.
Iowa City, Iowa 52240
United States of America

This person is:
- ☐ applicant only
- ☐ applicant and inventor
- ☒ inventor only *(If this check-box is marked, do not fill in below)*

Applicant's registration No. with the Office

State *(that is, country)* of nationality:

State *(that is, country)* of residence:

This person is applicant for the purposes of: ☐ all designated States ☐ all designated States except the United States of America ☐ the United States of America only ☐ the States indicated in the Supplemental Box

Name and address: *(Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country. The country of the address indicated in this Box is the applicant's State (that is, country) of residence if no State of residence is indicated below.)*

This person is:
- ☐ applicant only
- ☐ applicant and inventor
- ☐ inventor only *(If this check-box is marked, do not fill in below)*

Applicant's registration No. with the Office

State *(that is, country)* of nationality:

State *(that is, country)* of residence:

This person is applicant for the purposes of: ☐ all designated States ☐ all designated States except the United States of America ☐ the United States of America only ☐ the States indicated in the Supplemental Box

☐ Further applicants and/or (further) inventors are indicated on another continuation sheet.

2997

Sheet No. 4

Box No. V DESIGNATION OF STATES *Mark the applicable check-boxes below; at least one must be marked*

The following designations are hereby made under Rule 4.9(a):

Regional Patent

☒ **AP** ARIPO Patent: GH Ghana, GM Gambia, KE Kenya, LS Lesotho, MW Malawi, MZ Mozambique, SD Sudan, SL Sierra Leone, SZ Swaziland, TZ United Republic of Tanzania, UG Uganda, ZM Zambia, ZW Zimbabwe, and any other State which is a Contracting State of the Harare Protocol and of the PCT *(if other kind of protection or treatment desired, specify on dotted line)*: .....

.....

☒ **EA** Eurasian Patent: AM Armenia, AZ Azerbaijan, BY Belarus, KG Kyrgyzstan, KZ Kazakstan, MD Republic of Moldova, RU Russian Federation, TJ Tajikistan, TM Turkmenistan, and any other State which is a Contracting State of the Eurasian Patent Convention and of the PCT

☒ **EP** European Patent: AT Austria, BE Belgium, BG Bulgaria, CH & LI Switzerland and Liechtenstein, CY Cyprus, CZ Czech Republic, DE Germany, DK Denmark, EE Estonia, ES Spain, FI Finland, FR France, GB United Kingdom, GR Greece, HU Hungary, IE Ireland, IT Italy, LU Luxembourg, MC Monaco, NL Netherlands, PT Portugal, RO Romania, SE Sweden, SI Slovenia, SK Slovakia, TR Turkey, and any other State which is a Contracting State of the European Patent Convention and of the PCT

☒ **OA** OAPI Patent: BF Burkina Faso, BJ Benin, CF Central African Republic, CG Congo, CI Côte d'Ivoire, CM Cameroon, GA Gabon, GN Guinea, GQ Equatorial Guinea, GW Guinea-Bissau, ML Mali, MR Mauritania, NE Niger, SN Senegal, TD Chad, TG Togo, and any other State which is a member State of OAPI and a Contracting State of the PCT *(if other kind of protection or treatment desired, specify on dotted line)* .....

National Patent *(if other kind of protection or treatment desired, specify on dotted line)*:

☒	AE United Arab Emirates	☒	GM Gambia	☒	NZ New Zealand
☒	AG Antigua and Barbuda	☒	HR Croatia	☒	OM Oman
☒	AL Albania	☒	HU Hungary	☒	PH Philippines
☒	AM Armenia	☒	ID Indonesia	☒	PL Poland
☒	AT Austria	☒	IL Israel	☒	PT Portugal
☒	AU Australia	☒	IN India	☒	RO Romania
☒	AZ Azerbaijan	☒	IS Iceland	☒	RU Russian Federation
☒	BA Bosnia and Herzegovina	☒	JP Japan	☒	SC Seychelles
☒	BB Barbados	☒	KE Kenya	☒	SD Sudan
☒	BG Bulgaria	☒	KG Kyrgyzstan	☒	SE Sweden
☒	BR Brazil	☒	KP Democratic People's Republic of Korea	☒	SG Singapore
☒	BY Belarus	☒	KR Republic of Korea	☒	SK Slovakia
☒	BZ Belize	☒	KZ Kazakhstan	☒	SL Sierra Leone
☒	CA Canada	☒	LC Saint Lucia	☒	TJ Tajikistan
☒	CH and LI Switzerland and Liechtenstein	☒	LK Sri Lanka	☒	TN Tunisia
☒	CN China	☒	LR Liberia	☒	TM Turkmenistan
☒	CO Colombia	☒	LS Lesotho	☒	TR Turkey
☒	CR Costa Rica	☒	LT Lithuania	☒	TT Trinidad and Tobago
☒	CU Cuba	☒	LU Luxembourg	☒	TZ United Republic of Tanzania
☒	CZ Czech Republic	☒	LV Latvia	☒	UA Ukraine
☒	DE Germany	☒	MA Morocco	☒	UG Uganda
☒	DK Denmark	☒	MD Republic of Moldova	☐	US United States of America
☒	DM Dominica	☒	MG Madagascar	☒	UZ Uzbekistan
☒	DZ Algeria	☒	MK The former Yugoslav Republic of Macedonia	☒	VC Saint Vincent and the Grenadines
☒	EC Ecuador	☒	MN Mongolia	☒	VN Viet Nam
☒	EE Estonia	☒	MW Malawi	☒	YU Serbia and Montenegro
☒	ES Spain	☒	MX Mexico	☒	ZA South Africa
☒	FI Finland	☒	MZ Mozambique	☒	ZM Zambia
☒	GB United Kingdom	☒	NI Nicaragua	☒	ZW Zimbabwe
☒	GD Grenada	☒	NO Norway		
☒	GE Georgia				
☒	GH Ghana				

Check-boxes below reserved for designating States which have become party to the PCT after issuance of this sheet:

☐ ..... ☐ ..... ☐ .....

☐ ..... ☐ ..... ☐ .....

Precautionary Designation Statement: In addition to the designations made above, the applicant also makes under Rule 4.9(b) all other designations which would be permitted under the PCT except any designation(s) indicated in the Supplemental Box as being excluded from the scope of this statement. The applicant declares that those additional designations are subject to confirmation and that any designation which is not confirmed before the expiration of 15 months from the priority date is to be regarded as withdrawn by the applicant at the expiration of that time limit. *(Confirmation (including fees) must reach the receiving Office within the 15-month time limit.)*

Box No. VI PRIORITY CLAIM

The priority of the following earlier application(s) is hereby claimed:

Filing date of earlier application *(day/month/year)*	Number of earlier application	Where earlier application is:		
		national application: country	regional application:* regional Office	international application: receiving Office
item (1) 23 April 2002 (23.04.02)	60/374,831	US		
item (2)				
item (3)				
item (4)				
item (5)				

☐ Further priority claims are indicated in the Supplemental Box.

The receiving Office is requested to prepare and transmit to the International Bureau a certified copy of the earlier application(s) *(only if the earlier application was filed with the Office which for the purposes of this international application is the receiving Office)* identified above as:

☐ all items ☒ item (1) ☐ item (2) ☐ item (3) ☐ item (4) ☐ item (5) ☐ other, see Supplemental Box

** Where the earlier application is an ARIPO application, indicate at least one country party to the Paris Convention for the Protection of Industrial Property or one Member of the World Trade Organization for which that earlier application was filed (Rule 4.10(b)(ii)):* .....

.....

Box No. VII INTERNATIONAL SEARCHING AUTHORITY

Choice of International Searching Authority *(ISA) (if two or more International Searching Authorities are competent to carry out the international search, indicate the Authority chosen; the two-letter code may be used):*

ISA/US

Request to use results of earlier search; reference to that search
(if an earlier search has been carried out by or requested from the International Searching Authority):

Date *(day/month/year)* Number Country *(or regional Office)*

Box No. VIII DECLARATIONS

The following declarations are contained in Boxes Nos. VIII (i) to (v) (*mark the applicable check-boxes below and indicate in the right column the number of each type of declaration*): Number of Declarations

☐ Box No. VIII (i)	Declaration as to the identity of the inventor	:
☐ Box No. VIII (ii)	Declaration as to the applicant's entitlement, as at the international filing date, to apply for and be granted a patent	:
☐ Box No. VIII (iii)	Declaration as to the applicant's entitlement, as at the international filing date, to claim the priority of the earlier application	:
☐ Box No. VIII (iv)	Declaration of inventorship *(only for the purposes of the designation of the United States of America)*	:
☐ Box No. VIII (v)	Declaration as to non-prejudicial disclosures or exceptions to lack of novelty	:

Box No. IX CHECK LIST; LANGUAGE OF FILING

This international application contains:		This international application is accompanied by the following item(s) (*mark the applicable check-boxes below and indicate in right column the number of each item*):	Number of items
(a) the following number of sheets in paper form:			
request (including declaration sheets)	: 6	1. ☒ fee calculation sheet	: 1
description (excluding sequence listing part)	: 62	2. ☐ original separate power of attorney	:
claims	: 7	3. ☐ original general power of attorney	:
abstract	: 1	4. ☒ copy of general power of attorney; reference number, if any:	: 1
drawings	: 9	5. ☐ statement explaining lack of signature	
Sub-total number of sheets :	85	6. ☐ priority document(s) identified in Box No. VI as item(s):	:
Sequence listing part of description *(actual number of sheets if filed in paper form, whether or not also filed in computer readable form; see (b) below)*	: 36	7. ☐ translation of international application into (*language*):	:
Total number of sheets	: 121	8. ☐ separate indications concerning deposited microorganism or other biological material	:
(b) sequence listing part of description filed in computer readable form		9. ☒ sequence listing in computer readable form (indicate also type and number of carriers (diskette, CD-ROM, CD-R or other))	
(i) ☐ only (under Section 801(a)(i))		(i) ☐ copy submitted for the purposes of international search under Rule 13*ter* only (and not as part of the international application)	:
(ii) ☒ in addition to being filed in paper form (under Section 801(a)(ii))		(ii) ☐ *(only where check-box (b)(i) or (b)(ii) is marked in left column)* additional copies including, where applicable, the copy of the purposes of international search under Rule 13*ter*	:
Type and number of carriers (diskette, CD-ROM, CD-R or other) on which the sequence listing part is contained *(additional copies to be indicated under item 9(ii), in right column)*: diskette.....		(iii) ☒ together with relevant statement as to the identity of the copy or copies with the sequence listing part mentioned in the left column.	: 1
		10. ☒ other (*specify*): postcard, cheque, transmittal letter	: 3
Figure of the drawings which should accompany the abstract:		Language of filing of the international application: English	

Box No. X SIGNATURE OF APPLICANT, AGENT OR COMMON REPRESENTATIVE

Next to each signature, indicate the name of the person signing and the capacity in which the person signs (if such capacity is not obvious from the request).

Sheree Lynn Rybak, Agent

For receiving Office use only

1. Date of actual receipt of the purported international application:	2. Drawings: ☐ received: ☐ not received:
3. Corrected date of actual receipt due to later but timely received papers or drawings completing the purported international application:	
4. Date of timely receipt of the required corrections under PCT Article 11(2):	
5. International Searching Authority (if two or more are competent): ISA/	6. ☐ Transmittal of search copy delayed until search fee is paid

For International Bureau use only

Date of receipt of the record copy by the International Bureau:

302
704

The demand must be filed directly with the competent International Preliminary Examining Authority or, if two or more Authorities are with the one chosen by the applicant. The full name or two-letter code of that Authority may be indicated by the applicant on the line

IPEA/ US

PCT

CHAPTER II

DEMAND

under Article 31 of the Patent Cooperation Treaty:
The undersigned requests that the international application specified below be the subject of international preliminary examination according to the Patent Cooperation Treaty and hereby elects all eligible States (except where otherwise indicated).

For International Preliminary Examining Authority use only

Identification of IPEA	Date of receipt of DEMAND

Box No. I IDENTIFICATION OF THE INTERNATIONAL APPLICATION		Applicant's or agent's file reference CGL01/0145WO1 (6682-64349)
International application No. PCT/US03/12588	International filing date *(day/month/year)* 23 April 2003 (23/04/03)	(Earliest) Priority date *(day/month/year)* 23 April 2002 (23/04/02)
Title of invention POLYPEPTIDES AND BIOSYNTHETIC PATHWAYS		

Box No. II APPLICANT(S)

Name and address: *(Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country.)*	Telephone No.
CARGILL, INCORPORATED P. . Box 5624 Minneapolis, Minnesota 55440-5624 United States of America	Facsimile No.
	Teleprinter No.
	Applicant's registration No. with the Office

State *(that is, country)* of nationality:	State *(that is, country)* of residence:
US	US

Name and address: *(Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country.)*

State *(that is, country)* of nationality:	State *(that is, country)* of residence:

Name and address: *(Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country.)*

State *(that is, country)* of nationality:	State *(that is, country)* of residence:

☐ Further applicants are indicated on a continuation sheet.

International application No.
PCT/US03/12588

Box No. III AGENT OR COMMON REPRESENTATIVE; OR ADDRESS FOR CORRESPONDENCE

The following person is ☒ agent ☐ common representative

and ☐ has been appointed earlier and represents the applicant(s) also for international preliminary examination.

☒ is hereby appointed and any earlier appointment of (an) agent(s)/common representative is hereby revoked.

☐ is hereby appointed, specifically for the procedure before the International Preliminary Examining Authority, in addition to the agent(s)/common representative appointed earlier.

Name and address: *(Family name followed by given name; for a legal entity, full official The address must include postal code and name of country.)*

LEVINE, Edward L.
Cargill, Incorporated
P.O. Box 5624
Minneapolis, Minnesota 55440-5624
United States of America

Telephone No.
952.742.4767

Facsimile No.
952.742.6349

Teleprinter No.

Agent's registration No. with the Office
28,097

☐ Address for correspondence: *Mark this check-box where no agent or common representative is/has been appointed and the space above is used instead to indicate a special address to which correspondence should be sent.*

Box No. IV BASIS FOR INTERNATIONAL PRELIMINARY EXAMINATION

Statement concerning amendments*

1. The applicant wishes the international preliminary examination to start on the basis of:

☒ the international application as originally filed

the description ☒ as originally filed
☐ as amended under Article 34

the claims ☒ as originally filed
☐ as amended under Article 19 (together with any accompanying statement)
☐ as amended under Article 34

the drawings ☒ as originally filed
☐ as amended under Article 34

2. ☐ The applicant wishes any amendment to the claims under Article 19 to be considered as reversed.

3. ☐ The applicant wishes the start of the international preliminary examination to be postponed until the expiration of 20 months from the priority date unless the International Preliminary Examining Authority receives a copy of any amendments made under Article 19 or a notice from the applicant that he does not wish to make such amendments (Rule 69.1(d)). *(This check-box may be marked only where the time limit under Article 19 has not yet expired.)*

* Where no check-box is marked, international preliminary examination will start on the basis of the international application as originally filed or, where a copy of amendments to the claims under Article 19 and/or amendments of the international application under Article 34 are received by the International Preliminary Examining Authority before it has begun to draw up a written opinion or the international preliminary examination report, as so amended.

Language for the purposes of international preliminary examination: English

☒ which is the language in which the international application was filed.
☐ which is the language of a translation furnished for the purposes of international search.
☐ which is the language of publication of the international application.
☐ which is the language of the translation (to be) furnished for the purposes of international preliminary examination.

Box No. V ELECTION OF STATES

The applicant hereby elects all eligible States *(that is, all States which have been designated and which are bound by Chapter II of the PCT)*

excluding the following States which the applicant wishes not to elect:

304
304

Sheet No. 3.

International application No.
PCT/US03/12588

Box No. VI CHECK LIST

The demand is accompanied by the following elements, in the language referred to in Box No. IV, for the purposes of international preliminary examination:			For International Preliminary Examining Authority use only: received	not received
1. translation of international application	:	sheets	☐	☐
2. amendments under Article 34	:	sheets	☐	☐
3. copy (or, where required, translation) of amendments under Article 19	:	sheets	☐	☐
4. copy (or, where required, translation) of statement under Article 19	:	sheets	☐	☐
5. letter	:	sheets	☐	☐
6. other *(specify)*	:	sheets	☐	☐

The demand is also accompanied by the item(s) marked below:

1. ☒ fee calculation sheet
2. ☐ original separate power of attorney
3. ☐ original general power of attorney
4. ☒ copy of general power of attorney; reference number, if any: Agent Under Rule 90 and Request for Recordation of the Revocation and Appointment of an
5. ☐ statement explaining lack of signature
6. ☐ sequence listings in computer readable form
7. ☐ tables in computer readable form related to sequence listings
8. ☒ other *(specify)*: Transmittal Letter; copy of

Box No. VII SIGNATURE OF APPLICANT, AGENT OR COMMON REPRESENTATIVE

Next to each signature, indicate the name of the person signing and the capacity in which the person signs (if such capacity is not obvious from reading the demand).

M. Angela Parsons, Reg. No. 44,282
Agent for Applicant

For International Preliminary Examining Authority use only

1. Date of actual receipt of DEMAND:
2. Adjusted date of receipt of demand due to CORRECTIONS under Rule 60.1(b):
3. ☐ The date of receipt of the demand is AFTER the expiration of 19 months from the priority date and item 4 or 5, below, does not apply. ☐ The applicant has been informed accordingly.
4. ☐ The date of receipt of the demand is WITHIN the period of 19 months from the priority date as extended by virtue of Rule 80.5.
5. ☐ Although the date of receipt of the demand is after the expiration of 19 months from the priority date, the delay in arrival is EXCUSED pursuant to Rule 82.

For International Bureau use only

Demand received from IPEA on:

PATENT COOPERATION TREATY

From the INTERNATIONAL SEARCHING AUTHORITY

To:
JEFFREY J. SKELTON
CARGILL, INCORPORATED
LAW DEPARTMENT
P.O. BOX 5624
MINNEAPOLIS, MN 55440-5624

PCT

NOTIFICATION OF TRANSMITTAL OF THE INTERNATIONAL SEARCH REPORT OR THE DECLARATION

(PCT Rule 44.1)

	Date of Mailing (day/month/year) 21 OCT 2003
Applicant's or agent's file reference 6682-64349	FOR FURTHER ACTION See paragraphs 1 and 4 below
International application No. PCT/US03/12588	International filing date (day/month/year) 23 April 2003 (23.04.2003)
Applicant CARGILL, INCORPORATED	

1. ☒ The applicant is hereby notified that the international search report has been established and is transmitted herewith.

   Filing of amendments and statement under Article 19:
   The applicant is entitled, if he so wishes, to amend the claims of the international application (see Rule 46):

   When? The time limit for filing such amendments is normally two months from the date of transmittal of the international search report.

   Where? Directly to the International Bureau of WIPO, 34, chemin des Colombettes 1211 Geneva 20, Switzerland, Facsimile No.: (41-22) 740.14.35

   For more detailed instructions, see the notes on the accompanying sheet.

2. ☐ The applicant is hereby notified that no international search report will be established and that the declaration under Article 17(2)(a) to that effect is transmitted herewith.

3. ☐ With regard to the protest against payment of (an) additional fee(s) under Rule 40.2, the applicant is notified that:

   ☐ the protest together with the decision thereon has been transmitted to the International Bureau together with the applicant's request to forward the texts of both the protest and the decision thereon to the designated Offices.

   ☐ no decision has been made yet on the protest; the applicant will be notified as soon as a decision is made.

4. Reminder

   Shortly after 18 months from the priority date, the international application will be published by the International Bureau. If the applicant wishes to avoid or postpone publication, a notice of withdrawal of the international application, or of the priority claim, must reach the International Bureau as provided in Rules 90 *bis*.1 and 90 *bis*.3, respectively, before the completion of the technical preparations for international publication.

   Within 19 months from the priority date, but only in respect of some designated Offices, a demand for international preliminary examination must be filed if the applicant wishes to postpone the entry into the national phase until 30 months from the priority date (in some Offices even later); otherwise the applicant must, within 20 months from the priority date, perform the prescribed acts for entry into the national phase before those designated Offices.

   In respect of other designated Offices, the time limit of 30 months (or later) will apply even if no demand is filed within 19 months.

   See the Annex to Form PCT/IB/301 and, for details about the applicable time limits, Office by Office, see the *PCT Applicant's Guide*, Volume II, National Chapters and the WIPO Internet site.

Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US Commissioner for Patents P.O. Box 1450 Alexandria, Virginia 22313-1450 Facsimile No. (703)305-3230	Authorized officer Kathleen M Kerr Telephone No. (703) 308-0196

Form PCT/ISA/220 (April 2002) (See notes on accompanying sheet)

306

PATENT COOPERATION TREATY

PCT

INTERNATIONAL SEARCH REPORT

(PCT Article 18 and Rules 43 and 44)

Applicant's or agent's file reference 6682-64349	FOR FURTHER ACTION	see Notification of Transmittal of International Search Report (Form PCT/ISA/220) as well as, where applicable, item 5 below.
International application No. PCT/US03/12588	International filing date (day/month/year) 23 April 2003 (23.04.2003)	(Earliest) Priority Date (day/month/year) 23 April 2002 (23.04.2002)
Applicant CARGILL, INCORPORATED		

This international search report has been prepared by this International Searching Authority and is transmitted to the applicant according to Article 18. A copy is being transmitted to the International Bureau.

This international search report consists of a total of 6 sheets.

☒ It is also accompanied by a copy of each prior art document cited in this report.

1. Basis of the Report
   a. With regard to the language, the international search was carried out on the basis of the international application in the language in which it was filed, unless otherwise indicated under this item.
   ☐ the international search was carried out on the basis of a translation of the international application furnished to this Authority (Rule 23.1(b)).
   b. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of the sequence listing:
   ☒ contained in the international application in written form.
   ☒ filed together with the international application in computer readable form.
   ☐ furnished subsequently to this Authority in written form.
   ☐ furnished subsequently to this Authority in computer readable form.
   ☐ the statement that the subsequently furnished written sequence listing does not go beyond the disclosure in the international application as filed has been furnished.
   ☐ the statement that the information recorded in computer readable form is identical to the written sequence listing has been furnished.
2. ☐ Certain claims were found unsearchable (See Box I).
3. ☒ Unity of invention is lacking (See Box II).
4. With regard to the title,
   ☒ the text is approved as submitted by the applicant.
   ☐ the text has been established by this Authority to read as follows:
5. With regard to the abstract,
   ☒ the text is approved as submitted by the applicant.
   ☐ the text has been established, according to Rule 38.2(b), by this Authority as it appears in Box III. The applicant may, within one month from the date of mailing of this international search report, submit comments to this Authority.
6. The figure of the drawings to be published with the abstract is Figure No. ____
   ☐ as suggested by the applicant. ☒ None of the figures
   ☐ because the applicant failed to suggest a figure.
   ☐ because this figure better characterizes the invention.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US03/12588

Box I Observations where certain claims were found unsearchable (Continuation of Item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claim Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claim Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claim Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of Item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:
Please See Continuation Sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☒ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.: 1 and 26-33

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest
☐ The additional search fees were accompanied by the applicant's protest.
☐ No protest accompanied the payment of additional search fees.

308
24

INTERNATIONAL SEARCH REPORT

PCT/US03/12588

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING

This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fees must be paid.

Group I, claim(s) 1, drawn to methods of producing mumain by converting glucose to mumain.

Group II, claim(s) 2-23, 25, 41-43 and 55-57, drawn to methods of producing mumain using tryptophan.

Group III, claim(s) 2-23, 41-43, and 55, drawn to methods of producing mumain using indole-3-lactic acid.

Group IV, claim(s) 2-23, 25, and 55, drawn to methods of producing mumain using tryptophan and indole-3-lactic acid.

Group V, claim(s) 2-23, 24, 41-43, and 55, drawn to methods of producing mumain using indole-3-pyruvate.

Group VI, claim(s) 2-23, 41-43, 52, 53, and 55, drawn to methods of making mumain using 2-hydroxy 2-(indol-3-ylmethyl)-4-keto glutaric acid.

Group VII, claim(s) 26-33, drawn to cells for producing mumain comprising exogenous nucleic acid sequences.

Group VIII, claim(s) 34-35 and 38, drawn to known tryptophanase polypeptides having particular mutations.

Group IX, claim(s) 36-38, drawn to known beta-tyrosinase polypeptides having particular mutations.

Group X, claim(s) 39-40, drawn to methods of making mumain using an alanine derivative and an indole-3-pyruvate derivative.

Group XI, claim(s) 44-51, drawn to nucleic acids, vectors, recombinant cells and proteins related to SEQ ID NOs: 11 and/or 12, an aromatic aminotransferase from *Lactobacillus amylovorus*.

Group XII, claim(s) 54, drawn to a composition comprising 2-hydroxy 2-(indol-3-ylmethyl)-4-keto glutaric acid.

The inventions listed as Groups I-XII do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons.

The special technical feature of Group I is the particular substrate, glucose, to produce the particular product, mumain; no process steps are defined. Production of mumain is known in the art (Nakamura *et al.* Total Synthesis of Mumain. Organic Letters (2000) 2(19):2967-2970) from other substrates; therefore, it is the combination of glucose-to-mumain that comprises the special technical feature of the claim. No other claims share this special technical feature.

Groups II-VI and X all share the technical feature of Group I, that is making mumain. However, these Groups are use different starting material in the process and, thus, do not share the special technical feature of Group I that is using glucose.

The products of Group VII share the technical feature with Group I, that is making mumain. However, no requirement on the substrate used by the cells to make mumain is noted. Thus, the products of Group VII do not share a special technical feature with Group I.

The products of Groups VIII and IX share a technical feature with Group I, that is the polypeptide products can be used in the making of mumain. However, no requirement that the polypeptides must use glucose in the production of mumain and/or its intermediates is noted. Thus, the products of Groups VIII and IX do not share a special technical feature with Group I.

INTERNATIONAL SEARCH REPORT

PCT/US03/12588

The products of Group XI share a technical feature with Group I, that is the nucleic acids, vectors, host cells, and polypeptide products can be used in the making of mumain. However, as requirement that these products must use glucose in the production of mumain and/or its intermediates is noted. Thus, the products of Group XI do not share a special technical feature with Group I.

The product of Group XII shares a technical feature with Group I, that is the product is an intermediate in the disclosed production of glucose as mumain. However, the special technical feature of Group I does not include the disclosed method steps since mans are required in the claim. Thus, the intermediate product of Group XII does not share the special technical feature of Group I since it is neither glucose nor mumain.

Resp-Search Report
10-21-03
12-21-03
12-21-03

Foreign Art
10-21-03
1-21-04
1-21-04

I	12/21/03	1/21/04
I	12/21/03	1/21/04
I	SPE 11/21/03	SPE 11/21/03

310

INTERNATIONAL SEARCH REPORT

International application No.
PCT US03 12588

A. CLASSIFICATION OF SUBJECT MATTER
IPC(7) : C07D 309 18; C12P 17 10
US CL : 548 495; 435 121
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
U.S. : 548 495; 435 121

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
CAplus: mumin, polypeptide, protein

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	NAKAMURA et al. Total Synthesis of Mumin. Organic Letters. 2000, Vol. 2, No. 19, pages 2967-2970.	1
A	US 5,994,599 A (ABUSHANAB et al) 30 November 1999 (30.11.1999).	1
A	HOLZAPFEL et al. A simple cycloaddition approach to a racemate of the natural sweetner mumin. Synthetic Communications. 1994, Vol. 24, No. 22, pages 3197-3211.	1
A	HOLZAPFEL et al. The synthesis of a gamma-keto-alpha-amino acid, a key intermediate in the synthesis of mumin, a new natural sweetner. Synthetic Communications. 1993, Vol. 23, No. 18, pages 2511-2526.	1
A	VLEGGAAR et al. Structure elucidation of mumin, a high-intensity sweetner isolated from the plant Schlerochiton ilicifolius. J. Chem. Soc., Perkin Transactions 1: Organic and Bio-Organic Chemistry (1972-1999). 1992, Vol. 22, pages 3085-3088.	1
X	PATIL et al. Cloning, nucleotide sequence, overexpression, and inactivation of the Escherichia coli 2-keto-4-hydroxyglutarate aldolase gene. J. Bacteriol. January 1992, Vol. 174, No. 1, pages 103-107, especially abstract.	28-31

☒ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
04 September 2003 (04.09.2003)	21 OCT 2003
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US Commissioner for Patents P.O. Box 1450 Alexandria, Virginia 22313-1450 Facsimile No. (703)305-3230	Authorized officer Kathleen M Kerr Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet) (July 1998)

PCT/US03/12588

INTERNATIONAL SEARCH REPORT

C. (Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	GALKIN et al. Synthesis of optically active amino acids from alpha-keto acids with Escherichia coli cells expressing heterologous genes. Applied and Environmental Microbiology. December 1997, Vol. 63, No. 12, pages 4651-4656, entire document.	26-29, 31, and 33
X	DELUNA et al. NADP-Glutamate Dehydrogenase Isoenzymes of Saccharomyces cerevisiae: Purification, Kinetic Properties, and Physiological Roles. J. Biol. Chem. 23 November 2001, Vol. 276, No. 47, pages 43775-43783, entire document.	26-29 and 32

Form PCT/ISA/210 (second sheet) (July 1998)

312

CHAPTER I
PCT TELEPHONE MEMORANDUM
FOR
LACK OF UNITY OF INVENTION

PCT No.: PCT/US03/12588

Examiner: Kathleen M Kerr

Attorney spoken to: Paula DeGrandis

Date of call: 02 September 2003

☒ Amount of payment approved: $210

☒ Deposit account number to be charged: 02-4550

☐ Attorney elected to pay for ALL additional inventions

☒ Attorney elected to pay only for the additional inventions covered by

☒ Group(s): VII

– encompassing –

☒ Claim(s): 26-33

☐ Attorney elected NOT to pay for any additional inventions, therefore, only the first claimed invention (Group I) covered by Claim(s) _____ has been searched.

☒ Attorney was orally advised that there is no right to protest for any group not paid for.

☒ Attorney was orally advised that any protest must be filed no later than 15 days from the mailing of the Search Report (PCT/ISA/210).

Time Limit For Filing A Protest

Applicant is hereby given 15 days from the mailing date of this Search Report in which to file a protest of the holding of lack of unity of invention. In accordance with PCT Rule 40.2, applicant may protest the holding of lack of unity only with respect to the group(s) paid for.

Detailed Reasons For Holding Lack of Unity of Invention:
Please See Continuation Sheet

Note: A copy of this form must be attached to the Search Report.

313

International application No: PCT/US03/12588

ATTACHMENT TO CHAPTER I PCT TELEPHONE MEMORANDUM FOR LACK OF UNITY OF INVENTION

Continuation of Detailed Reasons For Holding Lack of Unity of Invention:

This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fees must be paid.

Group I, claim(s) 1, drawn to methods of producing monatin by converting glucose to monatin.

Group II, claim(s) 2-23, 25, 41-43 and 55-57, drawn to methods of producing monatin using tryptophan.

Group III, claim(s) 2-23, 41-43, and 55, drawn to methods of producing monatin using indole-3-lactic acid.

Group IV, claim(s) 2-23, 25, and 55, drawn to methods of producing monatin using tryptophan and indole-3-lactic acid.

Group V, claim(s) 2-23, 24, 41-43, and 55, drawn to methods of producing monatin using indole-3-pyruvate.

Group VI, claim(s) 2-23, 41-43, 52, 53, and 55, drawn to methods of making monatin using 2-hydroxy 2-(indol-3-ylmethyl)-4-keto glutaric acid.

Group VII, claim(s) 26-33, drawn to cells for producing monatin comprising exogenous nucleic acid sequences.

Group VIII, claim(s) 34-35 and 38, drawn to known tryptophanase polypeptides having particular mutations.

Group IX, claim(s) 36-38, drawn to known beta-tyrosinase polypeptides having particular mutations.

Group X, claim(s) 39-40, drawn to methods of making monatin using an alanine derivative and an indole-3-pyruvate derivative.

Group XI, claim(s) 44-51, drawn to nucleic acids, vectors, recombinant cells and proteins related to SEQ ID NOs: 11 and/or 12, an aromatic aminotransferase from *Lactobacillus amylovorus*.

Group XII, claim(s) 54, drawn to a composition comprising 2-hydroxy 2-(indol-3-ylmethyl)-4-keto glutaric acid.

The inventions listed as Groups I-XII do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons.

The special technical feature of Group I is the particular substrate, glucose, to produce the particular product, monatin; no process steps are defined. Production of monatin is known in the art (Nakamura *et al.* Total Synthesis of Monatin. Organic Letters (2000) 2(19):2967-2970) from other substrates; therefore, it is the combination of glucose-to-monatin that comprises the special technical feature of the claim. No other claims share this special technical feature.

Note: A copy of this form must be attached to the Search Report.

314 40

Groups II-VI and X all share the technical feature of Group I, that is making monatin. However, these Groups are use different starting material in the process and, thus, do not share the special technical feature of Group I that is using glucose.

The products of Group VII share the technical feature with Group I, that is making monatin. However, no requirement on the substrate used by the cells to make monatin is noted. Thus, the products of Group VII do not share a special technical feature with Group I.

The products of Groups VIII and IX share a technical feature with Group I, that is the polypeptide products can be used in the making of monatin. However, no requirement that the polypeptides must use glucose in the production of monatin and/or its intermediates is noted. Thus, the products of Groups VIII and IX do not share a special technical feature with Group I.

The products of Group XI share a technical feature with Group I, that is the nucleic acids, vectors, host cells, and polypeptide products can be used in the making of monatin. However, no requirement that these products must use glucose in the production of monatin and/or its intermediates is noted. Thus, the products of Group XI do not share a special technical feature with Group I.

The product of Group XII shares a technical feature with Group I, that is the product is an intermediate in the disclosed production of glucose to monatin. However, the special technical feature of Group I does not include the disclosed method steps since none are required in the claim. Thus, the intermediate product of Group XII does not share the special technical feature of Group I since it is neither glucose nor monatin.

Note: A copy of this form must be attached to the Search Report.

UNITED STATES RECEIVING OFFICE (RO/US) FEE CODING AND RECORDING SHEET

☐ ADD'L SHEETS

IDENTIFICATION OF THE INTERNATIONAL APPLICATION

INTERNATIONAL APPLICATION NUMBER	INTERNATIONAL FILING DATE
PCT/US03/12588	23.04.2003

APPLICANT (Name): CARGILL, INC.

	PAYMENTS				REFUNDS	
	Payment on Filing	Deposit Account	Deposit Account	To Deposit Account	To Deposit Account	
		02 4550				
Deposit Account		DATE: 02.09.2003	DATE:	DATE:	DATE:	
	☐ CASH/CHECK	☒ CASH/CHECK	☐ CASH/CHECK	☐ BY CHECK	☐ BY CHECK	
180		1604 $210.00				
181						
182						
800						
801						
802						
803						
	Total Paid:	Total Paid: $210.00	Total Paid:	Total Refunded:	Total Refunded:	
States included for 802:		802:	802:			
States included for 803:		803:	803:			

Date Mailed:

RO/US Authorization	RO/US Authorization	RO/US Authorization	RO/US Authorization	RO/US Authorization
	K.M.			

316

NOTES TO FORM PCT/ISA/220

These Notes are intended to give the basic instructions concerning the filing of amendments under Article 19. The Notes are based on the requirements of the Patent Cooperation Treaty, the Regulations and the Administrative Instructions under that Treaty. In case of discrepancy between these Notes and those requirements, the latter are applicable. For more detailed information, see also the *PCT Applicant's Guide*, a publication of WIPO.

In these Notes, "Article," "Rule" and "Section" refer to the provisions of the PCT, the PCT Regulations and the PCT Administrative Instructions, respectively.

INSTRUCTIONS CONCERNING AMENDMENTS UNDER ARTICLE 19

The applicant has, after having received the international search report, one opportunity to amend the claims of the international application. It should however be emphasized that, since all parts of the international application (claims, description and drawings) may be amended during the international preliminary examination procedure, there is usually no need to file amendments of the claims under Article 19 except where, e.g. the applicant wants the latter to be published for the purposes of provisional protection or has another reason for amending the claims before international publication. Furthermore, it should be emphasized that provisional protection is available in some States only.

What parts of the international application may be amended?

Under Article 19, only the claims may be amended.

During the international phase, the claims may also be amended (or further amended) under Article 34 before the International Preliminary Examining Authority. The description and drawings may only be amended under Article 34 before the International Preliminary Examining Authority.

Upon entry into the national phase, all parts of the international application may be amended under Article 28 or, where applicable, Article 41.

When? Within 2 months from the date of transmittal of the international search report or 16 months from the priority date, whichever time limit expires later. It should be noted, however, that the amendments will be considered as having been received on time if they are received by the International Bureau after the expiration of the applicable time limit but before the completion of the technical preparations for international publication (Rule 46.1).

Where not to file the amendments?

The amendments may only be filed with the International Bureau and not with the receiving Office or the International Searching Authority (Rule 46.2).

Where a demand for international preliminary examination has been/is filed, see below

How? Either by cancelling one or more entire claims, by adding one or more new claims or by amending the text of one or more of the claims as filed.

A replacement sheet must be submitted for each sheet of the claims which, on account of an amendment or amendments, differs from the sheet originally filed.

All the claims appearing on a replacement sheet must be numbered in Arabic numerals. Where a claim is cancelled, no renumbering of the other claims is required. In all cases where claims are renumbered, they must be renumbered consecutively (Administrative Instructions, Section 205(b)).

The amendments must be made in the language in which the international application is to be published.

What documents must/may accompany the amendments?

Letter (Section 205(b)):

The amendments must be submitted with a letter.

The letter will not be published with the international application and the amended claims. It should not be confused with the "Statement under Article 19(1)" (see below, under "Statement under Article 19(1)").

The letter must be in English or French, at the choice of the applicant. However, if the language of the international application is English, the letter must be in English; if the language of the international application is French, the letter must be in French.

ORGANIC LETTERS
2000
Vol. 2, No. 19
2967–2970

Total Synthesis of Monatin

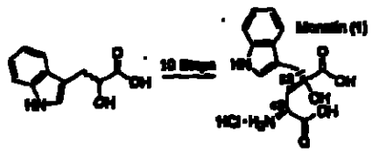
Kozo Nakamura,[†] Tracy J. Baker, and Murray Goodman*

Department of Chemistry and Biochemistry, University of California, San Diego, 0223 Pacific Hall, La Jolla, California 92093-0343

mgoodman@ucsd.edu

Received June 23, 2000

ABSTRACT

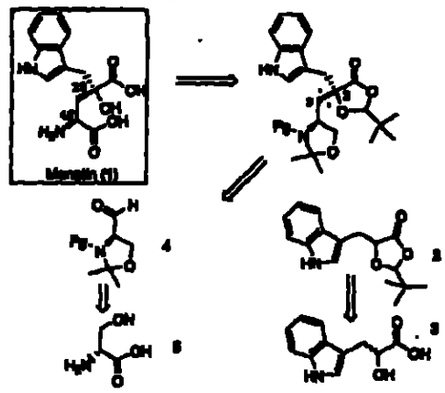


The naturally occurring sweetener Monatin, a diastereomer of Monatin, and a phenyl analogue of Monatin have been prepared and isolated in their enantiomerically pure forms.

The sweet, naturally occurring, unusual amino acid Monatin [1, (2*S*,4*S*)-4-hydroxy-4-(indol-3-ylmethyl)glutamic acid] was isolated in South Africa from *Schlerochiton ilicifolius*.[1] This small molecule exhibits a potent sweet taste, i.e., about 1000 times sweeter than sucrose on a weight basis[1] and is an attractive target to synthesize as a molecule exhibiting a sweet taste. The syntheses of racemates and an analogue of Monatin have been reported.[2] In this Letter, we report the synthesis of optically pure Monatin (1). For this undertaking, we utilized the strategy outlined in Scheme 1. In the retrosynthetic analysis shown, a bond is formed between a derivative of indolelactic acid (2) and an optically active protected oxazolidine 4. These reactants are readily formed from commercially available racemic indolelactic acid (3) and L-serine (5), respectively.

Monatin contains two asymmetric centers at C2 and C4. The C4 center has an α-proton which is easily epimerized under basic conditions. Therefore, we employed the highly reactive Garner aldehyde (4) to react with the enolate of compound 2, thus avoiding epimerization of aldehyde 4. With one asymmetric center fixed and the removal of the hydroxyl at C3, we obtained two diastereomers in our synthesis. The separation of the final intermediate 19a from its diastereomer 19b was readily accomplished.

Scheme 1. Strategy for the Synthesis of Monatin (1)[a]



[a] Pg = protecting group.

Before we undertook the synthesis of Monatin (1), we applied the strategy to the preparation of the phenyl analogue 13 in which a phenyl ring replaces the indole group. This enantioselective route began with commercially available D-phenyllactic acid (6, Scheme 2). In 1984, Seebach reported the enantiospecific alkylations of several 1,3-dioxolanones

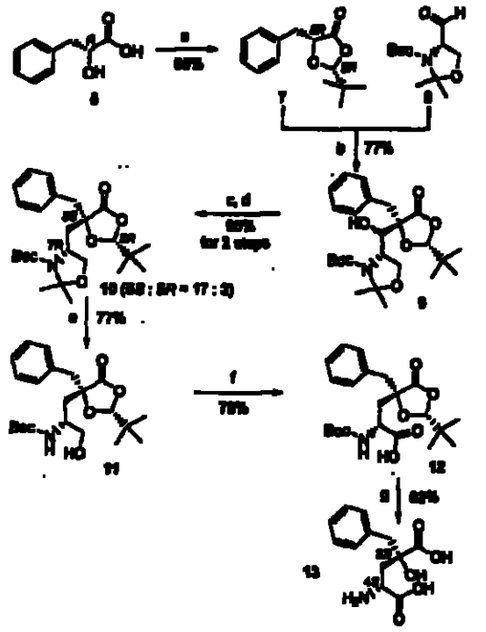
[†] Research Fellow of the Japan Society for the Promotion of Science.

(1) Vleggaar, R.; Ackermans, L. G. J.; Steyn, P. S. *J. Chem. Soc., Perkin Trans. 1* **1982**, *22*, 3095–3098.

(2) (a) Holmquist, C. W.; Hochstetler, K.; Olivier, *J. Synth. Commun.* **1994**, *24*(22), 3197–3211. (b) Abushanab, E.; Avramoglou, S. U.S. Patent 5,994,559, November 30, 1999. (c) Holmquist, C. W.; Olivier, *J. Synth. Commun.* **1993**, *23*(18), 2511–2526.

10.1021/ol006240q CCC: $19.00
Published on Web 08/29/2000

Scheme 2. Synthesis of a Phenyl Analogue of Monastin (13)



[a] (a) $(CH_3)_2CCHO$ (1.2 equiv), $BF_3{\cdot}Et_2O$ (1.2 equiv), Et_2O, −20 °C, 2 h; (b) **7** (1.0 equiv), **8** (1.1 equiv), LDA (1.2 equiv), THF, −78 °C, 2 h; (c) NaH (3.0 equiv), THF, −15 °C → rt, 40 min, CS_2 (3.0 equiv), 0 °C, 20 min, MeI (3.0 equiv), rt, 30 min; (d) *n*-Bu_3SnH (6.0 equiv), AIBN (3.0 equiv), Ph-Cl_2, Δ, 2 h, Ar; (e) PPTS (1.0 equiv × 3), EtOH/H_2O (95:5), Δ, 6 h; (f) PDC (6.0 equiv), DMF, rt, 24 h; (g) 0.1 N HCl/HCO_2H, Δ, 3 h. LDA = lithium diisopropylamide, AIBN = 2,2′-azobis(isobutyronitrile), PPTS = pyridinium *p*-toluenesulfonate, PDC = pyridinium dichromate, DMF = *N,N*′-dimethylformamide.

prepared from α-hydroxycarboxylic acids and pivalaldehyde through the *Li*-enolate.[3] This reaction has been applied to many total syntheses.[4] We used this method as a key step in the synthesis of a phenyl analogue of Monastin (13). The optically pure *cis*-dioxolanone **7** was obtained in 95% yield.[5] Alkylation at the α-carbon of compound **7** was unsuccessful using iodoalanine derivatives. We believe that the sterically hindered enolate of compound **7** cannot react with the bulky iodoalanine derivatives in a displacement reaction.[6] The Garner aldehyde (**8**)[7] derived from L-serine did react with the enolate of compound **7** generated with LDA at −78 °C to produce condensation product **9**. This molecule possesses a β-hydroxyl group which was removed by Barton deoxygenation[4,8] to provide the deoxygenated product **10** with excellent diastereoselectivity (17:3) as determined by ^1H NMR spectroscopy. Selective deprotection with PPTS in refluxing ethanol[9] afforded the Boc-protected β-amino alcohol **11** in good yield. The primary alcohol **11** was oxidized to carboxylic acid **12** with PDC without epimerization at the C4 position.[9] Compound **12** and the small amount of diastereomer were separable, and optically pure product **12** was obtained. We established by NMR spectroscopy that the major product **12** possessed the (2*S*,4*R*) configuration.[4,5] Final deprotection of the Boc and pivalidene protections of compound **12** was afforded by allowing the reaction to reflux in 0.1 N HCl/HCO_2H for 3 h. Subsequent purification provided the (2*S*,4*R*) phenyl analogue of Monastin (**13**) which did not possess a sweet taste.

Since separation of the diastereomers of the phenyl analogue **12** was successful, we applied this strategy to the preparation of Monastin (**1**, Scheme 3). This synthesis commences with commercially available racemic isoleucinic acid (**3**). The pivalidene derivatives **2** were obtained in good yield by following the same procedure as that reported above for the preparation of compound **7**. Reaction of the appropriate enantiomer of the Garner aldehyde (**14**) with the enolates of intermediates **2** produced compounds **15**. In this reaction, dimerization of intermediates **2** can readily occur. To avoid this side reaction, it was necessary to maintain the reaction temperature below −76 °C. Deoxygenation of isomers **15** produced the diastereomeric pair **16**.[10] Deprotection of the pivalidene group gave the hydroxymethylene molecules **17a** and **17b**[11] which were readily oxidized to the carboxylic acid structures **18a** and **18b**[12] using PDC. These reactions were successfully carried out using reaction conditions similar to those described above for the preparation of a phenyl analogue of Monastin, compound **13**.

The final deprotection steps used for the phenyl analogue could not be applied to Monastin (**1**) because decomposition of the indole group occurs under the acidic conditions employed. Compounds **18a** and **18b** were successfully converted to the diastereomeric mixture of lactams **19a** and

(2) (a) Seebach, D.; Naef, R. *Helv. Chim. Acta* **1981**, *64*, 2704. (b) Seebach, D.; Naef, R.; Calderari, G. *Tetrahedron* **1984**, *40*(8), 1313–1324.

(3) (a) Hiwa, H.; Ogawa, T.; Yamada, K. *Bull. Chem. Soc. Jpn.* **1990**, *63*, 3707–3708. (b) Rousseau, R. K., Jr.; Yoon, S. K.; Heckendorn, D. K. *J. Am. Chem. Soc.* **1991**, *113*, 9682–9684. (c) Bojack, G.; Romowski, H. *Tetrahedron* **1991**, *47*(44), 9179–9186. (d) McPherson, D. W.; Knapp, F. F., Jr. *J. Org. Chem.* **1996**, *61*, 8335–8337. (e) Visser, T. I.; Wanrik, A. V.; Jansen, T. J. H.; Visser, G. M.; Mark, T. W. V. D.; Kruse, J.; Essing, K.; Vanberg, W. *J. Med. Chem.* **1997**, *40*, 117–124.

(4) (a) Fenton, M.; Soulier, J. *Bull. Soc. Chim. Fr.* **1970**, 332. (b) Hoye, T. R.; Peterson, B. H.; Miller, J. D. *J. Org. Chem.* **1987**, *52*, 1351. (c) Pearson, W. H.; Chang, M. *J. J. Org. Chem.* **1987**, *52*, 1353. (d) Ostlund, J. Y.; Greiner, A. *J. Bull. Soc. Chim. Fr.* **1993**, *130*, 133. (e) Nicas, A. P.; Shao-Po, L. *J. Am. Chem. Soc.* **1995**, *117*, 6394–6395.

(6) (a) Forsyth, C. J.; Sabes, S. F.; Urbanek, R. A. *J. Am. Chem. Soc.* **1997**, *119*, 8181–8182. (b) Sabes, S. F.; Urbanek, R. A.; Forsyth, C. J. *J. Am. Chem. Soc.* **1998**, *120*, 2534–2543.

(7) (a) Garner, P.; Park, J. M. *J. Org. Chem.* **1987**, *52*, 2361. (b) Garner, P.; Park, J. M. *Org. Synth.* **1991**, *70*, 18.

(8) Barton, D. H. R.; McCombie, S. W. *J. Chem. Soc., Perkin Trans. 1* **1975**, 1574.

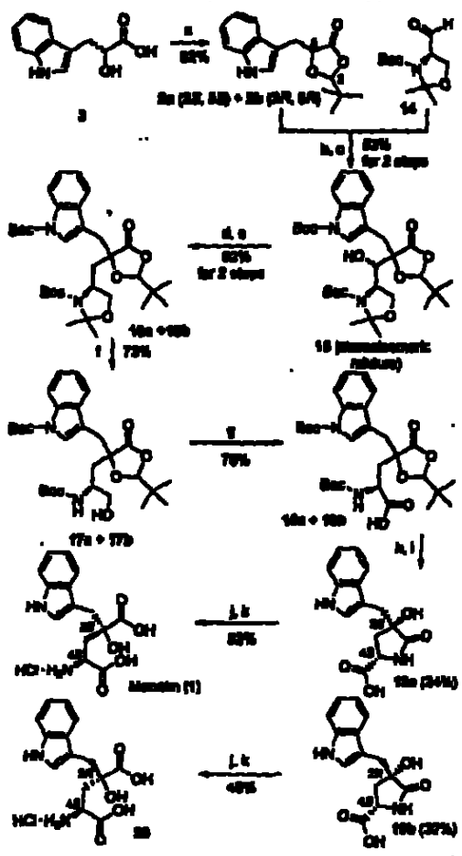
(9) (a) Jahn, U.; Ulbar, P.; Mann, A.; Wermuth, C.-G.; Steinberger, T.; Neuburg, B.; Brunk, G.; Durand, F. *J. Org. Chem.* **1991**, *56*, 5720. (b) Müller, M.; Mann, A.; Taddei, M. *Tetrahedron Lett.* **1995**, *36*, 3289. (c) D'Aniello, F.; Mann, A.; Taddei, M.; Wermuth, C.-G. *Tetrahedron Lett.* **1994**, *42*, 7773. (d) D'Aniello, F.; Falorni, M.; Mann, A.; Taddei, M. *Tetrahedron: Asymmetry* **1996**, *7*(4), 1217–1226.

(10) The diastereomeric ratio was not established.

(11) The ratio of diastereomers was approximately 3:2 as determined by ^1H NMR spectroscopy.

(12) The ratio of diastereomers was approximately 4:3 as determined by HPLC.

Scheme 3. Synthesis of Monatin (1)[a]



[a] (a) $(CH_3)_2CCHO$ (1.2 equiv), BF_3-Et_2O (1.2 equiv), Et_2O, −20 °C, 2 h; (b) 2 (1.2 equiv) 14 (1.0 equiv), LDA (1.3 equiv), THF, −78 °C, 2 h; (c) $(Boc)_2O$ (1.1 equiv), DMAP (0.1 equiv), CH_3CN, rt; (d) NaH (5.0 equiv), THF, −15 °C → rt, 40 min, CS_2 (5.0 equiv), 0 °C, 20 min; (e) *n*-Bu_3SnH (10.0 equiv), AIBN (5.0 equiv), Ph-CH_3, Δ, 2 h, Ar; (f) PPTS (0.8 equiv × 4), EtOH/H_2O (95:5), Δ, 6 h; (g) PDC (6.0 equiv), DMF, rt, 24 h; (h) 0.1 N HCl/HCO_2H, rt, 3 h; (i) 3 N KOH (10 equiv), 0 °C, 1 min; (j) 3 N NaOH (5 equiv), EtOH, Δ, 3 h; (k) HCl. DMAP = 4-(dimethylamino)pyridine.

19b[17] by treatment with 0.1 N HCl/HCO_2H at room temperature for 3 h followed by 3 N KOH at 0 °C for 1 min. In this reaction, only lactam formation was observed with no ring-opened products detected.

Diastereomers 19a and 19b were easily separated by RP-HPLC using a C18 column. Optically pure lactam 19a which possessed the same asymmetric configuration as Monatin (1) was refluxed in the presence of 3 N NaOH in ethanol. No

(13) The ratio of diastereomers was approximately 4:3 as determined by product yields.

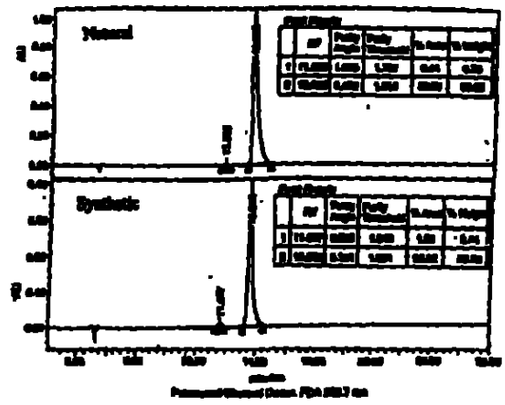


Figure 1. HPLC charts of natural and synthetic Monatin (1). The original scans are shown.

epimerization at C4 was detected by HPLC. The ring opening of lactam 19a after acidification led directly to synthetic Monatin (1) which chromatographed identically with natural Monatin (Figure 1). In addition, the ^1H NMR spectrum of naturally isolated Monatin is identical to the spectrum of synthetic Monatin (1, Figure 2), and HRMALDI established that the synthetic product possesses the same formula as natural Monatin. Last, the optical rotation $([\alpha]^{25}_D)$ of synthetic Monatin is consistent with that reported for natural Monatin [− 7.6 (lit.) vs − 8.8 (experimentally determined under the same conditions)]. Indeed, synthetic Monatin (1) exhibits a sweet potency equal to that of the natural product.[11] The diastereomer of Monatin, compound 20, obtained in this synthesis exhibits a slightly sweet taste which may be due to the presence of a trace amount of Monatin (1). Were

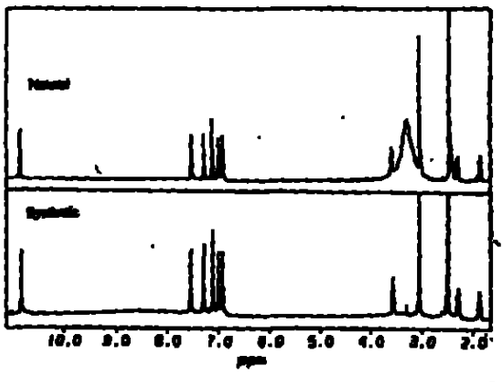


Figure 2. 500 MHz ^1H NMR spectra of natural and synthetic Monatin (1).

320
322

optically pure D-indolelactic acid available, only pure synthetic Monatin (1) would have been produced.

In conclusion, we have reported the total synthesis of Monatin (1), the diastereomer of Monatin (20) and a phenyl analogue of Monatin (13). The syntheses of these target molecules will prove to be most useful in our studies of the molecular basis of the sweet taste.

Acknowledgment. We thank Dr. S. Jiang, J. Kwak, L. Zhang, M. Tomioka, and J. Tanone for NMR, collection of physical data, and useful discussions. We are also grateful to Dr. Louis Ackerman for a sample of natural Monatin. This work was supported, in part, by Ajinomoto Co., Inc.

Supporting Information Available: Experimental procedures with spectroscopic data for compounds 1, 15, 19a, and 19b. This material is available free of charge via the Internet at http://pubs.acs.org.

OL000244Q

US005994559A

United States Patent [19]

Abushanab et al.

[11] Patent Number: 5,994,559

[45] Date of Patent: Nov. 30, 1999

[54] **SYNTHESIS OF MONATIN-A HIGH INTENSITY NATURAL SWEETENER**

[75] Inventors: **Elie Abushanab**, Peace Dale; **Subramanian Arumugam**, Exeter, both of R.I.

[73] Assignee: **The Board of Governors for Higher Education, State of Rhode Island and Providence Plantations**, Providence, R.I.

[21] Appl. No.: 09/131,885

[22] Filed: Aug. 6, 1998

[51] Int. Cl.[6] C07D 209/18

[52] U.S. Cl. 548/495

[58] Field of Search 548/495

[56] References Cited

U.S. PATENT DOCUMENTS

4,975,298	12/1990	Van Wyk et al. .	
5,128,164	7/1992	Van Wyk et al. .	
5,128,482	7/1992	Olivier et al. .	

FOREIGN PATENT DOCUMENTS

0 438 314 A1 7/1991 European Pat. Off. .

OTHER PUBLICATIONS

Holmquist, et al., A simple cycloaddition approach to a racemate of the natural sweetener monatin, Synthetic Communications, 24(22), 3197–3211 (1994).

Vleggaar, et al., Structure elucidation of monatin, a high-intensity sweetener isolated from the plant *Schlerochiton ilicifolius*, J. Chem. Soc. Perkin Trans., 3095–3098, (1992).

Primary Examiner—Joseph K. McKane
Assistant Examiner—John R. Dolan
Attorney, Agent, or Firm—Samuels, Gauthier & Stevens

[57] ABSTRACT

The high intensity natural sweetener monatin, an α-amino acid (1) is synthesized from readily available starting materials. Regiospecific ring opening of epoxide (9) with indole magnesium bromide yields lactone (11). Amination of (11) followed by saponification yields Monatin.

1 Claim, No Drawings

SYNTHESIS OF MONATIN-A HIGH INTENSITY NATURAL SWEETENER

BACKGROUND OF THE INVENTION

1. Field of the Invention

The invention relates to the synthesis of monatin.

2. Description of the Prior Art

The undesirable effects associated with the high consumption of sugar such as obesity and tooth decay have resulted in an intensive search for natural and synthetic substitutes. Although a number of non-nutritive sweeteners have been approved for public consumption, such as saccharin, cyclamate, acesulfame-K and aspartame, none fulfill the requirements of being safe, non-caloric, stable over a wide range of pHs and temperatures with a long shelf-life, and no aftertaste. The concern about long term effects of some of these products has resulted in the search for high intensity sweeteners from natural sources.

There are a number of naturally occurring compounds such as stevioside and glycyrrhizin that are used as sweeteners in Europe and Japan but none have been approved for use as such in the United States. Monatin is a high intensity, low-caloric sweetener that was isolated from the bark of the roots of *Schlerochiton ilicifolius*, found in the northern Transvaal region of South Africa. Its relative sweetness was determined to be 1400 times that of sucrose. It was found to create highly acceptable blend with other sweeteners such as aspartame in a variety of flavors with or without carbonation. The limited amounts in the dried bark (0.007%) has spurred efforts to develop synthetic methodologies to produce ample supplies of the sweetener and to prepare analogs to assess the relationship between structure and sweetness.

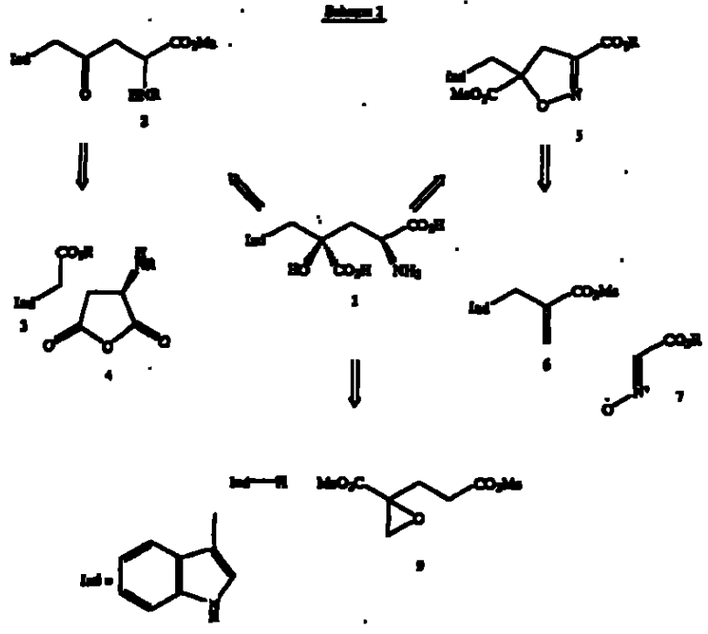
Structural and stereochemical analyses of monatin established it to be (2S,4S) 2-amino-4-carboxy-4-hydroxy-5-(3-indolyl)pentanoic acid 1. There are two reported syntheses of monatin that are retrosynthetically outlined in Scheme 1. the first deals with the preparation of a key keto ester 2 from indole acetic acid and an activated L-aspartate, Holmquist, C. W.; Olivier, *Synth. Commun.*, 1993, 23, 2511. The second is based upon a 2+3 cycloaddition reaction between indolyl acrylate 6 and furanocarbe oxide 7 to form isomonatins 5 from which monatin is obtained, Holmquist, C. W.; Rinehardbarger, K.; Olivier, *Synth. Commun.*, 1994, 24, 7025. Both are multistep routes and suffer from low yields in certain steps.

SUMMARY OF THE INVENTION

Broadly the invention comprises an economic, versatile and scalable synthesis of monatin that is adaptable for the preparation of chiral isomers and new analogs for taste evaluations.

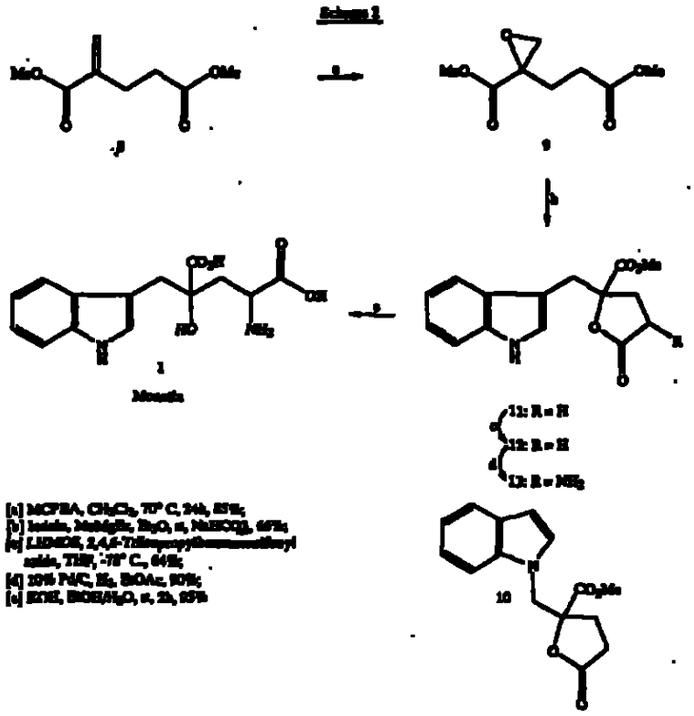
Referring to Scheme I, retrosynthetic analyses of 1 to indole, whose regiospecific condensation with an epoxy ester 9 generated the required substitution at C3.

Scheme 1



The synthetic pathway to monatin 1 is outlined in Scheme 2.

Scheme 1



[a] MCPBA, CH_2Cl_2, 70° C, 24h, 85%;
[b] Indole, Mg(Hg)Br, Et_2O, rt, $NaHCO_3$, 66%;
[c] LiHMDS, 2,4,6-Triisopropylbenzenesulfonyl azide, THF, -78° C., 64%;
[d] 10% Pd/C, H_2, EtOAc, 90%;
[e] EtNF, EtOH/H_2O, rt, 2h, 95%

Dimethyl-2-methylene glutarate 8 was obtained by BU_3P catalyzed dimerization of methyl acrylate, Rauhut, M. M.; Currier, H., U.S. Pat. No. 3,074,999, 1963; and Nagata, N.; Ogawa, M.; Naito, T., Japan Kokai 75868816, 1973. Oxidation of 8, with m-chloroperbenzoic acid furnished epoxide 9 in 85% isolated yield.

Regiospecific ring opening of epoxide 9, Ghosh, A.; Wang, W.; Freeman, J. P.; Althaus, J. S.; Von Voigtlander, P. F.; Scahill, T. A.; Mizsak, S. A.; Szmuszkovicz, J., *Tetrahedron*, 1991, 47, 8653, with indole magnesium bromide followed by in situ base catalyzed lactonization with $NaHCO_3$ furnished two products in 12 and 66% yield whose structures were assigned based on ^1H NMR analysis. The minor 10 was the result of N-alkylation having no NH signal and a proton doublet at C3 (δ 6.58). The major product proved to be the desired lactone 11 with an exchangeable indole proton at δ 9.25 and an AB quartet at 3.38 ppm for the methylene protons at C3.

The α-deprotonation of lactone 11 with lithium hexamethyldisilazide followed by azide transfer from the hindered electrophile 2,4,6-triisopropylbenzenesulfonyl azide, Evans, D. A.; Evrad, D. A.; Rychnovsky, S. D.; Fruh, T.; Whittingham, W. G.; DeVries, K. M., *Tetrahedron Lett.*, 1992, 33, 1189, (trisyl azide; prepared in quantitative yield from trisyl chloride) afforded azido lactone 12 as a 3:2 diastereomeric mixture. The most distinctive ^1H NMR feature of 12 is the difference in the chemical shifts of the azido protons (δ 3.36 & 4.40 ppm), both appearing as doublet of doublets. It is presumed that the more deshielded proton has cis stereochemistry to the methoxycarbonyl group.

Catalytic hydrogenation of azido lactone 12 gave amino lactone 13 in nearly quantitative yield as a 3:2 diastereomeric mixture. Base catalyzed hydrolysis of amino lactone 13 gave racemic monatin 1.

It is evident from this synthesis that chiral isomers of monatin can be obtained following the same route described in Scheme 2, by using an optically active epoxide 9. This can be obtained by AD-mix oxidation of olefin 8, Cha, J. K.; Kim, N. S., *Chem Rev.*, 1995, 95, 1761; and Kolb, H. C.; VanNieuwenhze, M. S.; Sharpless, K. B., *Chem Rev.*, 1994, 94, 2483, followed by Mitsunobu dehydration of the resultant diol, Abushanab, E.; Vemishetti, P.; Leiby, R. W.; Singh, H. K.; Mikkilineni, A. B.; Wu, D. C. J.; Saibaba, R.; Panzica, R. P., *J. Org. Chem.*, 1988, 53, 2598. The newly created chiral center can induce asymmetry in the conversion of 11 to 12 leading to optically pure monatin isomers.

DESCRIPTION OF THE PREFERRED EMBODIMENT

Experimental

General

All reactions were run in oven-dried glassware, sealed with a rubber septum, and stirred with a magnetic stirring bar under N_2. Unless otherwise noted, materials were obtained from commercial suppliers and were used without

purification. Ether was dried with $LiAlH_4$ and distilled under nitrogen. THF was distilled from sodium benzophenone ketyl just prior to use. Flash chromatography was performed on Fisher Scientific silica gel 230–400 mesh. Melting points are reported uncorrected.

Methyl 4,5-epoxy-4-methoxycarbonylpentanoate (9)

To an ice cold solution of the olefin 8 (3.44 g, 20 mmol) in CH_2Cl_2 (100 mL) was added 55% m-chloroperbenzoic acid (7.54 g, 24 mmol). After stirring the reaction mixture at room temperature for 24 h, it was diluted with ethyl acetate (300 mL), washed successively with saturated $NaHCO_3$ (150 mL), 10% $NaHSO_3$ (150 mL), saturated $NaHCO_3$ (150 mL), and brine, and dried ($MgSO_4$). The residue obtained after evaporation of ethyl acetate was purified by distillation to afford pure epoxide 9 (3.20 g, 85%); bp. 110–112° C./4 mm; ^1H NMR (300 MHz, $CDCl_3$ (δ 2.01–2.10 (m, 1H), 2.38–2.54 (m, 3H), 2.84 (d, 1H, J=5.8 Hz), 3.10 (d, 1H, J=5.8 Hz), 3.67 (s, H), 3.77 (s, H); ^{13}C NMR (75 MHz, $CDCl_3$) δ 26.55, 29.42, 51.72, 51.99, 52.62, 55.89, 170.36, 1792.96. Anal. Calcd for $C_8H_{12}O_5$: C, 51.06; H, 6.38. Found C, 51.05; H, 6.46.

4-(3-Indolylmethyl)-4-methoxycarbonyl-γ-butyrolactone (11)

To a well-stirred solution of indole (2.90 g, 24.8 mmol) in Et_2O (10 mL) under nitrogen was added dropwise a solution of methyl magnesium bromide (3.0 M, 8.3 mL, 25 mmol) at room temperature. The rate of addition was adjusted to maintain a gentle reflux. After completion of the addition, the resulting solution was stirred another 45 min at room temperature. A solution of epoxide 9 (4.70 g, 25 mmol) in Et_2O (10 mL) was then added dropwise, followed by 20 mL of Et_2O and left stirring overnight (12 h). The reaction was quenched slowly by the dropwise addition of a saturated solution of $NaHCO_3$ (30 mL), filtered and the white precipitate was washed repeatedly with EtOAc (4×50 mL). The combined organic layer was dried and evaporated under reduced pressure to leave a brown residue. This was purified by silica gel column chromatography eluting with EtOAc-hexanes (2:8) to afford pure 10 (0.80 g, 12%). A sample was crystallized from ethyl acetate/hexanes: mp. 100–102° C.; ^1H NMR (90 MHz, $CDCl_3$) δ 2.22–2.44 (m, 2H), 2.72–3.40 (m, 2H), 3.62 (AB_q, 2H, J=10.2 Hz), 3.81 (s, 3H), 6.58 (d, 1H, J=3.3 Hz), 7.16–7.74 (m, 4H), 8.36 (d, 1H J=8.4 Hz). Anal. Calcd for $C_{15}H_{15}NO_4$: C, 65.93; H, 5.50; N, 5.13. Found: C, 65.89; H, 5.46; N, 5.12. Further elution with EtOAc-hexanes (3:7) gave pure indole lactone 11 (4.42 g, 66%). Analytical sample was crystallized from acetonitrile to afford a white solid: mp. 168–170° C., ^1H NMR (300 MHz, CD_3CN) δ 2.19–2.49 (m, 4H), 3.38 (ABq, 2H, J=15.1 Hz), 3.69 (s, H), 7.04–7.16 (m, 3H), 7.40 (d, 1H, J=8.0 Hz), 7.59 (d, 1H, J=7.6 Hz); 9.25 (bra, 1H, D_2O Exchangeable); ^{13}C NMR (75 MHz, CD_3CN) δ 28.97, 31.40, 33.60, 53.41, 88.34, 109.28, 112.47, 120.03, 120.29, 122.75, 125.85, 129.13, 137.38, 173.23, 177.13. Anal. Calcd for $C_{15}H_{15}NO_4$: C, 65.93; H, 5.50; N, 5.13. Found: C, 66.10; H, 5.49; N, 5.22.

2-Azido-4-(3-indolylmethyl)-4-methoxycarbonyl-γ-butyrolactone (12)

To a solution of LHMDS (1.0 M, 24.2 mL, 24.2 mmol) in THF was added indole lactone 11 (3.00 g, 11 mmol) in freshly distilled THF (33 mL) at −78° C. The light yellow solution was stirred for 40 min, and triisopropylbenzenesulfonyl azide (3.40 g, 11 mmol) in 33 mL dry THF was added at −78° C. via syringe. The reaction mixture was stirred for 20 min before quenching with acetic acid (49.5 mmol, 2.83 mL). The reaction mixture was allowed to warm to room temperature over 30 min before being diluted with saturated NH_4Cl solution. This solution was stirred for 5 min and extracted with EtOAc (2×75 mL). The combined organic layers were dried ($MgSO_4$), and the solvent was removed under reduced pressure. The crude oil was flash chromatographed on silica gel (1:2, EtOAc-hexanes eluant) to afford 2.21 g (64%) of pure azide lactone 12 as a mixture of diastereomers (3:2; SS/RR:SR/RS). A sample was crystallized from ethyl acetate/hexanes: mp. 110–112° C., SS/RR racemate; ^1H NMR (300 MHz, $CDCl_3$) δ 2.19 (dd, 1H, J=14.0, 7.3 Hz), 2.63 (dd, 1H, J=14.0, 9.3 Hz), 3.34 (dd, 1H, J=9.0, 2.6 Hz), 3.39 (AB_q, 2H, J=15.3 Hz), 3.84 (s, 3H), 7.10 (d, J=2.4 Hz), 7.14–7.24 ((m, 2H), 7.37 (d, 1H, J=8.3 Hz), .7.63 (d, 1H, J=7.8 Hz), 8.32 (bra 1H, D_2O Exchangeable).

^{13}C NMR (75 MHz, $CDCl_3$) δ 32.57, 35.82, 53.24, 56.84, 85.16, 106.97, 111.64, 118.36, 120.37, 122.67, 125.04, 127.47, 135.99, 171.12, 172.14. SR/RS racemate: ^1H NMR (300 MHz, $CDCl_3$) δ 2.13 (dd, 1H, J=13.4, 110 Hz), 2.79 (dd, 1H, J=13.4, 8.9 Hz), 3.45 (Ab_q, 2H, J=15.3 Hz), 3.74 (s, 3H), 4.40 (dd, 1H, J=10.9, 9.0 Hz), 7.13 (d, 1H, J=2.3 Hz) 7.14–7.24 (m, 2H), 7.40 (d, 1H, J=8.0 Hz), 7.58 (d, 1H, J=7.5 Hz), 8.23 (bra, 1H, D_2O Exchangeable). Anal. Calcd for $C_{15}H_{14}N_4O_4$: C, 57.33; H, 4.46; N, 17.8. Found: C, 57.15; H, 4.62; N, 17.78.

2-Amino-4-(3-indolylmethyl)-4-methoxycarbonyl-γ-butyrolactone (13)

To a solution of the azide 12 (0.63 g, 2.0 mmol) in EtOAc (30 mL) was added 10% Pd/C (0.12 g). The flask was charged with H_2 (1 atm) and the solution was stirred at room temperature. After 2 h the catalyst was filtered off and the solvent was removed in vacuo. Purification of the residue by flash chromatography using EtOAc as eluent gave amino lactone 13 (0.52 g, 90%) as a mixture of diastereomers (3:2; SS/RR:SR/RS). A sample was crystallized from ethyl acetate/hexanes to afford a white solid: mp. 188–189° C.; SS/RR racemate: ^1H NMR (500 MHz, CD_3CN) δ 1.54 (bra, 2H D_2O Exchangeable), 2.16 (dd, 1H, J=13.2, 9.3 Hz), 2.71 (dd, 1H, J=13.2, 9.2 Hz), 3.34 (AB_q, 2H, J=15.0 Hz), 3.66 (s, 3H), 3.67–3.70 (m, 1H), 7.05–7.15 (m, 2H), 7.39 (d, 1H, J=7.9 Hz), 7.56 (d, 1H, J=8.0 HZ), 9.24 (bra, 1H, D_2O Exchangeable). SR/RS racemate: δ 1.54 (bra. 2H, D_2O Exchangeable), 2.06 (t, 1H, J=12.3 Hz), 2.79 (dd, 1H, J=12.9, 9.6 Hz), 3.17 (t, 1H, J=9.2 Hz), 3.38 (ABq, 2H, J=15.1 Hz), 3.71 (s, 3H), 7.05–7.15 (m, 2H), 7.40 (d, 1H, J=8.0 Hz), 7.59 (d, 1H, J=7.9 Hz), 9.28 (bra, 1H, D_2O Exchangeable). Anal. Calcd for $C_{15}H_{16}N_2O_4$: C, 62.50; H, 5.56; N, 9.72. Found: C, 62.35; H, 5.38; N, 9.62.

2-Amino-4-carboxy-4-hydroxy-5-(3-indolyl)pentanoicacid

A solution of amino ester 13 (0.25 g, 0.87 mmol) and potassium hydroxide (0.11 g, 1.96 mmol) in 5 mL 80% ethanol was stirred for 2 h at room temperature. This solution was passed through an ion-exchange resin (Biorad, AG 50W-X8, 3 mL) and the resin was rinsed with a 15 mL 80% ethanol solution. Removal of the solvent in vacuo afforded a pale yellow crystalline product monazin 1 (0.24 g, 95%) as racemates. SS/RR racemate: ^1H NMR (300 MHz, D_2O) δ 1.90 (dd, 1H, J=13.5, 8.7 Hz), 2.23–2.38 (m, 1H), 3.11 (AB_q, 2H, J=14.6 Hz), 3.10–3.17 (m, 1H), 7.12–7.22 (m, 2H), 7.24 (s, 1H), 7.46–7.49 (m, 1H), 7.70 (d, 1H J=7.7 Hz). SR/RS racemate: ^1H NMR (300 MHz, D_2O) δ 2.26–2.39 (m, 1H), 2.79 (dd, 1H, J=13.2, 7.4 Hz), 3.19 (AB_q, 2H, J=13.5 Hz), 4.11 (dd, 1H, J=9.1, 5.0 Hz), 7.12–7.22 (m, 2H), 7.30 (s, 1H), 7.46–7.49 (m, 1H), 7.72 (d, 1H, J=7.8 Hz).

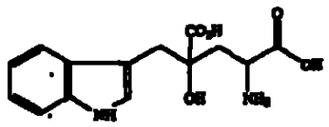
The foregoing description has been limited to a specific embodiment of the invention. It will be apparent, however, that variations and modifications can be made to the

invention, with the attainment of some or all of the advantages of the invention. Therefore, it is the object of the appended claims to cover all such variations and modifications as come within the true spirit and scope of the invention.

Having described our invention, what we now claim is:

1. A method for the synthesis of

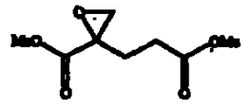
(1)



which comprises:

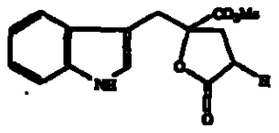
opening the ring of the active epoxy ester

(9)



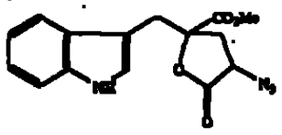
by reacting with indole magnesium bromide followed by in situ base catalyzed lactonization to produce

(11)



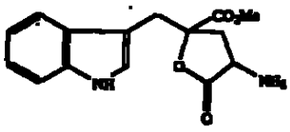
effecting α-deprotonation of lactone (11) with lithium hexamethyldisilazide followed by oxide transfer from the hindered electrophile 2,4,6-triisopropylbenzenesulfonyl azide to produce

(12)



catalytically hydrogenating (12) to produce

(13)



and

effecting base catalyzed hydrolysis of (13) to produce (1).

* * * * *

SYNTHETIC COMMUNICATIONS, 24(22), 3197-3211 (1994)

A SIMPLE CYCLOADDITION APPROACH TO A RACEMATE OF THE NATURAL SWEETENER MONATIN

Cedric W. Holzapfel*

Department of Chemistry, Rand Afrikaans University, Johannesburg, South Africa,

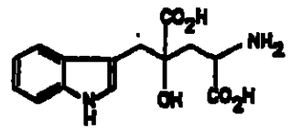
and Karl Blechschmidt and Johan Olivier

MATTEK, CSIR, P.O. Box 395, Pretoria, South Africa.

Abstract: The reaction of methyl 2-(3-indolylmethyl)-acrylate 2 with carbethoxyformonitrile oxide 3, followed by saponification and reduction, gives mainly racemic monatin 1. Preliminary attempts to adapt the 2+3 cycloaddition to a new chiral synthesis of monatin, is described.

* To whom correspondence should be addressed

We have recently described an efficient route to a key intermediate in the synthesis of the natural sweetener monatin 1, starting from L-aspartic acid[2]. In addition to a chiral synthesis of monatin, we required a simple method for the preparation of the racemate of monatin for the purpose of a structure/sweetness relationship study. We now report the synthesis of racemic monatin using a (2+3) cycloaddition as the key step.

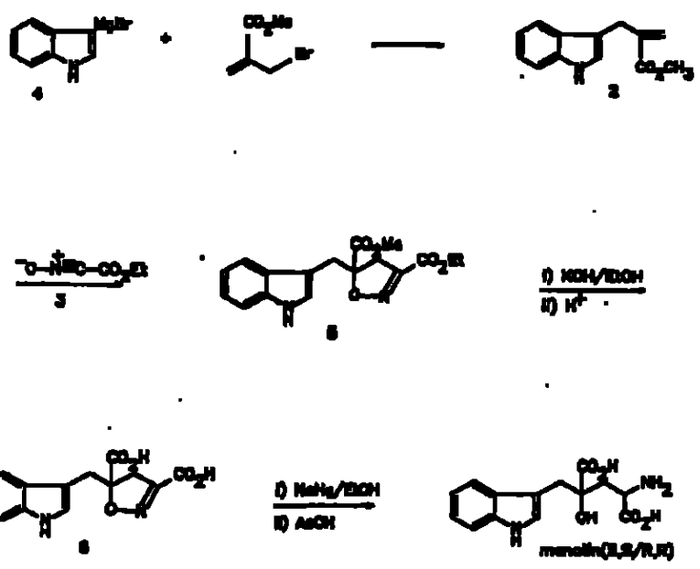


1

As starting materials for the synthesis of racemic 1 we selected the acrylate 2 and carbethoxyformonitrile oxide 3 (CEFNO)[3]. Methyl 2-(3-indolylmethyl)-acrylate 2 was prepared by the coupling of the indole Grignard 4 with methyl bromomethacrylate[4]. The nitrile oxide 3a was generated *in situ* from the corresponding chloro-oxime prepared from the glycine ethyl ester[5].

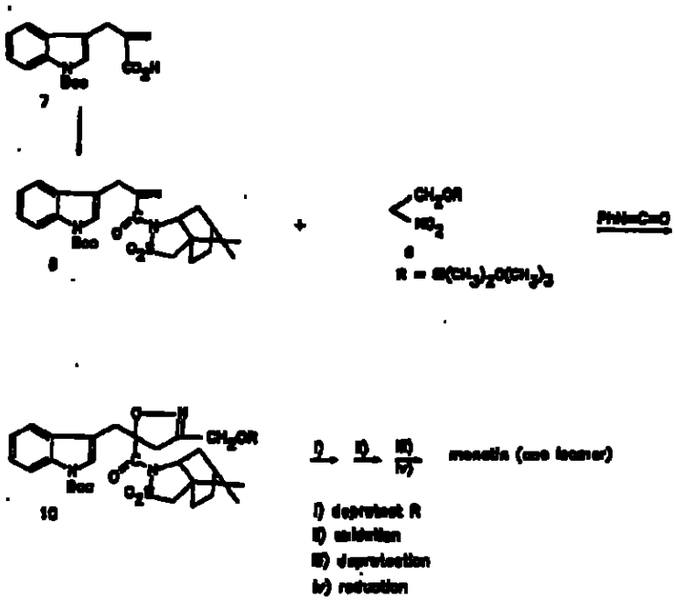
Reaction of 2 with CEFNO proceeded smoothly in the presence of triethylamine to afford the diester 5 in excellent yield. Subsequent saponification of diester 5 yielded the diacid 6. Reduction of the isoxazoline

ring system of 6 with sodium amalgam in ethanol afforded a mixture of the four stereo-isomers of monatin. Treatment of an aqueous solution of this mixture of isomers with acetic acid resulted in the selective precipitation the S,S/R,R isomeric mixture of monatin isomers (scheme 1). This S,S/R,R mixture of monatin isomers showed no apparent loss of sweetness when compared to the S,S natural monatin isomer.



Scheme 1

We are presently engaged in attempts to develop this approach into a second route for the synthesis of chiral monatin. In this regard the N-t-Boc 2-(3-indolylmethyl)-acrylic acid 7 was transformed into the chiral amide 8 by reaction with Oppolzer's sultam[7]. Compound 8 was reacted with the nitrile oxide 3. None of the expected products of (2+3) cycloaddition could be detected. Compound 8 was then reacted with the nitrile oxide generated



Scheme 2

in situ from O-allyl protected nitroethanol 9. This reaction proceeded with excellent chiral induction to yield a single isoxazoline diastereomer 10 in low yield (scheme 2). This low yield is probably due to steric hindrance in the cycloaddition transition state due to the presence of the bulky t-butyl dimethylsilyl O-protecting group. This rationalization is supported by the fact that reaction of 8 with the nitrile oxide generated *in situ* from nitroethane proceeded at a markedly higher rate and furnished the corresponding cyclo-adduct in good yield[7]. We envisage completion of the monatin synthesis to involve similar saponification and reduction steps as for the non-chiral synthesis as well as the transformation of the protected alcohol into the corresponding acid.

EXPERIMENTAL

General. ^1H and ^{13}C n.m.r. spectra were recorded on a Bruker WM-500 or AM-300 spectrometer. Mass spectra were taken on a Finnigan MAT 90 double focussing spectrometer. Optical rotations were recorded on a Perkin-Elmer 241 polarimeter. M.p.'s were recorded on a Kofler hot-stage and are uncorrected. Oppolzer's sultam was obtained from Aldrich in the D-form.

Preparation of methyl 2-(3-indolylmethyl)-acrylate (2)

Ethyl magnesium bromide was prepared by adding ethyl bromide (1.3 g, 12 mmol) in 2 ml dry THF to magnesium turnings (260 mg, 12 mmol) in 4 ml dry THF and stirring the mixture for 1.5 hours at room temperature. A solution of indole (1.17 g, 10 mmol) in 2 ml dry THF was added slowly by syringe and stirring continued for 20 minutes at room temperature. Methyl bromomethacrylate (1.90 g, 11 mmol) in 1.5 ml dry THF was added dropwise at -20°C. The reaction mixture was stirred at room temperature for 30 minutes and quenched with water.

The THF was removed *in vacuo* and the aqueous residue extracted 3 times with ethyl acetate, the ethyl acetate layers combined and dried over anhydrous magnesium sulphate, filtered and the solvent removed *in vacuo* to afford a yellow syrup. Column chromatography (toluene: ethyl acetate [19:1]) gave unreacted indole (230 mg, 19.6%) followed by the indole ester 2 (1.36 g, 63.2%) as pale brown crystals, m.p. 46°-47°C; ^{1}H n.m.r. ($CDCl_3$, δ): 3.75 (3H, s), 3.77 (2H, s), 5.48 (1H, m), 6.18 (1H, m), 7.01-7.99 (5H, m, indole); ^{13}C n.m.r. ($CDCl_3$, δ): 27.56 (Ind-$\underline{C}H_2$), 51.79 (O-$\underline{C}H_3$), 111.12 (C=$\underline{C}$), 112.66 ($\underline{C}$=C), 118.98-139.46 (indole), 167.85 ($\underline{C}$O); M^+ 215 (Expected for $C_{13}H_{13}NO_2$: M, 215); Anal. Calcd for

$C_{20}H_{20}NO_2$: C, 72.59; H, 6.08; N, 6.50. Found: C, 73.12; H, 6.19; N, 6.58.

Preparation of isoxazoline diester (5)

To a solution of methyl 2-(3-indolylmethyl)-acrylate 2 (65 mg, 0.3 mmol) and triethylamine (101 mg, 1.0 mmol) stirring in 1 ml dry chloroform at 25°C was dropwise added by syringe ethyl chloro(hydroxylimino)acetate 3 (91 mg, 0.6 mmol) over a period of 4 hours. The reaction mixture was stirred overnight at room temperature. The reaction mixture was washed with water (10 ml), the organic phase dried over anhydrous magnesium sulphate, filtered and concentrated *in vacuo* to afford a yellow syrup. Column chromatography (toluene: ethyl acetate [17:3]) afforded the pure cycloaddition product 5 as an oil that solidified and was recrystallised from ethyl acetate/toluene to afford 96 mg (0.29 mmol, 96.8%), m.p. 84°-85°C; ^{1}H n.m.r. ($CDCl_3$, δ): 1.27 (3H, t), 3.20 (1H, d, J=20.3 Hz,), 3.41 (1H, d, J=25.3 Hz, Ind-CH_2), 3.49 (1H, d, J=25.3 Hz, Ind-CH_2), 3.59 (1H, d, J=20.3 Hz,), 3.75 (3H, s, O-CH_3), 4.24 (2H, q), 7.09-8.26 (5H, m, indole); ^{13}C n.m.r. ($CDCl_3$, δ): 13.94 (CH_3-CH_2), 31.32 (Ind-CH_2), 40.48 (CH_2), 59.04 (O-CH_2), 62.11 (CH_2-CH_3), 76.99 (CH_2-C), 91.87a (C-4), 107.68-135.81 (indole), 151.58 (CO),

171.30 (CO); M^+ 330 (Expected for $C_{17}H_{18}N_2O_5$: M, 330); Calculated accurate mass for $C_{17}H_{18}N_2O_5$: 330.1215, Found: 330.1215; Anal. Calcd for $C_{17}H_{18}N_2O_5$: C, 61.81; H, 5.49; N, 8.48. Found: C, 61.93; H, 5.40; N, 8.53.

Preparation of racemic monatin

A solution of isoxazoline diester *5* (218 mg, 0.66 mmol) and potassium hydroxide (99 mg, 1.50 mmol) in 5 ml 80% ethanol was stirred for one hour at room temperature. This solution was passed through an ion-exchange resin (Biorad, AG 50W-X8, 3ml) and the resin rinsed with a 15 ml 80% ethanol solution. Removal of the solvent *in vacuo* afforded the monatin diacid (164 mg, 0.56 mmol, 86.2%) as a yellow, crystalline product *6*. Because of the light sensitivity of the diacid it was directly converted to the final product.

A solution of diacid *6* (164 mg, 0.56 mmol) in 5 ml 80% ethanol was treated with sodium amalgam (4.09 g, 8.89 mmol sodium) at room temperature for 5 hours. The reaction mixture was separated from the mercury by filtration and passed through cation exchange resin (Biorad, AG 50W-X8, 20 ml) and washed with 15 ml 80% ethanol. The product was removed from the resin by eluting with aqueous ammonia (80 ml, 5%) to

afford pure monatin as a mixture of four stereoisomers (156 mg, 0.52 mmol, 92.8%). The mixture of four stereoisomers (150 mg, 0.51 mmol) was dissolved in 1 ml water and 0.1 ml acetic acid added. A product precipitated and the procedure repeated twice to produce a product (S,S/R,R racemic mixture) identical to natural monatin by n.m.r. (80 mg), m.p. 212°-214°C; ^{1}H n.m.r. (D_2O, δ): 1.96 (1H, dd, J=15.2 Hz, J=11.7 Hz, CH_2-CH), 2.58 (1H, dd, J=15.2 Hz, J=1.9 Hz, CH_2-CH), 2.99 (1H, d, J=14.5 Hz, Ind-CH_2), 3.19 (1H, d, J=14.5 Hz, Ind-CH_2), 3.55 (1H, dd, J=11.7 Hz, J=1.9 Hz, CH_2-CH), 7.04-7.65 (4H, Indole); ^{13}C n.m.r. (D_2O, δ): 34.50 (CH_2-CH), 37.34 (Ind-CH_2), 52.89 (CH_2-CH), 79.40 (CH_2-C), 108.28-135.03 (Indole), 173.37 (CO), 179.19 (CO), M$^+$ 292 (Expected for $C_{14}H_{16}N_2O_5$: M, 292).

Preparation of 2-(3-Indolylmethyl)-acrylic acid

To a solution of methyl 2-(3-indolylmethyl)-acrylate 2 (230 mg, 1.07 mmol) dissolved in 80% ethanol (4 ml) was added potassium hydroxide (150 mg, 2.69 mmol) and the solution stirred for 5 hours at room temperature. The reaction mixture was passed through ion-exchange resin (Biorad 50W2X8 [200-400 mesh], 2 ml) to afford the pure 2-(3-indolylmethyl)-acrylic acid after evaporation of the solvent. The product was recrystallized from ethyl acetate/hexane to afford the product as slightly

335

red crystals, Anal. Calcd for $C_{20}H_{11}NO_3$: C, 71.62; H, 5.50; N, 6.96. Found: C, 71.24; H, 5.52; N, 6.91.

Preparation of N-t-Butyloxycarbonyl 2-(3-indolylmethyl)-acrylic acid (7)

Di-t-butyldicarbonate (352 mg, 1.61 mmol) dissolved in 1 ml dry acetonitrile was added by syringe to a solution of methyl 2-(3-indolylmethyl)-acrylic acid (217 mg, 1.079 mmol) and 20 mg DMAP stirring in 5 ml dry acetonitrile at room temperature. After 2 hours at room temperature the solvent was evaporated and the product purified by column chromatography (toluene: ethyl acetate: acetic acid [100:50:1]) to afford 370 mg (1.22 mmol, 76.3%) as a solid 7, m.p. 97°-99°C; ^{1}H n.m.r. ($CDCl_3$, δ): 1.67 (9H, s, t-Boc), 3.71 (2H, s), 5.83 (1H, m), 6.40 (1H, m), 7.19-8.11 (5H, indole); ^{13}C n.m.r. ($CDCl_3$, δ): 26.82 (Ind-$\underline{C}H_2$), 28.21 (C[$\underline{C}H_3$]$_3$), 83.57 (C[CH_3]$_3$), 115.29 (C=$\underline{C}$), 117.33 ($\underline{C}$=C), 119.20-137.78 (indole), 149.73 (N-$\underline{C}$O), 172.09 ($\underline{C}$O), Anal. Calcd for $C_{17}H_{19}NO_4$: C, 67.77; H, 6.31; N, 4.65 Found C, 67.75; H, 6.35; N, 4.64.

Preparation of camphor sultam acrylamide (8)

Thionyl chloride (0.21 ml, 2.92 mmol) was dissolved in 1 ml dry ether and added at 0°C by syringe to a solution of N-t-butyloxycarbonyl 2-(3-indolylmethyl)-acrylic acid 7 (86 mg, 0.29 mmol) dissolved in 2 ml dry

ether. The solution was refluxed for 10 hours and the solvent removed *in vacuo*.

D-Camphor sultam (60 mg, 0.27 mmol) dissolved in 1 ml dry toluene was added by syringe to sodium hydride (14 mg, 0.33 mmol, 55% in oil) under nitrogen and the reaction mixture stirred for 30 minutes at room temperature. The acid chloride prepared above was dissolved in 2 ml dry toluene and added by syringe to the sultam anion reaction mixture at room temperature. The reaction mixture was stirred for a further 90 minutes at room temperature.

Ice-water (10 ml) was added to the reaction mixture and the product extracted with a further 20 ml toluene. The organic phase was dried over anhydrous magnesium sulphate, filtered and the solvent removed *in vacuo*. The product was isolated by column chromatography (ethyl acetate: toluene [1:20]) and the product isolated as colourless crystals 8 (110 mg, 0.22 mmol, 75.8%), m.p. 164°-166°C; ^{1}H n.m.r. ($CDCl_3$, δ): 0.98 (3H, s, sultam-CH_3), 1.19 (3H, s, sultam-CH_3), 1.40 (2H, m, sultam), 1.64 (9H, s, t-Boc), 1.85-2.10 (5H, m, sultam), 3.39 (1H, d, J=13.6 Hz, Ind-CH_2), 3.50 (1H, d, J=13.6 Hz, Ind-CH_2), 3.64 (1H, d, J=17.3 Hz, 8-CH_2), 3.80 (1H, d, J=17.3 Hz, 8-CH_2), 4.06 (1H, m, sultam), 5.41 (1H, m, C=CH_2),

5.79 (1H, m, C=C$\underline{H}_2$), 7.13-8.12 (5H, m, indole); ^{13}C n.m.r. ($CDCl_3$, δ): 19.85 (sultam-$\underline{C}H_2$), 21.25 (sultam-$\underline{C}H_2$), 26.45 (sultam-$\underline{C}H_2$), 27.56 (sultam-$\underline{C}H_2$), 28.20 (C[$\underline{C}H_3$]$_2$), 33.23 (sultam-$\underline{C}H$), 38.36 (Ind-$\underline{C}H_2$), 45.21 (sultam-$\underline{C}H_2$), 47.69 (sultam-$\underline{C}$[CH_3]$_2$), 48.01 (sultam-$\underline{C}$), 59.57 (sultam-$\underline{C}H$), 65.56 (S-$\underline{C}H_2$), 83.58 (C[CH_3]$_2$), 115.06 (C=$\underline{C}$), 116.33 ($\underline{C}$=C), 119.88-125.27 (Indole), 143.23 (N-$\underline{C}O_2$), 170.87 (N-$\underline{C}O$); M^+ 498 (Expected for $C_{23}H_{34}N_2O_5S$: M, 498); Calculated accurate mass for $C_{23}H_{34}N_2O_5S$: 498.2188, Found: 498.2187; Anal. Calcd for $C_{23}H_{34}N_2O_5S$: C, 65.03; H, 6.87; N, 5.61. Found: C, 65.15; H, 7.63; N, 5.44.

Preparation of silyl-protected nitroethanol (9)

Nitroethanol (0.1 g, 0.93 mmol, 85% solution), imidazole (0.16 mg, 2.33 mmol) and t-butyl dimethylsilylchloride (0.17 mg, 1.1 mmol) were dissolved in 0.5 ml dry DMF and heated to 35°C for 10 hours under nitrogen. The product was extracted with ethyl acetate, washed with water, the organic phase dried over anhydrous magnesium sulphate, filtered and the organic solvent removed *in vacuo*. Distillation under reduced pressure removed the excess t-butyl dimethylsilylchloride as well as the starting material nitroethanol. Column chromatography (ethyl acetate: hexane [1:2]) removed the imidazole and afforded the product 9 (0.15 g, 0.73 mmol,

78.6%) as a colourless liquid, ^{1}H n.m.r. ($CDCl_3$, δ): 0.04 (6H, s, $2xCH_3$), 0.84 (9H, s, t-butyl), 4.11 (2H, t), 4.43 (2H, t); ^{13}C n.m.r. ($CDCl_3$, δ): -5.6q ($2xCH_3$), 18.15s ($C[CH_3]$), 25.66q (t-butyl), 50.54t, 77.64t.

Coupling of protected nitroethanol with camphor sultam acrylamide (8)

Indole sultam acrylate 8 (150 mg, 0.32 mmol) and phenyl isocyanate (0.08 ml, 0.69 mmol) were dissolved in 2 ml dry benzene and stirred under nitrogen at room temperature. Silyl protected nitroethane 9 (70 mg, 0.34 mmol) and triethylamine (1 drop) were dissolved in 2 ml dry benzene and added over a period of 3 hours to the reaction mixture at room temperature.

The reaction mixture was stirred for 36 hours at room temperature and the solvent removed *in vacuo*. The product was isolated by column chromatography (ethyl acetate: toluene [1:3]) to afford the product 10 30 mg (0.04 mmol, 13.6%) as a colourless oil, $[\alpha]^{25}_D$ -3.6 (c, 0.54); ^{1}H N.m.r. ($CDCl_3$) δ -0.08 (s, 3H, Si-CH_3), -0.03 (s, 3H, Si-CH_3), 0.78 (s, 9H, Si-$C[CH_3]$), 0.80-1.91 (m, sultam), 1.62 (s, 9H, O-$C[CH_3]$), 3.11 (d, 1H; J=17.6 Hz, CH_2-isoxazoline ring), 3.16 (d, 1H, J=14.4 Hz, S-CH_2), 3.46 (d, 1H, J=13.5 Hz, Ind-CH_2), 3.51 (d, 1H, J=13.5 Hz, Ind-CH_2), 3.52 (d, 1H, J=14.4 Hz, S-CH_2), 3.56 (d, 1H, J=17.6 Hz, CH_2-isoxazoline ring),

3.90 (m, 1H, sultam), 4.25 (d, 1H, J=13.1 Hz, O-CH$_a$), 4.32 (d, 1H, J=13.1 Hz, O-CH$_a$), 7.17-8.07 (m, 5H, indole); ^{13}C N.m.r. ($CDCl_3$) δ -5.52 (Si-CH$_3$), -5.52 (Si-CH$_3$), 1.02 (Si-C[CH$_3$]$_3$), 19.85 (sultam-CH$_2$), 20.99 (sultam-CH$_2$), 25.68 (Si-C[CH$_3$]$_3$), 26.26 (sultam-CH$_2$), 28.23 (O-C[CH$_3$]$_3$), 33.21 (sultam-CH), 39.41 (sultam-CH$_2$), 38.93 (ind-CH$_2$), 43.72 (isoxazoline-CH$_2$), 44.94 (sultam-CH$_2$), 47.54 (sultam-C[CH$_3$]$_3$), 48.20 (sultam), 54.26 (sultam-CH), 58.28 (O-CH$_2$), 67.22 (S-CH$_2$), 77.25 (O-C), 89.47 (C[CH$_3$]$_3$), 89.90 (C=N), 113.56-135.55 (indole), 159.27 (N-CO$_2$), 173.37 (N-CO).

REFERENCES AND NOTES

1. This work was done as part of a Phd project at the Rand Afrikaans University.
2. Paulsen, L., *Gazz. Chim. Ital.*, 1939, 69, 322.
3. Holzapfel, C.W. and Olivier, J., *Synth. Comm.*, 1989, 19, 2511.
4. Drewes, S.E.; Loizou, G.; Roos, G.H.P., *Syn. Comm.*, 1987, 17, 291.
5. Skinner, G.S., *JACS*, 1924, 46, 731.
6. Oppolzer, W., Wills, M., Starkemann, C. and Bernardinelli, G.,

Tet. Lett., 1990, 31, 4117.

7. Holmquist, C.W., Olivier, J., unpublished results.

(Received in the UK 19 April 1994)

SYNTHETIC COMMUNICATIONS, 23(18), 2511-2526 (1993)

THE SYNTHESIS OF A γ-KETO-α-AMINO ACID, A KEY INTERMEDIATE IN THE SYNTHESIS OF MONATIN, A NEW NATURAL SWEETENER

Cedric W. Holzapfel*

Department of Chemistry, Rand Afrikaans University, Johannesburg, South Africa,

and Johan Olivier*

MATIEK, CSIR, P.O. Box 395, Pretoria, South Africa.

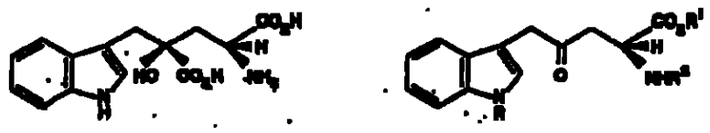
Abstract: An efficient method was developed for the synthesis of the keto-amino acid 2, a key intermediate in the synthesis of the novel sweet compound, monatin 1. Preparation of 2 entails coupling of a suitably protected indole acetate anion to an aspartic acid derivative.

Monatin, (Indol-3-yl)-2-amino-4-carboxy-4-hydroxypentanoic acid 1, is a sweet tasting α-amino acid isolated by ion chromatography from the roots of the plant *Schlerochiton ilicifolius*[2]. Monatin was evaluated by a taste panel

* To whom correspondence should be addressed

and the relative sweetness established as 1400 and 1200 times that of a 5 and 10% (w/v) sucrose solution, respectively[3]. As a sweetener monatin performs well, the pronounced body giving flavour enhancement, in common with other protein-based sweeteners. Monatin creates highly acceptable blends with other sweeteners such as aspartame in a variety of flavours and with or without carbonation.

Structural and stereochemical analysis established the structure of monatin as the novel γ-hydroxy-γ-indolyl glutamic acid 1, the published absolute stereochemistry[3] being 2S,4S. The preparation of γ-keto-α-amino acids of high optical purity have been widely explored[3,4]. Baldwin's recent



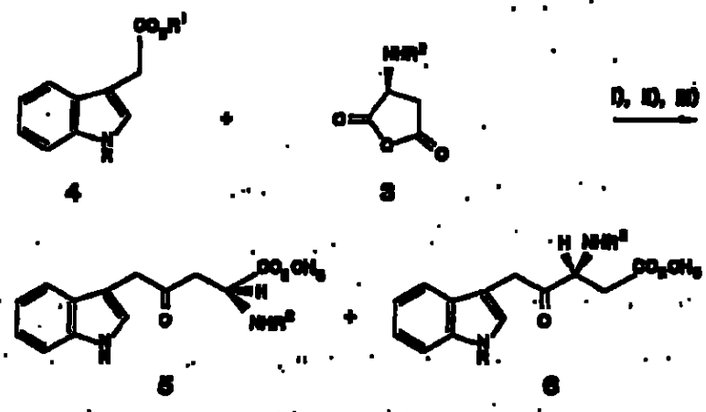
1 2

FIG 1

publication[5] of a new and efficient synthesis of γ-keto-α-amino acids from appropriate β-lactams prompts us to describe our own approach to these compounds.

Our approach to the synthesis of monatin involves the intermediary of the γ-keto-α-amino acid derivative of type 2. We now describe the efficient

synthesis of such compounds from L-aspartic acid. Ketone intermediate 2 was prepared by coupling of a suitably protected L-aspartic anhydride 3[6] to the N-protected indolyl acetic acid ester 4. Thus, the LDA generated anion of 7[7] was reacted with anhydrides 3, to afford coupling products which were subjected to hydrogenolysis in the presence of Pd/charcoal (affecting removal of the benzyloxycarbonyl group) followed by treatment with diazomethane to furnish the isomeric esters 5 and 6 (scheme 1). The results summarized in table 1 shows that selectivity of attack of the anion on the two carbonyls of the



4 3

5 6

	4a	4b	3a	3b	3c	5a/6a	5b/6b	6a
R	t-Boc	benzyl	-	-	-	t-Boc	benzyl	t-Boc
R^1	benzyl	t-butyl	-	-	-	-	-	-
R^2	-	-	t-Boc	Cbz	acetyl	t-Boc	Cbz	acetyl

I) LDA , -78°C , THF ; II) H_2/Pd charcoal (5a, 6a, 6c) ; p-toluenesulfonic acid , benzene (5b , 6b) ; III) CH_2N_2

scheme 1

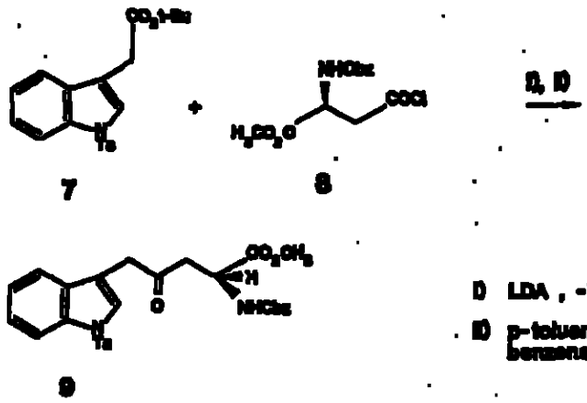
Table. The effect of anhydride protecting groups on the ratio of compounds 5 and 6.

Protecting group	Ratio (5/6)*	Yield (%)
t-Boc	50:50	70
Cbz	35:65	60
Acetyl	<5:>95	62

* Determined by chromatographic isolation.

anhydride is influenced by the nature of the protecting group (viz. N-acetyl, N-Cbz, N-t-Boc) on the anhydride. All three protecting groups have an electron withdrawing effect on the amine, increasing the electrophilicity of the carbonyl group and therefore its susceptibility to attack by the anion. This is exemplified by the N-acetyl protecting group where the ratio of 5/6 is <5:>95 (5c was not fully characterised). However, the bulkier N-t-Boc protecting group directs more of the attack of the anion to the less hindered side of the anhydride resulting in a ratio of 5/6 of 50:50. Compound 5a was prepared in an overall yield of 35% from the N-t-Boc anhydride.

A more efficient synthesis of 2 involved coupling of the anion of t-butyl 1-p-toluenesulfonylindole-3-acetate 7 to the acid chloride generated from α-methyl N-carbobenzoxy-L-aspartate 8[6]. p-Toluenesulfonic acid catalysed deprotection of the t-butyl ester was followed by spontaneous decarboxylation of the acid to afford the required γ-keto-α-amino acid 9 in good yield (scheme 2).



7 8 9

scheme 2

The formation of the γ-keto-α-amino acid derivative may proceed via the *in situ* formation of the β-lactam. No evidence for the intermediacy of a β-lactam

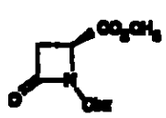


FIG 2

(FIG 2) could be obtained. Direct coupling to the acid chloride or reaction via a ketene intermediate cannot be excluded. In this regard we found[9] that the reaction of the anion of 7 with protected β-lactams did furnish the γ-keto-α-amino acid as expected, but in considerable lower yield than the corresponding reaction with the acid chloride. Nevertheless these results indicate Baldwin's γ-keto-α-amino acid synthesis can be applied to nucleophiles other than sulfone stabilised carbon nucleophiles.

The completion of the monatin synthesis requires the addition of a carboxylic acid group to the ketone moiety. The conversion of ketone intermediate 2 to the corresponding protected cyanohydrin and subsequent hydrolysis to monatin will be described in a following paper.

EXPERIMENTAL

General. ^1H and ^{13}C N.m.r. spectra were recorded on a Bruker WM-500 or AM-300 spectrometer. IR spectra were recorded on a Perkin-Elmer 883 spectrometer. Mass spectra were taken on a Finnigan MAT 90 double focussing mass spectrometer. Optical rotations refer to solutions in chloroform and were recorded on a Perkin-Elmer 241 polarimeter. M.p.'s were recorded

on a Kofler hot-stage and are uncorrected. Merck silica gel 60 was used for column chromatography.

Synthesis of benzyl 1-t-butyloxycarbonylindole-3-acetate (4a)

A solution of benzyl indole-3-acetate[10] (2.0 g, 7.5 mmol), di-t-butyl dicarbonate (1.7 g, 8.3 mmol) and N,N-dimethylaminopyridine (DMAP) (10 mg) was stirred at room temperature in 10 ml dry acetonitrile for 3 hours. The solvent was removed *in vacuo* and the product extracted with diethyl ether, washed with dilute hydrochloric acid followed by a saturated sodium bicarbonate solution. The ether layer was dried over anhydrous magnesium sulphate and the solvent removed *in vacuo* to afford 4a (2.8 g, 100%) as a slightly yellow oil: ν_{max} 2950-3050, 1750-1715; ^{1}H N.m.r. ($CDCl_3$) δ 1.67 (s, 9H, t-Boc), 3.73 (s, 2H, Ind-CH_2), 5.15 (s, 2H, O-CH_2), 6.85-8.20 (m, 9H, aromatic).

Synthesis of t-butyl 1-benzylindole-3-acetate (4b)

A solution of 1-benzylindole-3-acetic acid[11] (10.0 g, 37.7 mmol), t-butanol (10.6 ml, 0.11 mol) and DMAP (0.5 g) dissolved in 100 ml dry dichloromethane, was stirred at 0°C. Dicyclohexylcarbodiimide (8.6 g, 41.6 mmol) was added over a period of 5 minutes from a solid addition funnel, the ice bath was removed and the reaction mixture stirred for 3 hours at room

temperature. The precipitated dicyclohexylurea was removed by filtration and the filtrate washed with dilute hydrochloric acid and saturated sodium bicarbonate solution. The organic layer was dried over anhydrous magnesium sulphate and the solvent removed *in vacuo*. The product was isolated by column chromatography (ethyl acetate-hexane) and crystallized from ethyl acetate-hexane to afford 4b (4.3 g) as yellow crystals, m.p. 59°-60°C; ^{1}H N.m.r. ($CDCl_3$) δ 1.44 (s, 9H, t-Butyl), 3.67 (s, 2H, Ind-CH_2), 5.27 (s, 2H, N-CH_2), 7.09-7.62 (m, 9H, aromatic); ^{13}C N.m.r. ($CDCl_3$) δ 28.07 (t-butyl), 32.61 ($\underline{C}H_2$), 49.96 ($\underline{C}H_2$), 80.57 ($\underline{C}(CH_3)_3$), 108.32-137.57 (aromatic), 171.34 ($\underline{C}O$); Calcd for $C_{22}H_{23}NO_5$: C, 78.50; H, 7.16; N, 4.36. Found C, 78.03; H, 7.89; N, 4.62.

Synthesis of keto-amino acids (5a-b, 6a-c). General procedure

A solution of diisopropylamine (1.1 mmol, 0.15 ml) in 5 ml dry THF was stirred under nitrogen at -78°C. n-Butyllithium (1.1 mmol, of a 1.5 N solution in hexane) was added by syringe and the solution stirred for 10 minutes at -78°C. LDA formation was allowed to proceed for 30 minutes at room temperature. The LDA solution was cooled to -78°C and the protected indole acetate (1 mmol), dissolved in 2 ml dry THF, added by syringe. The solution was stirred for 30 minutes at -78°C followed by addition by syringe of the corresponding protected anhydride (0.5 mmol) in 2 ml dry THF. After stirring of the solution at -78°C for 2 hours, 1N HCl was added and the pH

adjusted to 3. THF was removed *in vacuo* and the aqueous phase extracted with ethyl acetate, the combined organic layers dried over anhydrous magnesium sulphate and the solvent removed *in vacuo*. The product was isolated by column chromatography, dissolved in 15 ml methanol and 25 mg palladium/charcoal added. The mixture was hydrogenated for 1 hour at 1 atmosphere pressure. The mixture was filtered, and the solution treated with diazomethane. The product was isolated by column chromatography.

The following derivatives were prepared by this method.

Methyl 2-(t-butyloxycarbonylamino)-5-(N-t-butyloxycarbonylindolyl-3')-4-oxopentanoate (5a) and methyl 3-(t-butyloxycarbonylamino)-5-(N-t-butyloxycarbonyl-3')-4-oxopentanoate (6a)

Reaction with 1.41 g N-t-butyloxycarbonyl-L-aspartic anhydride[6] and 4.80 g benzyl 1-t-butyloxycarbonylindole-3-acetate, afforded 1) compound 5a in 30 % yield (oil): $[\alpha]_D^{25}$ +27.9 (c, 0.39); ^{1}H N.m.r. ($CDCl_3$) δ 1.38 (s, 9H, t-Boc), 1.64 (s, 9H, t-Boc), 2.99 (dd, 1H, J=18.1 Hz, J=4.3 Hz, CH_β-CH), 3.19 (dd, 1H, J=18.1 Hz, J=4,3 Hz, CH_β-CH), 3.62 (s, 3H, O-CH_3), 3.74 (s, 2H, Ind-CH_2), 4.46 (m, 1H, CH_α-CH), 5.44 (d, 1H, NH), 7.19-8.11 (m, 5H, indole); ^{13}C N.m.r. ($CDCl_3$) δ 28.15, 39.55 (t-Boc), 43.55 (CH_β-CH), 49.48 (CH_α-CH), 52.47 (O-CH_3), 80.00, 83.70 ($C(CH_3)_3$), 112.57-135.42 (Indole), 149.47, 155.43 (N-CO), 171.70 (CO-OCH_3), 206.00 (ketone); Calcd accurate mass for $C_{24}H_{32}N_2O_7$: 460.2209,

Found: 460.2187, and

ii) compound 6a in 30% yield, m.p. 88°-89° C (ethyl acetate-hexane); $[\alpha]^{25}_D$ -39.6 (c, 0.44); ^{1}H N.m.r. ($CDCl_3$) δ 1.45, 1.64 (s, 9H, t-Boc), 2.77 (dd, 1H, J=17.1 Hz, J=4.9 Hz, CH_2-CH), 2.98 (dd, 1H, J=17.1 Hz, J=4.9 Hz, CH_2-CH), 3.64 (s, 3H, O-CH_3), 3.94 (d, 1H, J=17.3 Hz, Ind-CH_2), 4.00 (d, 1H, J=17.3 Hz, Ind-CH_2), 4.58 (m, 1H, CH_2-CH), 5.66 (d, 1H, NH), 7.18-8.11 (Indole); ^{13}C N.m.r. ($CDCl_3$) δ 28.16, 35.61 (t-Boc), 51.98 (O-CH_3), 55.41 (CH_2-CH), 80.43, 83.54 ($C(CH_3)_3$), 112.59-135.38 (Indole), 149.53, 155.36 (N-CO), 171.96 (CO-OCH_3), 205.30 (ketone); Anal Calcd for $C_{29}H_{33}N_2O_7$: C, 62.60; H, 6.95; N, 6.08. Found C, 61.97; H, 6.82; N, 6.13.

Methyl 2-(benzyloxycarbonylamino)-5-(N-benzylindolyl-3')-4-oxopentanoate (5b) and dimethyl3-(benzyloxycarbonylamino)-5-(N-benzylindolyl-3')-4-oxo-pentanoate (6b)

Reaction with 0.77 g N-carbobenzoxy-L-aspartic anhydride[9] and 2 g t-butyl 1-benzylindole-3-acetate, afforded the coupled products in an overall yield of 56 %. Deprotection of the t-butyl ester with p-toluenesulfonic acid in refluxing benzene, was followed by esterification with diazomethane as previously described, to give

i) 5b in 19 % yield (oil): $[\alpha]^{25}_D$ +12.8 (c, 0.28); ^{1}H N.m.r. ($CDCl_3$) δ 2.99 (dd, 1H, J=18.3 Hz, J=4.0 Hz, CH_2-CH), 3.23 (dd, 1H, J=18.3 Hz, J=4.0

Hz, CH_2-CH), 3.59 (s, 3H, O-CH_3), 3.79 (s, 2H, Ind-CH_2), 4.52 (m, 1H, CH_2-CH), 5.05 (s, 2H, O-CH_2), 5.26 (s, 2H, CH_2-phenyl), 7.04 - 7.49 (m, 15H, aromatic); 13C N.m.r. ($CDCl_3$) δ 39.86 (CH_2-CH), 40.94 (Ind-CH_2), 49.95 (O-CH_3), 52.46 (O-CH_3), 52.65 (CH_2-CH), 69.79 (CH_2-phenyl), 116.75 - 139.66 (aromatic), 155.98 (CO-CH_2-phenyl), 171.46 (CO-OCH_3), 207.02 (ketone); Calcd accurate mass for $C_{29}H_{28}N_2O_5$: 484.1998, Found: 484.1970, and

II) 6b in 37% yield (oil): $[\alpha]^{25}_D$ +0.00 (c, 0.16); ^1H N.m.r. ($CDCl_3$) δ 2.77 (dd, 1H, J=16.9 Hz, J=4.7 Hz, CH_2-CH), 2.96 (dd, 1H, J=16.9 Hz, J=4.5 Hz, CH_2-CH), 3.58 (s, 3H, O-CH_3), 4.00 (s, 2H, CH_2-phenyl), 4.66 (m, 1H, CH_2-CH), 5.06 (d, 1H, J=13.0 Hz, Ind-CH_2), 5.10 (d, 1H, J=13.0 Hz, Ind-CH_2), 5.25 (s, 2H, O-CH_2), 7.04 - 7.51 (m, 15H, aromatic); ^{13}C N.m.r. ($CDCl_3$) δ 35.65 (CH_2-CH), 36.14 (Ind-CH_2), 50.05 (O-CH_3), 51.97 (CH_2-phenyl), 55.47 (CH_2-CH), 109.81 - 137.37 (aromatic), 156.50 (CO-CH_2), 171.80 (CO-OCH_3), 205.31 (ketone); Calcd accurate mass for $C_{29}H_{28}N_2O_5$: 484.1998, Found: 484.1970.

Methyl 3-(acetylamino)-5-(N-t-butyloxycarbonylindolyl-3')-4-oxopentanoate (6c).

Reaction with 1 g N-acetyl-L-aspartic anhydride[6] and 4.64 g benzyl 1-t-butyloxycarbonylindole-3-acetate, afforded compound 6c in 25 % yield (oil):

$[\alpha]^{25}_D$ -18.0 (c, 0.2); ^{1}H N.m.r. ($CDCl_3$) δ 1.63 (s, 9H, t-Boc), 2.03 (s, 3H, N-acetyl), 2.79 (dd, 1H, J=15.0 Hz, J=4.9 Hz, CH_2-CH), 2.97 (dd, 1H, J=15.0 Hz, J=4.5 Hz, CH_2-CH), 3.64 (s, 3H, O-CH_3), 3.89 (d, 1H, J=17.1 Hz, Ind-CH_2), 3.95 (d, 1H, J=17.1 Hz, Ind-CH_2), 4.89 (m, 1H, CH_2-CH), 6.70 (s, 1H, NH), 7.17 - 8.13 (m, 5H, indole); ^{13}C N.m.r. ($CDCl_3$) δ 28.16 (t-Boc), 35.16 (CH_2-CH), 35.63 (Ind-CH_2), 52.08 (CH_2-CH), 54.14 (O-CH_3), 83.70 (C-[CH_3]$_3$), 112.35 - 136.49 (Indole).

Synthesis of t-butyl 1-p-toluenesulfonylindole-3-acetate (7)

A solution of 1-p-toluenesulfonylindole-3-acetic acid[12] (7.2 g, 22.1 mmol), phosphoric acid (1.0 g) and boron trifluoride/etherate (1.3 ml) was stirred in 140 ml dry dichloromethane at room temperature. Isobutylene gas was bubbled through the solution in small bursts for two hours and the solution stirred for a further two hours at room temperature. The reaction mixture was poured into an aqueous sodium bicarbonate solution, the organic layer was separated and the solvent and polyisobutylene removed *in vacuo*. The brown residue was purified by column chromatography (toluene-ethyl acetate) and the product recrystallized from toluene-hexane to afford 6.7 g 7 as pale yellow crystals, m.p. 80-81°C; ^{1}H N.m.r. ($CDCl_3$) δ 1.42 (s, 9H, t-butyl), 2.31 (s, 3H, Tosyl), 3.59 (s, 2H, Ind-CH_2), 7.17-8.00 (9H, m, aromatic); ^{13}C N.m.r. ($CDCl_3$ δ 21.44 (Tosyl), 27.94 (t-butyl), 32.37 (Ind-CH_2), 81.15 (C(CH_3)$_3$), 113.64-144.98 (aromatic), 169.60 (CO); M^+ 385

(Calcd for $C_{22}H_{23}NO_4S$: M, 385); Anal Calcd for $C_{22}H_{23}NO_4S$: C, 65.43; H, 6.01; N, 3.63. Found C, 65.67; H, 6.03; N, 3.72.

Synthesis of α-methyl N-Carbobenzoxy-L-aspartic acid chloride (8)

α-*Methyl* N-Carbobenzoxy-L-aspartate[3] (0.05 g, 0.17 mmol) was dissolved in 10 ml dry THF and stirred under nitrogen at 0°C. Thionyl chloride (0.1 ml, 0.9 mmol) was added by syringe at 0°C and the solution refluxed for one hour. The solvent was removed *in vacuo* and the product 8 crystallized, m.p. 56°-59°C; ^1H N.m.r. ($CDCl_3$) δ 3.48 (dd, 1H, J=18.5 Hz, J=3.7 Hz, CH-C$\underline{H}_2$), 3.56 (dd, 1H, J=18.5 Hz, J=3.7 Hz, CH-C$\underline{H}_2$), 3.74 (s, 3H, O-C$\underline{H}_3$), 4.58 (m, 1H, C$\underline{H}$-CH$_2$), 5.10 (s, 2H, O-C$\underline{H}_2$), 5.72 (d, 1H, NH), 7.30-7.35 (m, 5H, aromatic); ^{13}C N.m.r. ($CDCl_3$) δ 48.75 ($\underline{C}H_2$-CH), 50.56 (CH$_2$-$\underline{C}$H), 53.10 (O-$\underline{C}H_3$), 67.33 (O-$\underline{C}H_2$), 127.99-135.75 (aromatic), 155.67 (N-$\underline{C}$O), 169.70 ($\underline{C}$O-Cl), 171.86 ($\underline{C}$O-CH$_2$); Anal. Calcd for $C_{13}H_{14}ClNO_5$: C, 52.08; H, 4.67; N, 4.67. Found C, 52.76; H, 4.96; N, 5.10.

Synthesis of methyl 2-(benzyloxycarbonylamino)-5-(p-toluenesulfonyl-indolyl-3')-4-oxopentanoate (9)

A solution of diisopropylamine (1.3 g, 13.0 mmol) in 10 ml dry THF was stirred under nitrogen at -78°C. n-Butyllithium (13.0 mmol, 8.1 ml, of a 1.6 N solution in hexane) was added by syringe and the solution stirred for

10 minutes at -78°C. After 30 minutes at room temperature the LDA solution was cooled to -78°C and t-butyl N-p-toluenesulfinyl indole-3-acetate (5.0 g, 13.0 mmol), dissolved in 5 ml dry THF, added by syringe. The solution was stirred for 30 minutes at -78°C and α-methyl N-Carbobenzoxy-L-aspartic acid chloride (1.8 g, 6.2 mmol), dissolved in 10 ml dry THF, added slowly by syringe. After stirring of the solution at -78°C for 2 hours, 1N HCl was added and the pH adjusted to 3. THF was removed *in vacuo* and the aqueous phase extracted with ethyl acetate, the combined organic layers dried over anhydrous magnesium sulphate and the solvent removed *in vacuo*. The product was isolated by column chromatography, affording 1.8 g of the coupling product as an oil. This product (1.8 g, 2.8 mmol) was dissolved in 10 ml dry toluene, p-toluenesulfonic acid (34 mg, 0.2 mmol) added, and the mixture refluxed for 6 hours under nitrogen. Removal of the solvent *in vacuo* was followed by column chromatography to afford 9 in 55 % yield (based on the acid chloride) as a colourless oil: ^{1}H N.m.r. ($CDCl_3$) δ 2.30 (s, 3H, Tosyl), 2.96 (dd, 1H, J=18.0 Hz, J=4.2 Hz, CH_2-CH), 3.18 (dd, 1H, J=18.0 Hz, J=4.2 Hz, CH_2-CH), 3.59 (s, 3H, O-CH_3), 3.71 (s, 2H, Ind-CH_2), 4.53 (m, 1H, CH_2-CH), 5.06 (s, 2H, O-CH_2), 5.70 (d, 1H, NH), 7.17-7.97 (m, 14H, aromatic); ^{13}C N.m.r. ($CDCl_3$) δ 21.47 (Tosyl), 36.65 (CH_2-CH), 39.24 (Ind-CH_2), 49.89 (CH_2-CH), 52.57 (O-CH_3), 64.94 (O-CH_2), 113.71-144.97 (aromatic), 155.93 (CO-CH_2-phenyl), 171.21 (CO-OCH_3), 205.12 (ketone); M^+ 548 (Calcd for $C_{29}H_{28}N_2O_5S$: M, 548).

Acknowledgements: This investigation was supported by the CSIR, P.O. Box 395, Pretoria, South Africa.

REFERENCES AND NOTES

1. This work forms part of a PhD research project at the Rand Afrikaans University.
2. Vleggaar, R.; Ackerman, L.G.J.; and Steyn, P.S., *J.Chem.Soc. Perkin Trans.I*, 1992, 3095.
3. Jackson, R.F.W.; James, K.; Wythes, M.J.; and Wood, A., *J. Chem.Soc., Chem.Comm.*,1989, 644.
4. Mooiweer, H.H.; Hiemstra, K.W.A.; Hiemstra, H.; and Speckamp, W.N., *Tetrahedron*, 1990, 46, 2991.
5. Baldwin, J.E.; Adlington, R.M.; Godfrey, C.R.A.; Gollins, D.W.; Smith, M.L.; and Russel, A.T., *Synlett*, 1993, 51.
6. The synthesis of the relevant protected aspartic anhydride derivatives are described in the following journals: a) Mathias, L.J.; Vaidya, R.A., *J. Mississippi Acad. Sciences*, 1984, XXIX, 13. b) Bergmann, M.; Zervas, L., *Chem.Ber.*, 1932, 65, 1192. c) Barker, C.C., *J.Chem.Soc.*, 1953, 453.
7. Adam, W.; Takayama, K., *J.Org.Chem.*, 1980, 45, 447.

8. N-Carbobenzoxy-α-methyl-L-aspartate was conveniently prepared in three steps from L-aspartic acid: a) Ariyoshi, Y.; Yamatani, T.; Uchiyama, N.; Sato, N., *Bull. Chem. Soc. Jap.*, 1972, 45, 2208. b) Kovacs, J.; Kovacs, H.N.; Ballina, R., *J.Am.Chem.Soc.*, 1969, 85, 1839. c) Bodanszky, M.; Bodanszky, A., *The Practice of Peptide Synthesis*, 1984, 14.

9. Olivier, J.; Holzapfel, C.W., Unpublished results.

10. Williams, K.; Halpern, B., *Chem.Comm.*, 1974, 727.

11. Chapman, R.F.; Phillips, N.I.J.; Ward, R.S., *Tetrahedron*, 1985, 41, 5229.

12. 1-p-Toluenesulfonylindole-3-acetic acid was prepared analogues to the 1-benzylindole-3-acetic acid derivative described in 11.

(Received in the UK 07 April 1993)

Structure Elucidation of Monatin, a High-intensity Sweetener Isolated from the Plant *Schlerochiton ilicifolius*

Robert Vleggaar,[a] Louis G. J. Ackerman[a,b] and Pieter S. Steyn[b]
[a] *Department of Chemistry, University of Pretoria, Pretoria 0002, South Africa*
[b] *Division of Food Science and Technology, CSIR, PO Box 395, Pretoria 0001, South Africa*

The structure of monatin, a high-intensity sweetener isolated from *Schlerochiton ilicifolius* was elucidated by ^1H and ^{13}C NMR spectroscopy as 4-hydroxy-4-(indol-3-ylmethyl)glutamic acid. The $2R^*,4R^*$ relative configuration of the two chiral centres in monatin was determined by NOE studies on a derivative, methyl 2-(indol-3-ylmethyl)-4-(2,4-dinitroanilino)-5-oxo-2,3,4,5-tetrahydrofuran-2-carboxylate. The assignment of the 2*S* configuration to monatin, through application of the Clough–Lutz–Jirgensons rule, established the 2*S*,4*S* configuration for the compound.

The detrimental effects of high consumption of sucrose and glucose-based sugars, such as obesity and tooth decay in humans,[1-3] have resulted in the present day extensive research on synthetic, non-nutritive sweeteners with the aim to develop a safe, high-intensity, non-calorific sweetener, stable for a wide range of pH and temperatures, with a long shelf-life and no aftertaste. None of the current products on the market, *e.g.* saccharin, cyclamate, acesulfame-K and aspartame fulfil all these requirements. [illegible] concern about long-term effects of [illegible] has resulted in a search for high-intensity sweeteners from [illegible] sources.[4] Three of the more promising products are the [illegible] 22 200 dalton protein obtained from African [illegible], glycyrrhizin, a [illegible] obtained from the roots of *Glycyrrhiza glabra*,[6] and stevioside, [illegible] obtained from the herb, *Stevia rebaudiana*.[7]

The rocky hills [illegible] north-eastern Transvaal are the habitat of a spiny-leaved [illegible] shrub, *Schlerochiton ilicifolius*, that grows to a height of [illegible] 2 m. The high-intensity sweetener isolated from the [illegible] roots [illegible] *S. ilicifolius* was [illegible] plant.

The [illegible] roots of *S. ilicifolius* were extracted with water and the [illegible] treated with AG50W-X8 cation ion exchange resin [illegible]. The basic components bound to the resin were [illegible] with aqueous ammonia. The combined [illegible] were freeze-dried and the residual material purified by gel filtration on consecutively (i) Biogel P2 and (ii) Sephadex G10 (2 ×). The sweet tasting fractions were combined and freeze-dried to yield monatin, [illegible] −49.6 as a mixture of salts in which the sodium salt predominated (> 95%). Treatment of an aqueous solution of monatin salt with acetic acid followed by the addition of ethanol gave the free acid of monatin 1 as fine [illegible] crystals, m.p. 216–220 °C, which analysed for $C_{14}H_{16}N_2O_5 \cdot \frac{1}{2}H_2O$. The compound gave positive Ehrlich and ninhydrin colour reactions indicative of the presence of a 3-unsubstituted indole and an amino acid moiety, respectively. The UV spectrum of monatin was typical of an indole and had λ_{max} 279 (ε 5435).

The structure elucidation of monatin is based on a detailed study of the highfield ^1H and ^{13}C NMR spectral data of the compound (Table 1). The signal at δ_H 7.192 (s) was assigned to the C-2 proton of the putative indole moiety. The proton–proton connectivity pattern followed from the first-order analysis of the remaining multiplets as well as $^1H\{^1H\}$ decoupling experiments and established the presence of three fragments A, B and C in monatin.

Fragment A. The proton chemical shifts and coupling constants of this four-proton spin system are typical of the protons of an indole moiety.[8]

Fragment B. The chemical shift values of the resonances at δ_H 3.243 (10-H_a) and 3.051 (10-H_b) and the coupling constant (*J* 14.3 Hz) of this isolated two-proton spin system suggest that C-10 is bonded to an sp^2 hybridized carbon atom.

Fragment C. The signal at [illegible] (*J* 11.5 and 1.8) serves as the [illegible] ABX spin system and is indicative of the [illegible] amino acid. The chemical shifts of the AB part [illegible] 2.030) [illegible] methylene [illegible].

The multiplicities of the [illegible] ^{13}C resonances were deduced from the [illegible] NMR spectrum. The ^{13}C resonances were partly assigned by correlation of the proton-bearing carbon atoms with specific proton resonances in a two-dimensional (^{13}C, ^1H) chemical shift correlation experiment.[10,11] The connectivities between the different structural units followed from the two- and three-bond (^{13}C, ^1H) connectivity patterns [illegible] $^{13}C\{^1H\}$ selective [illegible] (SPI) [illegible] (see Table 2).[12-14]

The [illegible] signal at δ_H 7.192 (d) is assigned to the C-2 proton of the indole ring [illegible] (^{13}C, ^1H) coupling constant [illegible] π-pulse at a position 3.0 Hz [illegible] of the 2-H [illegible] an SPI experiment affected the resonances at δ_C 110.31 (s, C-3), 137.86 (s, C-8) and 129.23 (s, C-9). The assignment of these resonances is based on a comparison with the chemical shift values of other indole compounds, *e.g.* tryptophan.[15] The link between fragment B and the indole ring was established by the (^{13}C, ^1H) connectivity pattern determined for the C-10 methylene protons in an SPI experiment (see Table 2). Irradiation in separate experiments of each of the C-10 protons affected the resonances assigned to C-2, C-3 and C-9 and locates the C-10 methylene group at C-3 of the indole ring. In addition the C-10 methylene protons must be linked to quaternary carbon atoms two- and three-bonds removed as the resonances at δ_C 81.41 (s) and 181.18 (s) are also affected in these SPI experiments. The link between fragments B and C was established by the correlation observed between the C-12 methylene protons and the resonance at δ_C 81.41 (s, C-11) in an SPI experiment. On the basis of the two- and three-bond (^{13}C, ^1H) connectivity pattern the structure as shown in 1 is assigned to monatin.

A less likely alternative structure 2 for monatin in which the hydroxy and amino groups are interchanged, is excluded for a number of reasons. Although the chemical shift values of the

Table 1 ^1H and ^{13}C NMR data for mycestin unit 2 and the derivative 3

	1				3		
Atom	δ_H[a]	J(CH)/Hz	δ_C[b]	J(HH)/Hz	δ_H	δ_C[c]	J(HH)/Hz
2	126.69 (d)	182.2	4.202 (d)		126.62 (d)	7.371 (d)	—
3	110.31 (d)	—	—	—	108.12 (d)	—	—
4	120.46 (d)	159.0	2.836 (d)	7.9	120.65 (d)	7.294 (d)	7.9
5	120.35 (d)[d]	159.4	2.111 (ddd)[e]	8.0, 2.0	119.60 (d)	7.189 (ddd)	7.9, 8.0
6	122.76 (d)[d]	159.6	2.178 (ddd)[e]	8.0, 8.0	122.54 (d)	7.325 (ddd)	8.0, 8.1
7	112.79 (d)	168.3	2.439 (d)	8.1	112.32 (d)	7.502 (d)	8.1
8	137.86 (s)	—	—	—	137.39 (s)	—	—
9	128.29 (s)	—	—	—	125.68 (s)	—	—
10	36.50 (t)	128.2	3.263 (d)	14.3	33.58 (t)	3.513 (d)	14.9
			3.001 (d)	14.3		3.509 (d)	14.9
11	61.41 (s)	—	—	—	58.86 (s)	—	—
12	38.31 (t)	130.2	2.651 (dd)	13.3, 1.7	36.79 (t)	2.775 (dd)	10.5, 13.2
			2.886 (dd)	13.3, 11.7		2.889 (dd)	9.2, 13.1
13	54.39 (d)	144.2	3.168 (dd)	11.4, 1.5	51.32 (d)	3.916 (dd)	
14	173.38 (s)	—	—		171.67 (s)[f]	—	—
15	181.10 (s)	—	—		172.94 (s)[f]	—	—
16					—	6.371 (d)	9.5
17					126.26 (d)[g]	—	—
18					128.05 (d)[g]	—	—
19					124.17 (d)	8.016 (d)	2.6
20					128.28 (s)[g]	—	—
21					130.61 (s)	8.888 (dd)	2.1, 0.4
22					115.37 (d)	8.703 (d)	2.6
OMe					52.21 (q)	3.788 (s)	—

[a] Relative to internal dioxane at δ_H 67.86 in D_2O. [b] Relative to internal dioxane at δ_C 3.770 in D_2O. [c] Solvent [illegible] [d] Assignments may be interchanged. [e] [illegible] [f] [illegible] [g] [illegible]

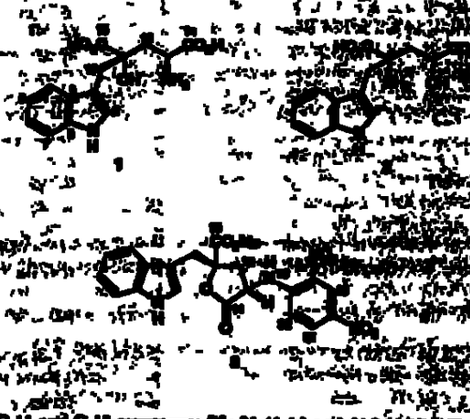


Table 2 Two- and three-bond ^1H–^{13}C connectivities in mycestin unit 2

	δ_H		Correlated ^{13}C	
1	3.263	(3-H)	110.31 (s)	(C-3)
			120.35 (d)	(C-5)
			137.86 (s)	(C-8)
2	3.263	(10-H$_A$)	128.29 (s)	(C-9)
			126.69 (d)	(C-2)
			110.31 (s)	(C-3)
3	3.001	(10-H$_B$)	137.86 (s)	(C-8)
			128.29 (s)	(C-9)
			126.69 (d)	(C-2)
			110.31 (s)	(C-3)
			61.41 (s)	(C-11)
4	2.651	(12-H$_A$)	173.38 (s)	(C-14)
			61.41 (s)	(C-11)
			54.39 (d)	(C-13)
5	2.886	(12-H$_B$)	181.10 (s)	(C-15)
			61.41 (s)	(C-11)
			54.39 (d)	(C-13)
6	3.168	(13-H)	173.38 (s)	(C-14)
			61.41 (s)	(C-11)
			38.31 (t)	(C-12)

C-11 and C-13 resonances [δ_C 61.41 (s) and 54.39 (d)] although against structure 2 the value of J(CH) of 144.2 Hz for C-13 favours other structures.

Additional evidence in favour of structure 1 was obtained from the pH dependence of the ^1H chemical shifts of mycestin. The ^1H chemical shift of the α-proton of α-amino acids is pH dependent.[14] Under basic conditions this proton resonates shifts upfield and under acidic conditions it moves downfield compared to its position under neutral conditions. Although the other protons behave in a similar fashion the effect is less pronounced. The data for selected protons of mycestin are collated in Table 3. From Table 3 it can be seen that 13-H of structure 1 appears at δ_H 3.018 (dd, J 11.6 and 1.8 Hz) but moves upfield by 0.781 ppm at pH 13 and downfield by 0.349 ppm at pH 1. These pH dependent shifts are not be accommodated by the alternative structure 2 and indicated the structure 1 for mycestin. Although the C-12 protons also experience substantial shifts at pH 13 this is due to the cumulative effect of the two carboxylic acid groups.

A derivative of mycestin was prepared by reaction of the amino group of mycestin with Sanger's reagent (2,4-dinitrofluorobenzene) to give the *N*-2,4-dinitrophenyl (DNP) derivative. The product was converted into a dimethyl ester by treatment with an ether solution of diazomethane. Elemental analysis and accurate mass determination of the molecular ion at *m/z* 454, gave the molecular formula of the ester 3 as $C_{21}H_{18}N_4O_9$. This molecular formula as well as the resonances at δ_H 3.788 (s) and δ_C 52.21 (q) in the ^1H and ^{13}C NMR spectra (see Table 1) indicates the presence of a single methyl ester

Table 3 pH Dependence of the chemical shifts of monatin 1

	δ_H				
pH	10-H_a	10-H_b	12-H_a	12-H_b	13-H
Neutral	3.243	3.051	2.611	2.006	3.418
pH 1	3.245	3.084	2.608	2.083	3.767
$\Delta\delta$[a]	−0.002	−0.033	0.003	−0.077	−0.349
pH 13	3.077	2.887	2.207	1.597	2.637
$\Delta\delta$[b]	0.166	0.164	0.404	0.409	0.781

[a] $\Delta\delta$ = δ(neutral) − δ(pH 1). [b] $\Delta\delta$ = δ(neutral) − δ(pH 13).

group in the product 3. The formation of the product 3 can be rationalised by the involvement of the hydroxy group in the formation of a lactone ring by reaction with either the C-11 or the C-13 methoxycarbonyl group, depending on the structure 1 or 2 proposed for monatin. The appearance of the 16-NH proton as a doublet (*J* 9.5 Hz) at δ_H 6.371 as a result of vicinal coupling with the C-13 proton confirms the structure 1 for monatin as no such coupling is possible for the lactone derived from the alternative structure 2.

The relative configuration of the C-2 and C-4 chiral centres in monatin 1 was deduced from the (^{1}H, ^{1}H) NOE connectivity pattern established for the lactone 3 in a number of homonuclear ^{1}H{^{1}H} NOE experiments. Irradiation at the resonance position of one of the C-12 methylene protons (δ_H 2.773, *J* 10.3 and 13.2 Hz) resulted in an NOE only for the other C-12 methylene proton at δ_H 3.251 (*J* 9.2 and 13.1 Hz). In contrast, irradiation of the δ_H 3.193 resonance resulted in the observation of NOEs for 12-H (δ_H 2.773), 13-H (δ_H 5.010) and both C-10 methylene protons (δ_H 3.313 and 3.590). The implication of these results is that the C-11 proton and the indolylmethyl group at C-11 are located on the same face of the lactone ring, *i.e.* are *cis* orientated. Consequently, the lactone 3 must have the 11*R**,13*R** configuration and thus monatin 1 the 2*R**,4*R** configuration.

The 4*S* configuration of monatin followed from the molecular rotations calculated from the observed specific rotations of [α]$_D$ −40.6 (H_2O) and −7.6 (HCl 1 mol dm^{-3}). This finding is in line with the Clough–Lutz–Jirgensons rule[14–17] if it is assumed that the contribution of the C-4 centre to the molecular rotation is not influenced by the change in solvent. This assumption is justified as shown by the literature values for the contributions of the C-2 and C-4 centres to the molecular rotation calculated[18,19] from the observed specific rotations for (2*S*,4*R*)- {[α]$_D$ −30.3 (H_2O); −4.3 (HCl; 0.2 mol dm^{-3})} and (2*S*,4*S*)-4-hydroxy-4-methylglutamic acid {[α]$_D$ +0.5 (H_2O); +23.2 (HCl; 0.2 mol dm^{-3})} isolated from natural sources.[19–21] This result in conjunction with the known relative configuration (see above), establishes the 2*S*,4*S* configuration for monatin.

The sweetness of monatin salt was determined by an experienced ten-member taste panel trained according to established procedures.[22] Panel members were selected for their ability to detect and evaluate sweetness at recognition threshold values, *i.e.* the concentration at which the sensation of sweetness is just discernible. The panel determined the recognition threshold value of sucrose as 0.22–0.44% (w/v) which compares well with the accepted value of 0.3% (w/v). The monatin salt's relative sweetness at a recognition threshold value of 2.75 × 10^{-4}% (w/v) was determined as 800 times that of sucrose (calculated on the lowest value found for sucrose). In addition the panel found that monatin salt exhibited a very slight liquorice aftertaste. Subsequently the relative sweetness of the monatin salt was established as 1400 and 1200 times that of a 5 and 10% (w/v) sucrose solution, respectively. These findings make monatin a very attractive prospect as a high intensity sweetener.[23]

Experimental

M.p.s were determined on a Reichert Kofler hot-stage apparatus and are uncorrected. UV absorptions were measured on a Unicam SP8-100 spectrometer. Infrared spectra were recorded on a Beckman Acculab 8 spectrometer as KBr discs or for solutions in chloroform. Mass spectra were recorded on a Varian MAT 212 double focussing mass spectrometer. NMR spectra were recorded on a Bruker WM-500 (11.7 T) or AC-300 (7.0 T) spectrometer. Optical rotations, measured at 20 °C on a Perkin-Elmer 241 polarimeter, are recorded in units of 10^{-1} deg cm^2 g^{-1}.

Isolation of Monatin 1.—Roots of *Schlerochiton ilicifolius* (160 kg) were collected in the vicinity of the Waterberg in the north-western Transvaal, air-dried and milled to a coarse powder (80 kg). Portions of the ground material (10 kg) were stirred overnight with water (25 dm^3) and the resulting slurry pressed in a mechanical press to obtain an aqueous extract which was filtered to remove fine particles. The extract was stirred in portions (2 dm^3) with Bio-Rad cation exchange resin (AG50WX-8, 500 cm^3) in the H^+ form. The resin was washed with water and [illegible] removed from the resin by stirring with [illegible] ammonia solution (5%; 1 dm^3). The combined ammonia extracts were freeze-dried to give the crude [illegible] component [illegible]. [illegible] was dissolved in water (2 dm^3) and portions [illegible] were applied to a [illegible] gel Sephadex column (100 × [illegible] cm) and eluted with [illegible] [illegible]. The [illegible] were combined and freeze-dried to give a crude product [illegible]. Portions (4.0 g) of this product were applied to a Sephadex G10 gel filtration column (200 × 2.5 cm) and eluted with water at a flow rate of 2 dm^3 min^{-1} [illegible] the sweet tasting fractions were combined and freeze-dried to give an impure crude product (0.8 g). This product was again applied to a Sephadex G10 column [illegible] portions (300 mg) and eluted with water at the same flow rate as previously. The sweet tasting fractions were combined and freeze-dried to give [illegible] 1 (1.73 g) as a mixture of salts, [α]$_D$ −40.6 (*c* 1.00 in H_2O) [illegible] which the sodium salt predominated (> 95%).

The free acid was [illegible] as follows. A [illegible] of [illegible] monatin salts (100 mg) was dissolved in water (1 cm^3) and glacial acetic acid (1 cm^3) was added. Ethanol (0.05 dm^3) was then added and upon storage overnight at room temperature the mixture gave fine [illegible] needle-like crystals of monatin or (2*S*,4*S*)-4-hydroxy-4-(indol-3-ylmethyl)glutamic acid 1, m.p. 216–220 °C (gas evolution), [α]$_D$ −7.6 [*c* 1.00 in HCl (1 mol dm^{-3})]; λ_{max}[NaOH (1 mol dm^{-3})]/nm 279 (ε 3385); ν_{max}(KBr)/cm^{-1} 3396 (NH), 3020, 1580 (CO_2H) and 1040 (Found: C, 56.6; H, 5.5; N, 9.3. $C_{14}H_{16}N_2O_5 \cdot \frac{1}{2}H_2O$ requires C, 56.6; H, 5.6; N, 9.4%).

Methyl (2S,4S)-2-(indol-3-ylmethyl)-4-(2,4-dinitroanilino)-5-oxo-2,3,4,5-tetrahydrofuran-2-carboxylate 3.—Monatin 1 (60 mg) and sodium hydrogen carbonate (20 mg) were stirred for 10 min in an ethanol–water mixture (1:1 v/v). A solution of 1-fluoro-2,4-dinitrobenzene (200 mg) in ethanol (1 cm^3) was added and the mixture stirred overnight at room temperature. It was then acidified (6 mol dm^{-3} HCl) to pH 2 and extracted twice with diethyl ether. The ether extract was treated with an ether solution of diazomethane [generated from Diazald® (*N*-methyl-*N*-nitrosotoluene-*p*-sulfonamide) (2.14 g)] and after 30 min the excess of diazomethane was decomposed by the addition of a few drops of acetic acid. The ether was evaporated

in a stream of nitrogen and the crude product purified by column chromatography on silica gel using ethyl acetate–hexane (1:3 v/v) as eluent. Crystallization from ethyl acetate–hexane yielded crystals of the *title lactone* 3 (25 mg), m.p. 174–175 °C, $[\alpha]_D^{25}$ +0.2 (*c* 1.00 in $CHCl_3$); $\nu_{max}(CHCl_3)/cm^{-1}$ 3350 (NH), 1780 (lactone) and 1740 (ester) (Found: C, 53.2; H, 4.8; N, 11.9%; M^+, 454.1114. [illegible] requires C, 53.2; H, 4.8; N, 12.3%; *M*, 454.1115).

Taste Evaluation of Aromatic Salts.—Water used to dissolve the compounds in the taste evaluation was doubly distilled and then passed through a series of Millipore deionizing filters. The sweetness of the compounds was assessed on a weight by weight basis. The taste panel consisted of 10 members with taste evaluation experience of foodstuffs and beverages, who were selected for their ability to recognize sweetness at the accepted threshold value for sucrose (0.3% w/v).

The recognition threshold value of sucrose was determined by the panel by tasting blind, sucrose solutions with the following concentrations ($\times 10^{-3}$): 0.2, 0.4, 0.8, 1.6, 3.2, 6.4 and 12.8 mol dm^{-3}. Panel members had to indicate at which concentration sweetness could be recognized. This test was repeated 10 times to ascertain the accuracy of the results. Analysis of the results indicated that the panel could detect sweetness in the 6.4–12.8 $\times 10^{-3}$ mol dm^{-3} range, which corresponds to a recognition threshold value of 0.22–0.44% (w/v). Following this procedure the recognition threshold value of the aromatic salt was determined [illegible] by tasting solutions over a concentration range of [illegible] cm^3.

The relative sweetness of the aromatic salt [illegible] and a 10% solution (w/v) [illegible] solutions of the compound at different [illegible] and selecting the concentration at which the [illegible] that of the sucrose solutions. From the known concentration of the selected sample of aromatic salt, the relative sweetness of aromatic salt was determined as 1400 and 1200 times that of sucrose, respectively.

Acknowledgements
The authors thank the Foundation for [illegible] for financial support (to E. V.).

References
1 L. O'Brien, Chem. Brit., 1991, 552.
2 T. H. Grenby, Chem. Brit., 1991, 342.
3 Anon., Food Technol. (Chicago), 1984, 133.
4 A. D. Kinghorn and D. D. Soejarto, CRC Crit. Rev. Plant Sci., 1986, 4, 79.
5 R. B. Iyengar, P. Smits, F. van der Ouderaa, H. van der Wel, J. van Brouwershaven, P. Ravestein, G. Richters and P. D. van Wassenaar, Eur. J. Biochem., 1979, 96, 193.
6 G. E. DuBois, J. Lubinski and C. Nieman, Food Chem., 1980, 28, 135.
7 G. Tomlin, Trends Anal. Chem., 1982, 1, 245.
8 C. J. Pouchert, The Aldrich Library of NMR Spectra, 2nd edn., Aldrich Chemical, Milwaukee, 1983.
9 R. G. Parker and J. D. Roberts, J. Org. Chem., 1970, 35, 996.
10 A. Bax, Two-dimensional Nuclear Magnetic Resonance in Liquids, Delft University Press, Delft, 1982; A. E. Derome, Modern NMR Techniques for Chemistry Research, Pergamon Press, Oxford, 1987.
11 A. Bax, in Topics in Carbon-13 NMR Spectroscopy, ed. G. C. Levy, Wiley, 1984, vol. 4, p. 197.
12 K. G. R. Pachler and P. L. Wessels, J. Magn. Reson., 1973, 12, 337; 1977, 28, 53.
13 W. Voelter, G. Jung, E. Breitmaier and E. Bayer, Z. Naturforsch., B, Chem. Sci., 1971, 26, 213.
14 G. C. K. Roberts and O. Jardetzky, Adv. Protein Chem., 1970, 24, 447; 455.
15 G. W. Clough, J. Chem. Soc., 1918, 113, 526.
16 O. Lutz and B. Jirgensons, Ber. Dtsch. Chem. Ges. B, 1931, 64, 1221.
17 J. P. Greenstein and M. Winitz, Chemistry of the Amino Acids, Wiley, 1961, vol. 1, pp. [illegible].
18 [illegible]
19 [illegible]
20 [illegible]
21 [illegible]
22 [illegible]
23 [illegible]

[illegible]

JOURNAL OF BACTERIOLOGY, Jan. 1992, p. 102–107
0021-9193/92/010102-06$02.00/0

Vol. 174, No. 1

Cloning, Nucleotide Sequence, Overexpression, and Inactivation of the *Escherichia coli* 2-Keto-4-Hydroxyglutarate Aldolase Gene

RAJKUMAR V. PATIL AND EUGENE E. DEKKER*

Department of Biological Chemistry, The University of Michigan, Ann Arbor, Michigan 48109-0606

Received 29 July 1991/Accepted 1 November 1991

Having previously determined the complete amino acid sequence of 2-keto-4-hydroxyglutarate aldolase from *Escherichia coli* (C. J. Vlahos and E. E. Dekker, J. Biol. Chem. 263:11603–11611, 1988), we amplified the gene that codes for this enzyme by the polymerase chain reaction using synthetic degenerate deoxyoligonucleotide primers. The amplified DNA was sequenced by subcloning the polymerase chain reaction products into bacteriophage M13; the nucleotide sequence of the gene was found to be in exact agreement with the amino acid sequence of the gene product. Overexpression of the gene was accomplished by cloning it into the pKK223-3 expression vector so that it was under control of the *tac* promoter and then using the resultant plasmid, pHP6, to transform *E. coli* DH5αF'IQ. When this strain was grown in the presence of isopropyl β-D-thiogalactopyranoside, aldolase specific activity in crude extracts was 80-fold higher than that in wild-type cells and the enzyme constituted approximately 30% of the total cellular protein. All properties of the purified, cloned gene product, including cross-reactivity with antibodies elicited against the wild-type enzyme, were identical with the aldolase previously isolated and characterized. A strain of *E. coli* in which this gene is inactivated was prepared for the first time by insertion of the kanamycin resistance gene cartridge into the aldolase chromosomal gene.

2-Keto-4-hydroxyglutarate (KHG) aldolase (4-hydroxy-2-oxoglutarate glyoxylate-lyase [EC 4.1.3.16; 4-hydroxy-2-oxoglutarate ⇌ pyruvate + glyoxylate]) of *Escherichia coli* is a novel trimeric enzyme (30). Past enzymological and radiotracer studies (7) led to the suggestion that KHG aldolase participates in regulation of the intracellular level of glyoxylate in *E. coli* (and possibly other prokaryotic) cells; i.e., glutamate, it is well established that it catalyzes a central reaction in hydroxyproline catabolism (19). This enzyme has been obtained in homogeneous form from extracts of *E. coli* (5), as well as bovine liver (3) and kidney (12). KHG aldolases from *E. coli* and bovine liver and kidney differ in size (M_r, 66,000, 120,000, and 140,000, respectively), and whereas the liver and kidney enzymes are homotetrameric, that from *E. coli* has a trimeric subunit structure. Another difference is that the liver and kidney enzymes are virtually nonstereospecific in cleaving or forming both optical isomers of KHG (13); the *E. coli* enzyme, in contrast, is stereoselective for L-KHG. Mechanistically, KHG aldolase from all sources so far examined is a class I or "lysine-type" aldolase, catalyzing the reaction via a mechanism involving Schiff base formation between KHG or pyruvate and the ε-amino group of an active-site lysine residue (14, 28). Most interestingly, and regardless of the source, KHG aldolase is bifunctional, efficiently catalyzing β-decarboxylation of oxalacetate as well as aldol cleavage or formation of KHG. The β-decarboxylase/aldolase activity ratios, however, are 0.2, 0.5, and 1.0, respectively, for the kidney, liver, and *E. coli* enzymes. The availability, therefore, of pure KHG aldolase preparations from a prokaryotic source and eukaryotic sources that are similar in some molecular and/or catalytic properties but significantly different in others provides an attractive system for studying enzyme structure-function interrelationships.

First in-depth studies are being carried out with KHG aldolase from *E. coli*. In previous work, we succeeded in determining the complete amino acid sequence of the protein (27); it consists of 213 amino acids with subunit and trimer molecular masses of 22 and 66 kDa, respectively. Selective chemical modification-inactivation studies, followed by peptide isolation and sequencing endeavors, enabled us to determine that glutamate residue 45 (28), arginine residue 49 (27, 29), and lysine residue 133 (27) are each required for catalytic activity. Such efforts await confirmation and, possibly, new insights by application of molecular biological techniques. This report describes our success in cloning the *kga* (KHG aldolase) gene that codes for the primary structure of the enzyme and is the first to describe the nucleotide sequence of this gene from any source. In addition, the gene was overexpressed in *E. coli*, allowing for large quantities of the enzyme to be formed, and the cloned fragment, which included a selectable marker, was used to inactivate the gene on the *E. coli* chromosome by double recombination.

MATERIALS AND METHODS

Bacterial strains, phages, plasmids, and other materials. *E. coli* DH5αF' [F' φ80d *lacZΔM15* Δ(*lacZYA-argF*)*U169* *recA1* *endA1* *hsdR17* (r_K^- m_K^+) *supE44* λ⁻ *thi-1* *gyrA* *relA1*], used for transfection with bacteriophage M13, and *E. coli* DH5αF'IQ [φ80d *lacZΔM15* Δ(*lacZYA-argF*)*U169* *recA1* *endA1* *hsdR17* (r_K^- m_K^+) *supE44* λ⁻ *thi-1* *gyrA* *relA1*/F' *proAB*⁺ *lacI*ᵠ*ZΔM15* *zzf*::Tn5 (Kmʳ)], used for transformation, were obtained from Bethesda Research Laboratories. *E. coli* JC7623 [F' *thr-1* *ara-14* *leuB6* Δ(*gpt-proA*)*62* *lacY1* *sbcC201* *rfx-33* *strB44* *galK2* λ⁻ *rac* *sbcB15* *hisG4* *rfbD1* *recB21* *recC22* *rpsL31* *hdgK51* *xyl-5* *mtl-1* *argE3* *thi-1*], obtained from the *E. coli* Genetic Stock Center, Yale University, New Haven, Conn., was used for transformation of linear DNA fragments. Phages M13mp18 and M13mp19 were used for sequencing, plasmid pKK223.3 was the expression vector for the cloned aldolase gene, and plasmid pUC4K was the source for the kanamycin resistance gene

* Corresponding author.

cartridge; Pharmacia provided these phages and plasmids. All restriction endonucleases, T_4 DNA ligase, 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside, and isopropyl-β-D-thiogalactopyranoside were purchased from Bethesda Research Laboratories. *Taq* DNA polymerase was obtained from United States Biochemical Corp., [α-^{35}S]dATP was from Amersham Corp., and low-melting-point agarose was from FMC Bioproducts. Goat anti-rabbit immunoglobulin G antibody and the reagents for polyacrylamide gel electrophoresis and immunoblotting were purchased from Bio-Rad Laboratories. All other materials were of the highest purity commercially available.

Oligonucleotides. The following single-stranded oligonucleotide primers were synthesized in the Biochemical Research Core Facility at The University of Michigan by the cyanoethyl phosphoramidite method with an Applied Biosystems, Inc., automated DNA synthesizer. Nucleotides placed in parentheses indicate the bases used in degenerate positions: primer 1, GAA TTC ATG AA (AG) AA(TC) TGG AA (AG) ACI (AT) (GC) I GCI GA; primer 2, A AGC TTA IA (AG) (TC)TT IGC ICC (TC)TC IAC IGC (TC)TC IC (TG) IGC. The sequence AAG CTTA at the 5′ end of primer 2 designates the recognition site for *Hin*dIII and the stop codon. These primers were used without high-performance liquid chromatographic purification. Deoxyinosine at the ambiguous third position of selected codons was used to reduce the number of possible structures in the mixed primers.

Polymerase chain reaction (PCR). PCR was performed with a Cetus/Perkin-Elmer DNA thermocycler using *E. coli* K-12 strain W-1485 genomic DNA as the template. A 100-μl PCR reaction mixture contained 1.0 μg of template DNA, 1.0 μg of each primer, each nucleotide at 0.25 mM, and 0.5 μl (2.5 U) of *Taq* DNA polymerase in 10 mM Tris·HCl buffer (pH 8.3) containing 50 mM KCl, 1.5 mM $MgCl_2$, and 0.01% (wt/vol) gelatin. The reaction mixture was covered with 100 μl of light mineral oil to avoid evaporation and then subjected to 35 cycles consisting of a 1.5-min denaturation period at 94°C, a 1-min annealing period at 55°C, a slow rise in 2 min to 72°C to stabilize the hybridized duplex, and a 3-min extension period at 72°C. After the PCR, reaction mixture was analyzed on a 1.2% low-melting-point agarose gel, the amplified product was purified from the melted gel by phenol-chloroform extraction as described by Sambrook et al. (23). The product was then digested with *Eco*RI and *Hin*dIII and cloned into pKK223-3 which had previously been digested with the same two restriction endonucleases.

Purification of plasmid DNA. Plasmid DNA was isolated and purified from fresh overnight cultures by the method of Rodriguez and Tait (21), except that the phenol-chloroform extraction was performed once before ethanol precipitation in 0.3 M sodium acetate.

Subcloning into bacteriophage M13 and DNA sequencing. Positive clones of PCR DNA harvested from cultures of *E. coli* DH5αF′ were purified, and the insert was excised by digestion with *Eco*RI and *Hin*dIII. The resulting fragments were ligated to the replicative forms of M13mp18 and M13mp19 DNAs that had been cut by *Eco*RI and *Hin*dIII and were used to transfect *E. coli* DH5αF′ competent cells prepared by the method of Hanahan (8). Single-stranded DNA from recombinant phage M13 was isolated as described by Sambrook et al. (23). DNA sequencing was performed by the dideoxy-chain termination method using the Klenow fragment of DNA polymerase and the Stratagene DNA sequencing kit with a synthetic 17-base universal primer. Segments of overlapping DNA sequences were generated by subcloning smaller fragments into M13 phage by using new restriction sites as they became evident in the aldolase sequence. Both strands of the *kga* gene were completely sequenced.

Inactivation of the *E. coli* chromosomal *kga* gene. Plasmid pDP6KS was constructed by inserting a purified 1,252-bp *Hin*cII fragment of pUC4K containing the kanamycin resistance gene into the single *Sma*I restriction site of pDP6 (i.e., in the coding region of the *kga* gene). Approximately 2 μg of this plasmid DNA was digested with *Scu*I, a restriction endonuclease that cuts the plasmid only in the vector DNA, and used to transform competent *E. coli* JC7623. Transformation mixtures were plated on Luria-Bertani agar containing 50 μg of kanamycin per ml. After the plates were incubated at 37°C for 16 h, 20 transformants were seen; 7 of these were sensitive to ampicillin.

Immunoblotting of proteins. Western immunoblot assays were performed by standard procedures (9) on soluble fractions prepared from 5-ml Luria broth cultures grown overnight. After being pelleted, the cells were suspended in 1 ml of 10 mM potassium phosphate buffer (pH 8.0) and the suspension was sonicated until a clear lysate was obtained. The supernatant fluids (50 μg of protein) were electrophoresed on sodium dodecyl sulfate (SDS)–8% polyacrylamide gels (17). After electrophoresis, the proteins were electroeluted onto nitrocellulose and the membranes were blocked with 5% powdered milk–1% bovine serum albumin. Subsequent reaction was carried out with crude rabbit antiserum (diluted 1,000-fold in blocking solution) raised against purified *E. coli* KHG aldolase. Antibodies against the aldolase were elicited by subcutaneously injecting the gene protein into three rabbits (~300 μg for each animal over an interval of about 8 weeks), and antiserum was subsequently prepared and collected as described by Harlow and Lane (9). Proteins electroeluted onto nitrocellulose were probed with purified goat anti-rabbit immunoglobulin antibody coupled with 3,000-fold-diluted horseradish peroxidase. The latter was then detected by using the color-developing reagent 4-chloro-1-naphthol (Bio-Rad Laboratories). Controls probed with preimmune serum were negative.

Preparation and assay of KHG aldolase. KHG aldolase was purified to homogeneity as described before (3), from extracts of *E. coli* DH5αF′ harboring the pDP6 plasmid, as well as from extracts of wild-type cells. The substrate KHG was prepared in addition from 4-hydroxyglutamate (19). Aldolase activity was measured colorimetrically by determining the amount of glyoxylate formed by cleavage of KHG after incubating the reaction mixture for 20 min at 37°C (4). One unit of enzyme activity is defined as the amount of protein that liberated 1.0 μmol of glyoxylate per 20 min.

RESULTS

Cloning of the coding region of the *kga* gene. Two oligodeoxynucleotide primers, corresponding to the known 10 N-terminal and the 10 C-terminal amino acids of *E. coli* KHG aldolase (27), were synthesized by using deoxyinosine at the third ambiguous position of selected codons. An *Eco*RI site was placed at the 5′ end of the N-terminal primer, and a seven-base nucleotide sequence, ATTCGAA, containing a stop codon and a *Hin*dIII site, was included at the 5′ end of the C-terminal primer to facilitate subsequent cloning of PCR products. The PCR reaction was carried out as described in Materials and Methods, by using *E. coli* K-12 strain W-1485 genomic DNA. Analysis of the PCR products

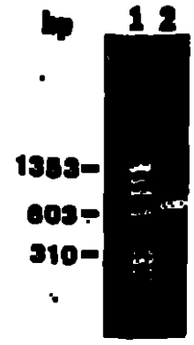


FIG. 1. Agarose gel electrophoretic analysis of PCR reaction products stained with ethidium bromide. Lanes: 1, DNA size markers; 2, PCR products in 2 μl of reaction mixture after 30 cycles. The PCR reaction was carried out with 1.0 μg of *E. coli* K-12 strain W-1485 genomic DNA and 1 μg each of the two primers as described in Materials and Methods.

by agarose gel electrophoresis showed a single major band of the predicted size (Fig. 1).

Subcloning of PCR products for sequencing into M13mp18 and M13mp19. DNA fragments obtained by digestion of the PCR products with *Eco*RI and *Hin*dIII were ligated into M13mp18 and M13mp19 replicative forms of DNA, previously cut by *Eco*RI and *Hin*dIII, and then used to transfect *E. coli* DH5αF′. White recombinants of M13, into which the 640-bp *Eco*RI-*Hin*dIII fragment had been inserted, were picked from plates containing 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside and isopropyl-β-D-thiogalactopyranoside. Single-stranded templates were prepared (23), and DNA sequencing was performed as described in the Stratagene DNA sequencing manual, by using the Klenow fragment of DNA polymerase with [α-^{35}S]dATP and a 17-base universal primer. As sequencing progressed, newly uncovered restriction endonuclease sites were used to clone smaller fragments to obtain the complete gene sequence. The complete nucleotide sequence was established from data for overlapping sequences on both strands. The complete nucleotide sequence determined in this manner was derived from both strands; it and the corresponding deduced amino acid sequence of *E. coli* KHG aldolase are shown in Fig. 2. The nucleotide sequence determined exactly confirmed the previously reported amino acid sequence of the protein (27). Nucleotide sequence analysis indicated that this gene displays highly selective codon usage (data not shown). For example, of the six possible leucine codons, only two (UUG and CUG) are used; likewise, the GGU codon is preferentially used for glycine and so also is CCG for proline. Overall, the G+C content is 57% and comparison of the codon usage in this gene with the codon preference statistics of weakly and strongly expressed *E. coli* genes (6) suggests that this aldolase has a high level of expression. A search of a DNA data base (GenBank) with MicroGenie software (Beckman Instruments) showed no similarities with any reported bacterial or mammalian protein.

Cloning of the kga gene into an expression vector. pKK223.3 is a prokaryotic gene expression plasmid that has the *tac* promoter and an ATG translation initiation codon (2) immediately following a unique *Eco*RI restriction site 4 bases downstream of an AGGA Shine-Dalgarno sequence. The presence of an *Eco*RI site at this position allows for

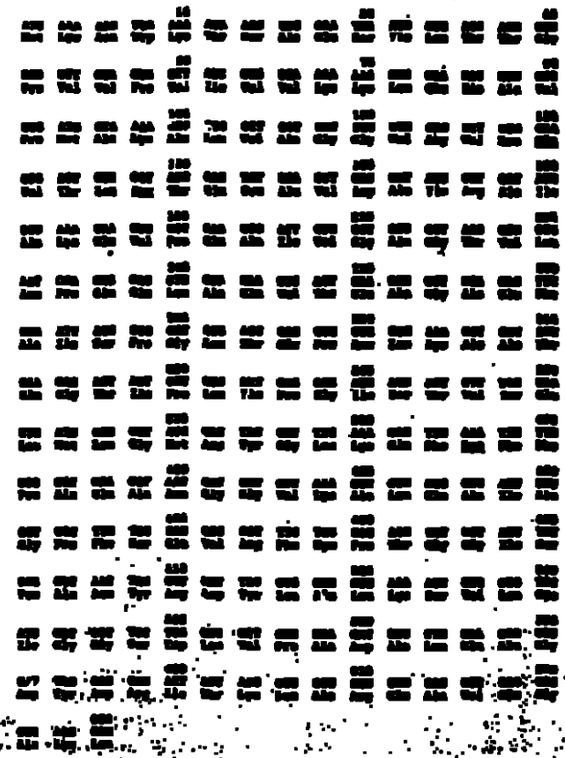


FIG. 2. Nucleotide and encoded amino acid sequences of the *E. coli kga* structural gene. Nucleotide numbers were assigned arbitrarily, beginning with the translation start codon, ATG. The Schiff base-forming lysine residue and the active-site arginine and glutamic acid residues are underlined.

direct insertion of a gene into this vector(s) an optimal distance from the *tac* promoter and in the correct reading frame. Incorporation of an *Eco*RI site and a *Hin*dIII site at the 5′ and 3′ ends, respectively, of the *kga* gene during amplification by PCR was advantageous for cloning of the amplified aldolase DNA directly into the pKK223.3 plasmid previously digested with *Eco*RI and *Hin*dIII (Fig. 3). This hybrid plasmid (designated pDP6) was used to transform *E. coli* DH5αF′IQ. A stop codon was included immediately after the aldolase C-terminal amino acid (i.e., leucine); the first stop codon encountered in pKK223.3 is 78 bases downstream from the C-terminal amino acid, which would give an expressed gene product having 26 additional amino acids at the C terminus.

Characterization of the kga gene product. Plasmid pDP6 was transformed into wild-type *E. coli* DH5αF′IQ and the proteins expressed were analyzed for the presence of the cloned aldolase gene product by SDS-polyacrylamide gel electrophoresis. Crude extracts prepared by sonication of this strain showed a major 22-kDa polypeptide (Fig. 4A, lanes 2 and 3), whereas a polypeptide of the corresponding size was barely detectable in extracts of the wild-type organism (Fig. 4A, lane 1). Under the electrophoretic conditions used, the expressed protein migrated in a manner identical to that of native KHG aldolase purified to homogeneity (Fig. 4A, lane 4). The specific activity of the enzyme in crude extracts of *E. coli* (pDP6) was found to be 80-fold

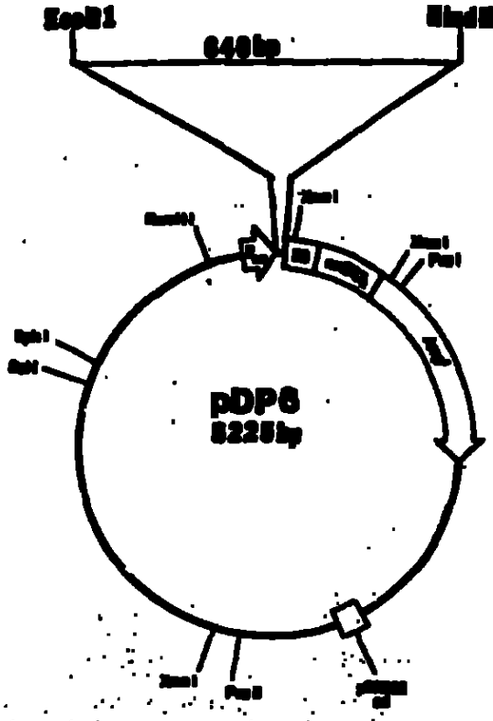


FIG. 3. Construction of plasmid pDP6. PCR-amplified double-stranded DNA was digested by endonucleases *Eco*RI and *Hin*dIII, giving a 640-bp insert containing the *E. coli* aldolase structural gene. This insert was ligated into the *Eco*RI-*Hin*dIII-cut pKK223.3 expression vector. The ligated mixture was used to transform *E. coli* DH5αF'IQ to select recombinant ampicillin-resistant plasmids.

FIG. 4. Overproduction of the KHG aldolase polypeptide in *E. coli* harboring plasmid pDP6. *(A)* SDS-polyacrylamide gel electrophoresis. Lanes: 1, sonicated extracts (50 μg) of wild-type *E. coli* without the plasmid; 2 and 3, sonicated extracts (50 μg) of wild-type *E. coli* carrying plasmid pDP6; 4, purified KHG aldolase (10 μg) from wild-type *E. coli*. (B) Western blot analysis of KHG aldolase as described in Materials and Methods. Lanes: 1, sonicated extracts of wild-type *E. coli* without the plasmid; 2 and 3, sonicated extracts of wild-type *E. coli* carrying plasmid pDP6; 4, purified KHG aldolase (10 μg) from wild-type *E. coli*.

higher than that of wild-type cells, and it was calculated that the cloned gene product constituted ~30% of the total cellular protein.

Past studies had shown that pure KHG aldolase from extracts from either *E. coli* (5) or mammalian sources (3, 12) differs in many fundamental respects from other class I (lysine-type) aldolases. Furthermore, although *E. coli* KHG aldolase has some properties similar to those of the same enzyme from liver or kidney, it is markedly dissimilar in others. As a consequence, we purified KHG aldolase to homogeneity from extracts of *E. coli*(pDP6) by the usual procedure (5) and compared it with the pure enzyme from wild-type *E. coli* cells. As Table 1 shows, the cloned enzyme has the very same properties. The difference in specific activity is not considered to be significant, since values ranging from the low 40s to the high 50s have routinely been obtained for the pure enzyme from wild-type cells. Subunit structure was established by determining its mass after dissociation and subsequent SDS-gel electrophoresis and also by visualization and mass determination of the species separated on SDS-gels after cross-linking with dimethyl suberimidate. The same cross-reactivity with antibodies raised against the purified wild-type enzyme was also observed (Fig. 4B). Collectively, these data clearly indicate that the cloned gene product is the same as the aldolase previously isolated and characterized.

Inactivation of the *kga* gene. We used the method of inserting a selectable marker into the cloned fragment of DNA and then reintroduced it into its original site in the genome by transforming the linear DNA into *E. coli* JC7623 (16). JC7623, a *recB21 recC22 sbcB15* strain, has been used frequently to transform linear DNA (1, 11, 15, 24, 29). Successful use of this technique when the fragment of homologous DNA used contained 0.3- and 2.4-kilobase regions flanking the insertion site has been reported (31). Plasmid pDP6KS was constructed by ligating the kanamycin resistance gene cartridge (1,252-bp *Hin*cII fragment of plasmid pUC4K) into the single *Sma*I site (which lies within the *kga* gene) of plasmid pDP6. After linearization and transformation of plasmid pDP6KS into *E. coli* JC7623 (as described in Materials and Methods), a new kanamycin-resistant strain, *E. coli* RE91, with a mutation in the chromosomal *kga* gene was obtained. Inactivation of the chromosomal gene was confirmed by showing that (i) the strain contained no detectable plasmid DNA; (ii) RE91 grown aerobically for 16 h at 37°C showed no detectable KHG aldolase activity (up to 4 mg of protein was analyzed), whereas good levels of activity (190 Klett units with 0.5 mg of protein) were seen with comparable extracts of strain JC7623; (iii) no inhibitory substance was judged to be present by assaying mixed extracts of strains RE91 and JC7623; and (iv) strain RE91 retained other auxotrophic markers present in parental strain JC7623, as determined by plating on appropriate media. The absence of the aldolase protein and, hence, any activity in strain RE91 was confirmed by Western blotting (Fig. 5A). Extracts of strain RE91 grown in Luria Bertani medium showed no cross-reactive material with antibodies raised against purified aldolase from *E. coli* (lane 3). In contrast, a distinct precipitin band was seen at the normal position for the aldolase protein (lane 2) with strain JC7623. Inactivation of the *kga* gene by insertion of the kanamycin resistance

TABLE 1. Comparative properties of native and cloned *E. coli* KHG aldolases

Enzyme	Sp act (U/mg)	Mol. mass (kDa)	Subunit structure	Optimal pH	K_m[a] (mM)	V_{max}[b]	OAA[c] decarboxylase/KHG aldolase activity ratio
Native	45.0	68	Homotrimer	8.4	2.4, 1.1	3.0, 1.8	1.0
Cloned	55.0	68	Homotrimer	8.4	2.4, 0.8	3.4, 2.1	1.1

[a] For KHG and oxaloacetate, respectively.
[b] Expressed as micromoles of product formed per minute per milligram of protein.
[c] OAA, oxaloacetate.

gene was also established by PCR amplification of the inactivated aldolase gene by using *E. coli* RE91 genomic DNA as the template together with the two primers utilized before in amplifying the *kga* gene from wild-type *E. coli*. Under conditions similar to those outlined earlier for amplification of the coding region of this gene (Fig. 1), the PCR product obtained showed a single band of 1,904 bp after analysis by agarose gel electrophoresis (Fig. 5B, lane 3) rather than 652 bp (the length of the uninterrupted aldolase-coding region, including restriction site linkers at the ends). When genomic DNA from strain JC7623 was used as a template for PCR, only a 652-bp product was amplified (Fig. 5B, lane 1). Analysis of the 1,904-bp DNA fragment by digestion with endonucleases gave a restriction map in full agreement with expectations (data not shown). Inactivation of the *kga* gene was further confirmed by determining the nucleotide sequence of the coding region for KHG aldolase from the *E. coli* RE91 genome. For this purpose, the 3.6-kb *Sal*I-*Bam*HI DNA fragment carrying the kanamycin gene cartridge was first subcloned into plasmid pUC18. The nucleotide sequence for the coding region of the aldolase was then obtained by using synthetic primers for the known end sequences of the kanamycin gene. This is the first report of successful gene inactivation wherein only 282- and 357-bp regions of homologous DNA flanked the inserted element. Having an *E. coli* mutant unable to synthesize a functional *kga* gene product in hand, we searched for a phenotype by monitoring the aerobic growth of strain RE91 on various carbon compounds by using $(NH_4)_2SO_4$ as the sole nitrogen source. In a series of experiments (data not shown), almost identical doubling times were observed for strains RE91 and K-12 in Luria Bertani medium and in minimal medium containing glucose, fructose, glycerol, or lactate. Obviously, a phenotype for the lack of KHG aldolase remains to be established.

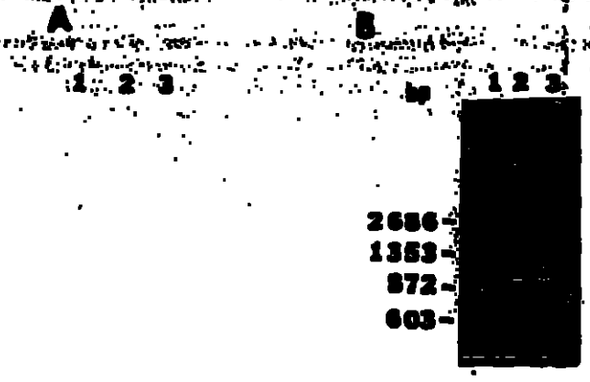


FIG. 5. (A) Western blot analysis of extracts of *E. coli* JC7623 and RE91. Lanes: 1, purified KHG aldolase (5.0 μg of protein) from wild-type *E. coli*; 2, sonicated extracts of *E. coli* JC7623 (50 μg of protein); 3, sonicated extracts of *E. coli* RE91 (50 μg of protein). (B) Agarose gel electrophoresis of PCR products stained with ethidium bromide. Lanes: 1, PCR products in 2 μl of a reaction mixture containing *E. coli* JC7623 genomic DNA as the template; 2, DNA size markers; 3, PCR products in 2 μl of a reaction mixture containing *E. coli* RE91 genomic DNA as the template.

DISCUSSION

Extensive previous research, primarily with fructose 1,6-bisphosphate aldolase, has led to the designation of class I versus class II aldolases and delineation of their characteristic properties (10, 18, 22). β-Decarboxylases (like acetoacetate decarboxylase) are also present as one of these two general types. KHG aldolase is somewhere in the middle of such a classification. Mechanistically, it is in class I regardless of the source (mammalian or bacterial). Although *E. coli* KHG aldolase has the added novel property of being a trimer, the enzyme as obtained in pure form from either *E. coli*, liver, or kidney, has a high level of β-decarboxylase activity toward oxaloacetate. Apparently, only KHG aldolase among well-studied class I aldolases exhibits such bifunctional activity. Since a Schiff base mechanism applies to both aldolases and β-decarboxylases of the lysine type, this unusual bifunctional property of KHG aldolase attracts detailed studies at the molecular level in light of the phylogenetic concept (22) that class I aldolases may be structurally resistant to preexisting enzyme such as β-decarboxylase. One goal, therefore, is to establish structure-function interrelationships of KHG aldolase from prokaryotic and eukaryotic sources. First efforts have concentrated on the pure enzyme protein from *E. coli*; some details of its structure and of amino acid residues that are catalytically essential have been uncovered (26–29).

This is the first report of cloning of the *kga* gene from any source. The methods and results outlined represent a necessary initial step for subsequent incisive studies that will establish at the molecular level the relationship of structure to function for this unusual enzyme. Cloning of this gene presented a challenge because of the difficulty in selecting mutants that lacked the KHG aldolase phenotype. Typical colony hybridization experiments to screen an *E. coli* library and examination of some 400 to 500 separate mutants of *E. coli* after Tn*10* transposon insertion did not achieve the desired goal. To overcome such hurdles, the coding region of the *E. coli* enzyme was cloned by PCR methodology. Success of this strategy was evidenced by DNA sequencing of the gene and by in vivo expression of the aldolase protein. Determining that the nucleotide sequence of the coding region for KHG aldolase from the *E. coli* RE91 genome exactly matched that of the PCR-prepared DNA indicated that no errors were made during the amplification procedure.

Efforts to sequence the DNA regions flanking the *kga* gene are now in progress. When the cloned gene was overexpressed under control of the *tac* promoter, markedly elevated levels of the gene product were found. Attainment of the goal of purifying large amounts of the homogeneous enzyme for possible crystallization and X-ray crystallographic studies of this trimeric protein now seems possible. Definitive structure-function interrelationships of *E. coli* KHG aldolase can now also be addressed by site-directed mutagenesis of specific amino acid residues. Furthermore, site-directed insertion of the kanamycin resistance gene into the *E. coli* chromosomal gene enabled us to isolate a mutant lacking KHG aldolase activity. This mutant strain will be useful in expressing and characterizing mutant forms of the enzyme obtained by site-directed mutagenesis, in cloning the complete gene with its own regulatory sequences, and in mapping its locus on the chromosome.

ACKNOWLEDGMENT

This research was supported by Public Health Service grant DK-03718 from the National Institutes of Health.

REFERENCES

1. Bennett, C. L., and S. R. Kushner. 1984. Exonucleases I, III, and V are required for stability of ColE1-related plasmids in *Escherichia coli*. J. Bacteriol. 157:661–664.
2. Bolivar, F. A., L. J. Caruthers, and N. Ramer. 1983. The *tac* promoter: a functional hybrid derived from the *trp* and *lac* promoters. Proc. Natl. Acad. Sci. USA 80:21–25.
3. Bolivar, F. G., R. D. Kohen, and S. R. Gordy. 1975. 2-Keto-4-hydroxyglutarate aldolase from bovine liver. Methods Enzymol. 42:283–289.
4. Bolivar, F. G., and G. Mattoo. 1982. Conversion of γ-hydroxyglutamate to glyoxylate and alanine: purification and properties of the enzyme system. J. Biol. Chem. 257:2215–2221.
5. Bolivar, F. G., R. Nakkharn, and S. R. Gordy. 1975. 2-Keto-4-hydroxyglutarate aldolase from *Escherichia coli*. Methods Enzymol. 42:283–289.
6. Gregory, E., and W. Fierz. 1982. Preferential codon usage in prokaryotic genes: the optimal codon-anticodon interaction energy and the selective codon usage in efficiently expressed genes. DNA 2:179–199.
7. Gruys, K. C., and H. E. Dekker. 1984. Methylcitrate formation in the NAD⁺-, CoA(III)-, and α-ketoglutarate dehydrogenase-dependent oxidation of 2-keto-4-hydroxyvalerate. Possible coupled role of this reaction with 2-keto-4-hydroxyglutarate aldolase activity in a pyruvate-catalyzed cyclic oxidation of glycolate. J. Biol. Chem. 259:10012–10020.
8. Hanahan, D. 1983. Studies on transformation of *Escherichia coli* with plasmids. J. Mol. Biol. 166:557–580.
9. Harlow, E., and D. Lane. 1988. Antibodies: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
10. Harris, C. K., R. D. Kohen, B. C. Tobin, and W. J. Reiter. 1969. The molecular characteristics of yeast aldolase. Biochemistry 8:2443–2454.
11. Inselburg, J. 1978. ColE1 plasmid mutants affecting growth of an *Escherichia coli recB recC sbcB* mutant. J. Bacteriol. 133:413–416.
12. Kilian, R. P. 1988. Ph.D. thesis. The University of Michigan, Ann Arbor.
13. Kohen, R. D., and F. E. Dekker. 1971. Variant properties of bovine liver 2-keto-4-hydroxyglutarate aldolase: its β-decarboxylase activity, lack of substrate stereospecificity, and structural requirements for binding substrate analogs. Biochim. Biophys. Acta 250:238–250.
14. Kohen, R. D., and F. E. Dekker. 197. 2-Keto-4-hydroxyglutarate aldolase of bovine liver. Schiff-base formation with 2-keto-4-hydroxyglutarate, pyruvate, and glycolate. Biochemistry 10:388–395.
15. Kushner, S. R., H. Nagaishi, and A. J. Clark. 1972. Indirect suppression of *recB* and *recC* mutations by exonuclease I deficiency. Proc. Natl. Acad. Sci. USA 69:1366–1370.
16. Kushner, S. R., H. Nagaishi, A. Templin, and A. J. Clark. 1971. Genetic recombination in *Escherichia coli*: the role of exonuclease I. Proc. Natl. Acad. Sci. USA 68:824–827.
17. Laemmli, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) 227:680–685.
18. Lebherz, H. G., and W. J. Rutter. 1973. A class I (Schiff base) fructose diphosphate aldolase of prokaryotic origin. Purification and properties of *Micrococcus aurogenes* aldolase. J. Biol. Chem. 248:1658–1669.
19. Meiers, U., and F. E. Dekker. 1983. Enzymatic steps in the conversion of γ-hydroxyglutamate to glyoxylate and alanine. J. Biol. Chem. 258:1660–1668.
20. Nakkharn, R., and R. E. Dekker. 1972. Purification, substrate specificity and binding, β-decarboxylase activity, and other properties of *Escherichia coli* 2-keto-4-hydroxyglutarate aldolase. J. Biol. Chem. 247:5979–5987.
21. Old, R. W., and S. B. Primrose. 1985. Recombinant DNA technology: An introduction. The Benjamin/Cummings Publishing Co., Inc., Menlo Park, Calif.
22. Rutter, W. J. 1964. Evolution of aldolases. Fed. Proc. 23:1248–1257.
23. Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
24. Shimizu, Y.-F., and S. Udenfriend. 1972. Degradation of hexuronic acids, 2-keto-3-deoxyhexonic acid aldolase by *Escherichia coli* K-12. J. Bacteriol. 110:1058–1063.
25. Vapnek, D., M. E. Alton, C. L. Bennett, and S. R. Kushner. 1976. Amplification in *Escherichia coli* of enzymes involved in genetic recombination: construction of hybrid ColE1 plasmids carrying the structural gene for exonuclease I. Proc. Natl. Acad. Sci. USA 73:3491–3495.
26. Vlahos, C. J., and F. E. Dekker. 1988. Amino acid sequence of the active-site tryptic glycopeptide from the lysine prosite of *Escherichia coli* 2-keto-4-hydroxyglutarate aldolase. J. Biol. Chem. 263:11092–11097.
27. Vlahos, C. J., and F. E. Dekker. 1988. The complete amino acid sequence and identification of the active-site tryptic peptide of *Escherichia coli* 2-keto-4-hydroxyglutarate aldolase. J. Biol. Chem. 263:11085–11091.
28. Vlahos, C. J., and F. E. Dekker. 1988. Active-site mutants of 2-keto-4-hydroxyglutarate aldolase from *Escherichia coli*. Bromopyruvate inactivation and labeling of glutamate 45. J. Biol. Chem. 263:20084–20089.
29. Vlahos, C. J., M. A. Gleissner, and F. E. Dekker. 1989. Evidence for an essential arginine residue in the active site of *Escherichia coli* 2-keto-4-hydroxyglutarate aldolase. Modification with 1,2-cyclohexanedione. J. Biol. Chem. 264:5483–5485.
30. Wang, J. K., F. E. Dekker, H. D. Lawhead, and H. C. Water. 1981. Physical and chemical evidence for the trimeric subunit structure of 2-keto-4-hydroxyglutarate aldolase from *Escherichia coli* K-12. J. Biol. Chem. 256:1795–1800.
31. Wilson, K. C., K. J. Elledge, J. H. Krueger, and G. C. Walker. 1983. Site-directed insertion or deletion mutagenesis with cloned fragments in *Escherichia coli*. J. Bacteriol. 155:1219–1221.

APPLIED AND ENVIRONMENTAL MICROBIOLOGY, Dec. 1997, p. 4651–4656 Vol. 63, No. 12
0099-2240/97/$04.00+0

Synthesis of Optically Active Amino Acids from α-Keto Acids with *Escherichia coli* Cells Expressing Heterologous Genes

ANDREY GALKIN,[1] LJUDMILA KULAKOVA,[1] TOHRU YOSHIMURA,[1] KENJI SODA,[2] AND NOBUYOSHI ESAKI[1]*

Institute for Chemical Research, Kyoto University, Uji, Kyoto-Fu 611,[1] and Faculty of Engineering, Kansai University, Suita, Osaka 564,[2] Japan

Received 20 June 1997/Accepted 19 September 1997

We describe a simple method for enzymatic synthesis of L and D amino acids from α-keto acids with *Escherichia coli* cells which express heterologous genes. L-amino acids were produced with thermostable L-amino acid dehydrogenase and formate dehydrogenase (FDH) from α-keto acids and ammonium formate with only an intracellular pool of NAD^+ for the regeneration of NADH. We constructed plasmids containing, in addition to the FDH gene, the genes for amino acid dehydrogenases, including L-, leucine dehydrogenase, alanine dehydrogenase, and phenylalanine dehydrogenase. L-Leucine, L-valine, L-norvaline, L-methionine, L-phenylalanine, and L-tyrosine were synthesized with the recombinant *E. coli* cells with high chemical yields (>80%) and high optical yields (up to 100% enantiomeric excess). Stereospecific conversion of various α-keto acids to D amino acids was also examined with recombinant *E. coli* cells containing a plasmid coding for the four heterologous genes of the thermostable enzymes D-amino acid aminotransferase, alanine racemase, L-alanine dehydrogenase, and FDH. Optically pure D enantiomers of glutamate and leucine were obtained.

The worldwide market value of amino acids is approximately 2 billion dollars annually (22), and the synthesis of optically active amino acids has been extensively studied. Microbial processes have been developed for large-scale industrial production of natural proteinogenic L amino acids by fermentation. However, several other methods have been developed for the production of unnatural D and L amino acids; these methods include the enzymatic resolution of racemic amino acid amides with D- or L-aminopeptidase (13).

A method for enzymatic synthesis of L amino acids from α-keto acids with NAD(P)-dependent amino acid dehydrogenases has been proposed (10, 19). However, the application of this method to industrial production of L amino acids has been hampered by the cost of coenzymes. A multienzyme reaction system for simultaneous coenzyme regeneration (Fig. 1) has been proposed to overcome this problem. L-Leucine has been produced from α-ketoisocaproate (2-oxo-4-methylpentanoic acid) with leucine dehydrogenase (LeuDH) from *Bacillus sphaericus*, formate dehydrogenase (FDH) from *Candida boidinii*, and NADH covalently bound to water-soluble polyethyleneglycol in a reactor with an ultrafiltration membrane (6, 28). The system is also applicable to the production of several other aliphatic L amino acids, such as L-valine, L-*tert*-leucine, and L-[^{15}N]leucine. The process has been successfully scaled up for industrial production of these L amino acids (10). A similar system has been developed with phenylalanine dehydrogenase (PheDH) and FDH for the production of L-phenylalanine (1, 2). The synthesis of L-β-chloroalanine by using L-alanine dehydrogenase (AlaDH), LeuDH, and PheDH in combination with FDH has also been studied (14). However, the instability of the amino acid dehydrogenases used in these processes has hampered the efficient operation of the systems. Improvement has been achieved through the use of thermostable enzymes (18).

The only commercially available preparation of FDH is the enzyme from *C. boidinii* and is not sufficiently stable. Recently, we cloned and expressed in *Escherichia coli* the genes for bacterial FDH, which is more stable than the yeast enzyme (8). Thermostable alanine racemase (AlaR) (25), AlaDH (15), and D-amino acid aminotransferase (DAAT) (EC 2.6.1.21) (26) have been developed, and an efficient system for conversion of α-keto acids to D amino acids by the coupling of these four enzymes has been described (9) (Fig. 1). This method is also applicable to the synthesis of ^{15}N-labelled D amino acids, which are expected to be valuable in the study of mammalian neural systems (7).

The industrial use of the systems mentioned above depends chiefly on the cost of the enzymes, although intact cells of microorganisms containing the enzymes can be used as catalysts in order to decrease costs (21). In most cases, however, additional genetic improvements through metabolic engineering are required, and new functional combinations are made by the rational transfer of pathways from one organism to another (4). The transfer of the ethanol pathway from *Zymomonas mobilis* to other enteric bacteria represents an example of this approach (12). In our system, various L and D amino acids can be produced from α-keto acids if two or four functional genes are introduced into one microorganism according to the schemes presented in Fig. 1. The simultaneous expression of all enzymes in a single cell should provide additional benefit for industrial applications; the intracellular pool of NAD^+ (supplied by the cell itself) could be used for NADH regeneration without any additional supplies. We describe here a simple method for the conversion of α-keto acids to L and D amino acids with recombinant *E. coli* cells which contain plasmids with heterologous genes necessary for biotransformations.

MATERIALS AND METHODS

Plasmids and other materials. Restriction enzymes and other DNA-modifying enzymes were purchased from Toyobo or Takara and were used according to the manufacturers' instructions. Leucine dehydrogenase from rabbit muscle was obtained from Sigma. *Bacillus stearothermophilus* AlaDH was obtained from Unitika, Osaka, Japan. NAD^+ and NADH were obtained from Serva, and α-keto acids were obtained from Sigma. All other chemicals were obtained from Nacalai Tesque or Wako Pure Chemicals (Japan). The oligonucleotides were purchased

* Corresponding author. Phone: 81-774-38-3240. Fax: 81-774-38-3248. E-mail: esaki@pclsp2.kuicr.kyoto-u.ac.jp.

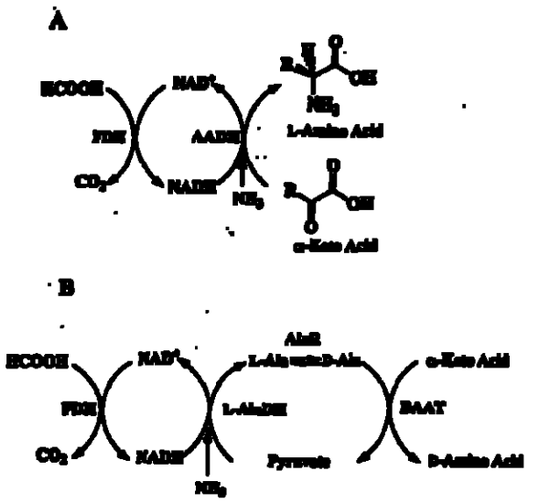


FIG. 1. Synthesis of L amino acids (A) and D amino acids (B) by coupling of enzyme reactions. AADH, L-amino acid dehydrogenase (i.e., AlaDH).

from Biologica, Nagoya, Japan. Vector plasmids pUC19 and pUC119 were used for gene cloning and expression.

DNA techniques. DNA sequencing was performed with an Applied Biosystems model 370A DNA sequencer and a dye-labelled terminator sequencing kit (Applied Biosystems). DNA synthesis by PCR was performed with a DNA Thermal Cycler (Perkin-Elmer) in 0.1 ml of a mixture containing each deoxynucleoside triphosphate at a concentration of 0.1 mM, 100 pmol of primers, 200 ng of template DNA, 0.01 ml of 10× reaction buffer (Takara), and 2.5 U of BcaBest DNA polymerase (Takara). The PCR program consisted of denaturation at 95°C for 1 min, annealing at 55°C for 1 min, and extension at 72°C for 2 min for a total of 20 cycles. Site-directed mutagenesis was carried out with a DNA mutagenesis kit from Amersham by using the protocol given by the supplier. All of the primers used are shown in Table 1.

Construction of pFDHLeuDH2, pFDHAlaDH, pFDHPheDH, and pFADA. Plasmids for the simultaneous expression of FDH and LeuDH (pFDHLeuDH2), FDH and AlaDH (pFDHAlaDH), FDH and PheDH (pFDHPheDH), and FDH, AlaDH, DAAT, and AlaR (pFADA) were constructed as shown in Fig. 2. Plasmids pULDH2, pICD301, pKPDH2, pMcFDH, pICT113p, and pMDalr3 encoding the genes for LeuDH (20), AlaDH (13), PheDH (24), FDH (5), DAAT (26), and AlaR (11), respectively, were prepared as described previously and used as template DNAs for PCR (Table 1). A unique SphI site was introduced into the FDH gene in pMcFDH by site-directed mutagenesis by using the oligonucleotide shown in Table 1. The LeuDH gene was cloned by PCR by using pULDH2 as a template and was introduced into pMcFDH at the SphI site to yield pFDHLeuDH2. Then, a 2.3-kb *Sal*I-*Eco*RI fragment including the structural genes for FDH and LeuDH derived from pFDHLeuDH2 was ligated with a 3.1-kb *Sal*I-*Eco*RI fragment derived from pFDH4 (27) to produce pFDHLeuDH2. pFDHAlaDH and pFDHPheDH were constructed in the same manner (Fig. 2). The *Xho*I-*Eco*RI fragment containing the DAAT gene was isolated from pICT113p and introduced into pFDHAlaDH to obtain pFAD. Then, the *Nde*I-*Eco*RI fragment encoding the structural gene for AlaR, which was produced with pMDalr3 by PCR, was introduced into pFAD to produce plasmid pFADA (Fig. 2).

Cultivation of recombinant *E. coli* strains and enzyme assays. Plasmid pFDHLeuDH2, pFDHAlaDH, pFDHPheDH, or pFADA was introduced into *E. coli* TG1 [F'traD36 *proAB lacI*^{q} Δ*lacZ M15* Δ(*lac-pro*) *thi hsdR sup*E], and the resulting transformant was cultivated in Luria-Bertani medium (14) supplemented with 50 mg of ampicillin per liter and 0.2 mM isopropyl-β-D-thiogalactopyranoside. The cells were collected at the beginning of the stationary phase of the culture, and cell extracts were prepared by sonication. FDH, AlaDH, PheDH, and LeuDH were assayed with a thermostatic model 3450HRS spectrophotometer by following the reduction of NAD$^+$ at 340 nm at 37°C by using the methods described previously (15, 20, 24, 27). DAAT was also assayed spectrophotometrically in the coupled reaction with lactate dehydrogenase (17). AlaR activity was measured in the direction from D- to L-alanine by monitoring the production of L-alanine with AlaDH (13). One unit of enzyme was defined as the amount of enzyme that catalyzed the formation of 1 μmol of product (NADH or NAD$^+$) in 1 min.

Production of L and D amino acids. The dried cells were collected as described above and washed with 0.85% NaCl. The washed cells (wet weight, 0.1 g) were suspended in 2 to 5 ml of 0.5 M ammonium formate (pH 7.5) containing 0.05 to 0.6 M α-keto acids. The standard reaction mixture for D amino acid synthesis contained 0.05 to 0.3 M α-keto acid, 0.5 M ammonium formate (pH 7.5), 5 to 10 mM pyruvate, and washed cells (wet weight, 0.2 g) in a total volume of 2 ml, and the reaction was performed at 37°C with reciprocal shaking. Aliquots of the reaction mixture were removed for analysis of substrates and products.

Identification of reaction products and determination of optical purity. The products obtained by reductive amination of α-ketoisocaproate, α-ketobutyrate, α-ketovalerate, α-keto-γ-thiomethylbutyrate, pyruvate, phenylpyruvate, *p*-hydroxyphenylpyruvate, and α-ketoglutarate with recombinant *E. coli* cells were identified with an amino acid analyzer (model L-8500; Hitachi, Tokyo, Japan) as leucine, valine, norvaline, methionine, alanine, phenylalanine, tyrosine, and glutamate, respectively. The optical purity (the enantiomeric excess) of the amino acids produced was determined at 25°C with a high-performance liquid chromatography (HPLC) column (MCI GEL CRS 10W; Mitsubishi, Tokyo, Japan). The solvent used was 0.2 to 2.0 mM $CuSO_4$ at a flow rate of 0.5 to 1 ml/min.

The consumption of α-keto acids was monitored by HPLC with an Ultron PS80H column (Shinwa Kako, Kyoto, Japan) at 60°C with 0.02 M perchloric acid (pH 2.0) at a flow rate of 0.8 ml/min.

RESULTS

Construction of plasmids for the production of L and D amino acids. Plasmid pFDHLeuDH2, used for simultaneous expression of FDH and LeuDH, was constructed in two steps, as shown in Fig. 2. Both LeuDH and FDH were efficiently expressed through the tandem *lac-tac* promoter of pFDHLeu

TABLE 1. Primers and template DNAs used for PCR amplification of the genes used for production of L and D amino acids

Gene	Primer	Template	Reference
FDH	5′-GCCCTGGGAATTCGAGCATGCTCAGACC-3′[a]	pMcFDH	11
LeuDH	Forward 5′-GGGCATGCGATTGAGGAGGAATGAAAG-3′[b]	pULDH2	21
	Reverse 5′-GCAGAATTCCATCCCTTATTTGTTGTT-3′[b]		
AlaDH	Forward 5′-GAAGCGGCATGCAAAGGAGGAAATG-3′[c]	pICD301	13
	Reverse 5′-GGCGGCTCCGAATTCAAGATCTCATCCGTGC-3′[c]		
PheDH	Forward 5′-AGCGGCATGCTTGTTGAGGAAGCGAAG-3′[d]	pKPDH2	22
	Reverse 5′-CATCAGTTTCATGAATTCTTTTTACCTCC-3′[d]		
DAAT	Forward 5′-TTAGCAGCGAAGATCTGAAGGATGTGGG-3′[e]	pICT113p	14
	Reverse 5′-AGCAAAACCAAAGAATTCTCACTGCAGATTATA-3′[e]		
AlaR	Forward 5′-GCGCTTTCAAGTATGCATAACAGGAAAGGC-3′[f]	pMDalr3	12
	Reverse 5′CTTTGTCTTTTGAATTCGATTATGCACTGC-3′[f]		

[a] The underlined site is the SphI restriction site.
[b] The underlined site is the *Eco*RI restriction site.
[c] The underlined sites are the *Xho*II and *Bgl*II restriction sites (in that order).
[d] The underlined site is the *Bgl*II restriction site.
[e] The underlined sites are the *Xho*II and *Pst*I restriction sites (in that order).
[f] The underlined site is the *Nde*I restriction site.

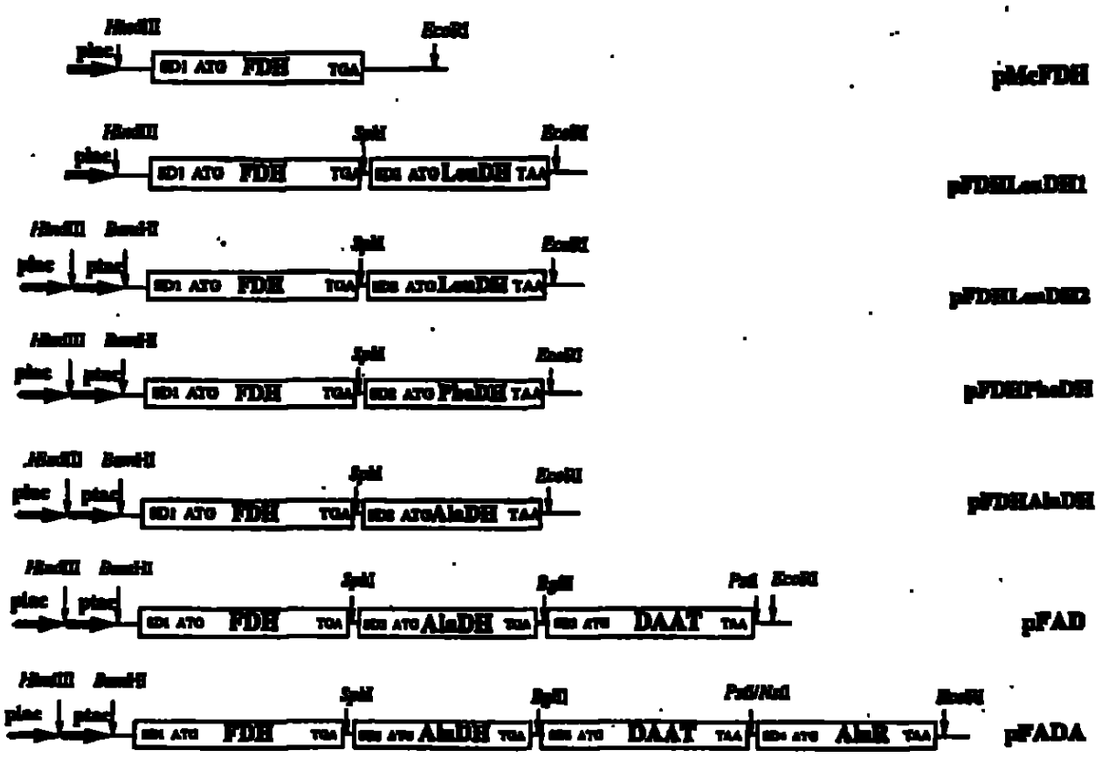


FIG. 2. Construction of the plasmids used for production of L and D amino acids by expression in *E. coli* cells.

DH2 in *E. coli* TG1. The specific activities in the clone cell extract were 1.1 U of FDH per mg and 8.2 U of LeuDH per mg. By comparing the specific activities of the homogeneous preparations of LeuDH and FDH with the specific activities in the extract, the amounts of FDH and LeuDH produced in the clone cells were each estimated to be about 7% of the total soluble protein. The levels of expression (i.e., specific activities in the cell extract) of AlaDH and PheDH were similar to those of LeuDH (1.2 U/mg for FDH and 7.3 U/mg for AlaDH with pFDHAlaDH and 1.0 U/mg for FDH and 6.7 U/mg for PheDH with pFDHPheDH). Plasmid pFADA, used for the simultaneous expression of all four genes required for the synthesis of D amino acids, was also constructed. The levels of expression of the four enzymes produced with pFADA were a little lower than the levels of expression of the enzymes described above (0.65 U of FDH per mg, 6.0 U of AlaDH per mg, 5.3 U of DAAT per mg, and 5.0 U of AlaR per mg).

Synthesis of L *amino acids.* LeuDH acts on various branched- and straight-chain amino acids and their α-keto analogs, and the relative rates of synthesis of L amino acids with resting cells of *E. coli* TG1 carrying pFDHLeuDH2 were similar to the rates of synthesis obtained with purified LeuDH from *Thermoactinomyces intermedius* (20) (data not shown). L amino acids with high optical purity were synthesized from α-keto acids by using ammonium formate (Table 2). L-Leucine produced from α-ketoisocaproate was crystallized in the reaction mixture during the incubation period (Fig. 3A). Crystals of L-norleucine also appeared when the concentration of this compound reached 0.14 M. L-Valine and L-norvaline were synthesized at a final concentration of 0.38 M, while the final concentrations of L-methionine and L-α-aminobutyrate were approximately 0.35 M.

L-Alanine was produced from pyruvate with *E. coli* TG1 which contained pFDHAlaDH (Table 2). The amount of L-alanine produced increased as the concentration of pyruvate increased (Fig. 3B). However, the overall yield of L-alanine decreased as the pyruvate concentration increased. Moreover, the optical purity of L-alanine decreased with prolonged incubation; the enantiomeric excess of L-alanine was 88% after 3 h, while it was only 80% after 10 h. This was probably due to the action of the AlaR produced by the host cells.

L-Phenylalanine and L-tyrosine were synthesized with *E. coli* TG1 which contained pFDHPheDH (Table 2). Because PheDH suffers from substrate inhibition at high concentrations of phenylpyruvate, the α-keto acid was added to the reaction mixture stepwise in several portions to keep its concentration no higher than 50 mM. Thus, the final concentration of L-phenylalanine was 0.3 M (50 g/liter) (Fig. 3C). L-Tyrosine was also synthesized from *p*-hydroxyphenylpyruvate in a similar manner, with a yield of 92%. L-Tyrosine crystallized in the incubation mixture due to its low solubility as the reaction proceeded. The optical purity of L-phenylalanine and L-tyrosine was 100%.

Synthesis of D *amino acids.* D enantiomers of glutamate and leucine were produced at high optical purities and high conversion rates with *E. coli* TG1 carrying pFADA (Table 2). The final concentration of D-glutamate produced was around 0.3 M

TABLE 2. Synthesis of L and D amino acids from α-keto acids with *E. coli* cells carrying pFDHLeuDH2, pFDHAlaDH, pFDHPheDH, or pFADA

Plasmid	Substrate	Product	Yield (%)[a]	Enantiomeric excess (%)[b]
pFDHLeuDH2	α-Ketoisocaproate	L-Leucine	97	100
	α-Ketocaproate	L-Norleucine	95	100
	α-Ketoisovalerate	L-Valine	95	100
	α-Ketovalerate	L-Norvaline	95	100
	α-Ketobutyrate	L-α-Aminobutyrate	88	100
	α-Keto-γ-thiomethylbutyrate	L-Methionine	88	100
pFDHAlaDH	Pyruvate (0.2 M)	L-Alanine	95	99
	Pyruvate (0.4 M)	L-Alanine	92	99
	Pyruvate (0.6 M)	L-Alanine	75	99
pFDHPheDH	Phenylpyruvate	L-Phenylalanine	95	100
	p-Hydroxyphenylpyruvate	L-Tyrosine	92	100
pFADA	α-Ketoglutarate	D-Glutamate	85	100
	α-Ketoisocaproate	D-Leucine	76	100
	α-Ketocaproate	D-Norleucine	70	96
	α-Keto-γ-thiomethylbutyrate	D-Methionine	80	96
	α-Ketoisovalerate	D-Valine	85	95
	α-Ketovalerate	D-Norvaline	90	95
	α-Ketobutyrate	α-Aminobutyrate	95	0
	Phenylpyruvate	D-Phenylalanine	15	ND[c]
	p-Hydroxyphenylpyruvate	D-Tyrosine	5	ND

[a] The yields were determined after 15 h of incubation.
[b] The optical purity was determined by HPLC.
[c] ND, not determined.

(Fig. 3D). Norvaline, valine, and α-aminobutyrate were also produced with high yields. However, α-aminobutyrate was synthesized as a racemic mixture because it is racemized by AlaR. D-Norvaline was obtained at an enantiomeric excess of only 35%. This compound was also racemized by the racemase, although at an extremely low rate. D-Valine, D-methionine, and D-norleucine also suffered from contamination by the antipodes at concentrations of 4, 5, and 6%, respectively. This was

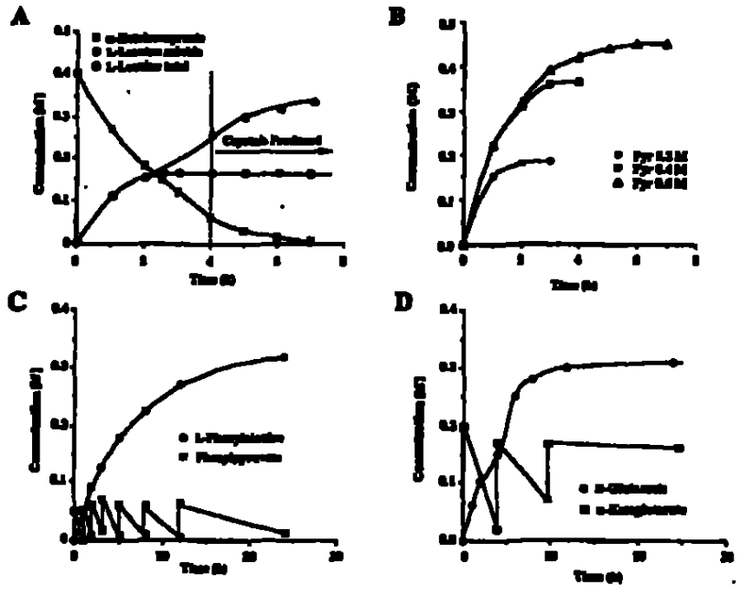


FIG. 3. Time course for the production of L-leucine (A), alanine (B), L-phenylalanine (C), and D-glutamate (D) with *E. coli* cells carrying pFDHLeuDH2 (A), pFDHAlaDH (B), pFDHPheDH (C), and pFADA (D), respectively. (A) After 4 h of incubation (dotted line) crystals of L-leucine appeared in the reaction mixture. Finally, the reaction mixture was solidified with the crystals. (B) The L-alanine produced had an optical purity of approximately 80%. (C) Sodium phenylpyruvate was added to the reaction mixture stepwise. (D) α-Ketoglutarate was added after 4 and 10 h of incubation at a final concentration of approximately 0.2 M.

probably due to the action of AlaDH, because these amino acids are not racemized by AlaR (9). D-Phenylalanine and D-tyrosine, which are poor substrates for DAAT, were synthesized with yields of less than 15%.

DISCUSSION

We show here the usefulness of FDH, DAAT, and AlaR, as well as L-amino acid dehydrogenases, such as LeuDH and PheDH, for the production of optically active amino acids. The simultaneous expression of all enzymes in a single cell allowed the efficient recycling of NADH. We found that the intracellular concentration of NAD^+ is high enough to produce amino acids at final concentrations of approximately 0.3 M. We were not, however, able to obtain production of higher concentrations, probably due to the degradation of the coenzyme in the cells, as has been reported previously (23).

Optically pure L-phenylalanine was synthesized with a yield of 99% with a mixture of acetone-dried cells of *B. sphaericus* and *C. boidinii* as sources of PheDH and FDH, respectively (2). In our study, a similar yield was obtained by using the cells of a single strain, *E. coli* TG1 carrying pFDHPheDH, and the cells were not treated with acetone, which increased the permeability of the cell membrane for the NADH added.

We used resting cells of *E. coli* as the catalyst. However, optically active amino acids were also produced in a growing cell system by using Luria-Bertani or M9 medium (16) supplemented with ammonium formate and α-keto acid (data not shown). Our preliminary experiments showed that α-keto acids, particularly branched-chain and long-chain α-keto acids, are toxic and inhibit the growth of *E. coli* when they are added at concentrations of only 15 to 30 mM. Therefore, the resting cell system is much more convenient than the growing cell system for producing high concentrations of amino acids.

We have shown previously by using the purified-enzyme system (9) that low optical purity of the D amino acids produced is mostly due to the action of AlaDH toward α-keto acids other than pyruvate. Similar results were obtained with the cell systems discussed here. *E. coli* used as a host should have several L-amino acid aminotransferases, which may decrease the optical purity of the D amino acids produced. Although no racemase activity was found in amino acids other than alanine in *E. coli*, the optical purity of the L-alanine produced in our system decreased due to the action of the AlaR of the host cells. Thus, yeasts, which inherently lack AlaR activity, or AlaR-deficient mutant strains of *E. coli* may facilitate enantiospecific synthesis of L-alanine when they are used as the hosts for AlaDH and FDH genes. Optically pure L-β-chloroalanine was synthesized with a 90% yield from β-chloropyruvate with purified preparations of AlaDH and FDH (14). Because AlaR is efficiently inactivated by β-chloroalanine (3), our present system in which *E. coli* TG1/pFDHAlaDH cells are used probably serves as an efficient tool for the production of optically pure L-β-chloroalanine.

As exemplified by the production of leucine enantiomers with 100% optical purity as shown here, our system should be particularly suitable for the synthesis of unnatural L and D amino acids with bulky side chains, such as *tert*-leucine, which are used as building blocks for potent human immunodeficiency virus protease inhibitors, antitumor agents, etc. (5). The L enantiomer of *tert*-leucine has been produced efficiently in a membrane reactor system by using purified LeuDH and FDH, and the D enantiomer has been obtained chemically only by using many complicated steps. Our system should serve as a simple method for the production of D-*tert*-leucine.

REFERENCES

1. Asano, Y., and A. Nakazawa. 1987. High yield synthesis of L-amino acids by phenylalanine dehydrogenase from *Sporosarcina ureae*. Agric. Biol. Chem. 51:2035–2036.
2. Asano, Y., A. Yamada, Y. Kato, K. Yamaguchi, Y. Hibino, K. Hirai, and K. Kondo. 1990. Enantioselective synthesis of (S)-amino acids by phenylalanine dehydrogenase from *Bacillus sphaericus*: use of natural and recombinant enzymes. J. Org. Chem. 55:5567–5571.
3. Badet, B., D. Roise, and C. T. Walsh. 1984. Inactivation of the *dadB Salmonella typhimurium* alanine racemase by D- and L-isomers of β-substituted alanines: kinetics, stoichiometry, active site peptide sequencing, and reaction mechanism. Biochemistry 23:5188–5194.
4. Bailey, J. E. 1991. Toward a science of metabolic engineering. Science 252:1668–1675.
5. Bommarius, A. S., M. Schwarm, K. Stingl, M. Kottenhahn, K. Huthmacher, and K. Drauz. 1995. Synthesis and use of enantiomerically pure tert-leucine. Tetrahedron Asymm. 6:2851–2888.
6. Bückmann, A. F., M.-R. Kula, R. Wichmann, and C. Wandrey. 1981. An efficient synthesis of high-molecular-weight NAD(H) derivatives suitable for continuous operation with coenzyme-dependent enzyme systems. J. Appl. Biochem. 3:301–315.
7. Fitz-Delliston, S., A. S. Gaillard, A. J. Cooper, K. C. Rosenspire, and E. Nieves. 1990. Short-term metabolic fate of L-[^{13}N]glutamate in the Walker 256 carcinosarcoma in vivo. Cancer Res. 50:4639–4644.
8. Galkin, A., L. Kulakova, V. Tishkov, N. Esaki, and K. Soda. 1995. Cloning of formate dehydrogenase gene from a methanol-utilizing bacterium *Mycobacterium vaccae* N10. Appl. Microbiol. Biotechnol. 44:479–483.
9. Galkin, A., L. Kulakova, H. Yamamoto, K. Tanizawa, H. Tanaka, N. Esaki, and K. Soda. 1997. Conversion of α-keto acids to D-amino acids by coupling of four enzyme reactions. J. Ferment. Bioeng. 83:299–300.
10. Hummel, W., and M.-R. Kula. 1989. Dehydrogenases for the synthesis of chiral compounds. Eur. J. Biochem. 184:1–13.
11. Inagaki, K., K. Tanizawa, B. Badet, C. T. Walsh, H. Tanaka, and K. Soda. 1986. Thermostable alanine racemase from *Bacillus stearothermophilus*: molecular cloning of the gene, enzyme purification, and characterization. Biochemistry 25:3268–3274.
12. Ingram, L. O., F. Alterthum, K. Ohta, and D. S. Beall. 1990. Genetic engineering of *Escherichia coli* and other enterobacteria for ethanol production. Dev. Ind. Microbiol. 31:21–30.
13. Kamphuis, J., E. M. Meijer, W. H. J. Boesten, T. Sonke, W. J. J. van den Tweel, and H. E. Schoemaker. 1992. New developments in the synthesis of natural and unnatural amino acids. Ann. N.Y. Acad. Sci. 672:510–527.
14. Kato, Y., K. Fukumoto, and Y. Asano. 1993. Enzymatic synthesis of L-β-chloroalanine using amino acid dehydrogenase. Appl. Microbiol. Biotechnol. 39:301–304.
15. Kuroda, S., K. Tanizawa, Y. Sakamoto, H. Tanaka, and K. Soda. 1990. Alanine dehydrogenases from two *Bacillus* species with distinct thermostabilities: molecular cloning, DNA and protein sequence determination, and structural comparison with other NAD(P)-dependent dehydrogenases. Biochemistry 29:1009–1015.
16. Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
17. Nishimura, K., K. Tanizawa, T. Yoshimura, N. Esaki, S. Futaki, J. M. Manning, and K. Soda. 1991. Effect of substitution of lysyl residue that binds pyridoxal phosphate in thermostable D-amino acid aminotransferase by arginine and alanine. Biochemistry 30:4072–4077.
18. Ohshima, T., C. Wandrey, M.-R. Kula, and K. Soda. 1985. Improvement for L-leucine production in a continuously operated enzyme membrane reactor. Biotechnol. Bioeng. 27:1616–1618.
19. Ohshima, T., and K. Soda. 1989. Biotechnological aspects of amino acid dehydrogenases. Int. Ind. Biotechnol. 9:5–11.
20. Ohshima, T., N. Nishida, S. Bakthavatsalam, K. Kataoka, H. Takada, T. Yoshimura, N. Esaki, and K. Soda. 1994. The purification, characterization, cloning and sequencing of the gene for a halostable and thermostable leucine dehydrogenase from *Thermoactinomyces intermedius*. Eur. J. Biochem. 222:305–312.
21. Roberts, S. M., N. J. Turner, A. J. Willetts, and M. K. Turner. 1995. The interrelationships between enzymes and cells, with particular reference to whole cell biotransformations using bacteria and fungi, p. 34–78. *In* M. R. Stanley (ed.), Introduction to biocatalysis using enzymes and micro-organisms. Cambridge University Press, New York, N.Y.
22. Rozzell, J. D. 1994. Introduction, p. 2. *In* J. D. Rozzell and F. Wagner (ed.), Biocatalytic production of amino acids and derivatives. John Wiley & Sons, Inc., New York, N.Y.
23. Simon, H. 1985. Microbial transformations using growing or resting cells, p. 157–161. *In* K. Drauz and H. Waldmann (ed.), Enzyme catalysis in organic synthesis. VCH Publishers, Inc., New York, N.Y.
24. Takada, H., T. Yoshimura, T. Ohshima, N. Esaki, and K. Soda. 1991. Thermostable phenylalanine dehydrogenase of *Thermoactinomyces intermedius*: cloning, expression, and sequencing of its gene. J. Biochem. 109:371–376.

25. Tsujimura, K., A. Obukhov, A. Schwibbert, K. Inagaki, H. Tanaka, and K. Soda. 1988. Thermostable alanine racemase from *Bacillus stearothermophilus*: DNA and protein sequence determination and secondary structure prediction. Biochemistry 27:1311–1316.
26. Tsujimura, K., S. Asano, Y. Misu, S. Kawashima, M. Kagamiyama, H. Tanaka, and K. Soda. 1988. The primary structure of thermostable D-amino acid aminotransferase from a thermophilic *Bacillus* species and its correlation with L-amino acid aminotransferases. J. Biol. Chem. 264:2453–2454.
27. Tishkov, V. I., A. G. Galkin, V. N. Gladilov, V. V. Kamzolov, and A. M. Egorov. 1992. Analysis of gene structure, optimization of expression in *E. coli* and properties of recombinant formate dehydrogenase of *Ancylobacter aquaticus* sp. 301. Biotechnology (Russia) 8:53–58.
28. Wichmann, R., C. Wandrey, A. F. Buckmann, and M.-R. Kula. 1981. Continuous enzymatic transformation in an enzyme membrane reactor with simultaneous NAD(H) regeneration. Biotechnol. Bioeng. 23:2789–2802.

THE JOURNAL OF BIOLOGICAL CHEMISTRY
© 2001 by The American Society for Biochemistry and Molecular Biology, Inc.
Vol. 276, No. 47, Issue of November 23, pp. 43775–43783, 2001
Printed in U.S.A.

NADP-Glutamate Dehydrogenase Isoenzymes of *Saccharomyces cerevisiae*

PURIFICATION, KINETIC PROPERTIES, AND PHYSIOLOGICAL ROLES*

Received for publication, August 20, 2001, and in revised form, September 17, 2001
Published, JBC Papers in Press, September 18, 2001, DOI 10.1074/jbc.M107966200

Alexander DeLuna‡, Amaranta Avendaño, Lina Riego, and Alicia González§

From the Departamento de Genética Molecular, Instituto de Fisiología Celular, Universidad Nacional Autónoma de México, Apartado Postal 70-242, México D.F. 04510, México

In the yeast *Saccharomyces cerevisiae*, two NADP$^+$-dependent glutamate dehydrogenases (NADP-GDHs) encoded by *GDH1* and *GDH3* catalyze the synthesis of glutamate from ammonium and α-ketoglutarate. The *GDH2*-encoded NAD$^+$-dependent glutamate dehydrogenase degrades glutamate producing ammonium and α-ketoglutarate. Until very recently, it was considered that only one biosynthetic NADP-GDH was present in *S. cerevisiae*. This fact hindered understanding the physiological role of each isoenzyme and the mechanisms involved in α-ketoglutarate channeling for glutamate biosynthesis. In this study, we purified and characterized the *GDH1*- and *GDH3*-encoded NADP-GDHs; they showed different allosteric properties and rates of α-ketoglutarate utilization. Analysis of the relative levels of these proteins revealed that the expression of *GDH1* and *GDH3* is differentially regulated and depends on the nature of the carbon source. Moreover, the physiological study of mutants lacking or overexpressing *GDH1* or *GDH3* suggested that these genes play nonredundant physiological roles. Our results indicate that the coordinated regulation of *GDH1*-, *GDH3*-, and *GDH2*-encoded enzymes results in glutamate biosynthesis and balanced utilization of α-ketoglutarate under fermentative and respiratory conditions. The possible relevance of the duplicated NADP-GDH pathway in the adaptation to facultative metabolism is discussed.

Like most free living microorganisms, the yeast *Saccharomyces cerevisiae* possesses amino acid biosynthetic pathways that allow the cell to use ammonium as sole nitrogen source. Ammonium utilization occurs exclusively via its incorporation into glutamate and glutamine (1), a process that can be achieved by two metabolic routes. One of them is constituted by the concerted action of glutamine synthetase and the *GLT1*-encoded glutamate synthase (2, 3). The other pathway is mediated by the NADP$^+$-dependent glutamate dehydrogenase (NADP-GDH)[1] (EC 1.4.1.4), a broadly distributed enzyme that catalyzes the reductive amination of α-ketoglutarate to form glutamate (4, 5). In *S. cerevisiae*, two genes (*GDH1* and *GDH3*) have been described whose products constitute NADP-GDH isoenzymes (6). Glutamate catabolism is achieved through a reaction catalyzed by a different but related enzyme, the *GDH2*-encoded NAD$^+$-dependent glutamate dehydrogenase (NAD-GDH) (EC 1.4.1.2), which determines glutamate degradation to ammonium and α-ketoglutarate (4, 7, 8).

S. cerevisiae is the first microorganism described in which the NADP-GDH activity is encoded by two genes (6); the physiological significance of this apparent redundancy is not clear. When this yeast is grown on glucose and ammonium as carbon and nitrogen sources, Gdh1p is the primary pathway for glutamate biosynthesis (6, 9, 10). It has also been shown that *GDH1* expression is regulated by the *HAP* system (11), which is known to control expression of genes involved in carbon metabolism and respiratory function (12). Null *gdh3Δ* mutants show no evident growth phenotype on glucose, and *GDH3*-dependent activity is negligible on this carbon source. Nevertheless, a biosynthetic role was established for *GDH3* in a double *gdh1Δ gln1Δ* mutant that grows on ammonium sulfate as sole nitrogen source by means of Gdh3p (6). Moreover, global analysis of transcription suggests that *GDH3* expression is influenced by the general nitrogen control system (13).

S. cerevisiae is able to grow using a variety of carbon sources under fermentative and respiratory conditions. This fact has stimulated discussion as to which specific mechanism allows α-ketoglutarate utilization for glutamate biosynthesis without impairing the integrity of the tricarboxylic acid cycle as an energy-providing system. In this regard, it has been shown that *Klebsiella aerogenes* strains overexpressing their *gdhA* gene coding for the biosynthetic NADP-GDH display an auxotrophy that is interpreted as a limitation for α-ketoglutarate and succinyl-coenzyme A (14). Accordingly, α-ketoglutarate modulates NADP-GDH activity so that fluctuations in the intracellular levels of tricarboxylic acid cycle intermediates would regulate glutamate biosynthesis. Indeed, it has been shown that the signal that coordinately regulates carbon and nitrogen metabolism in *Escherichia coli* depends on the intracellular levels of α-ketoglutarate and glutamine (15). Interestingly, the presence of Gdh3p has been found to be increased during diauxic transition in *S. cerevisiae* (16), suggesting a particular role of this enzyme in respiratory metabolism.

To understand the function of the duplicated NADP-GDH pathway present in *S. cerevisiae*, we purified both isoenzymes and studied their biochemical properties. Our results revealed

* This work was supported in part by Dirección General de Asuntos del Personal Académico, Universidad Nacional Autónoma de México (UNAM), Grant IN212898 and by Consejo Nacional de Ciencia y Tecnología Grant 31774. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ Recipient of a fellowship and a grant (PAEP-102815) from the Dirección General de Estudios de Posgrado, UNAM.

§ To whom correspondence should be addressed. Tel.: 52-56225631; Fax: 52-56225630; E-mail: amanjarr@ifisiol.unam.mx.

[1] The abbreviations used are: NADP-GDH, NADP$^+$-dependent glutamate dehydrogenase; NAD-GDH, NAD$^+$-dependent glutamate dehydrogenase; MM, minimal medium; YPD, yeast-peptone-dextrose; PCR, polymerase chain reaction; TLCK, *N*^{α}-*p*-tosyl-L-lysine chloromethyl ketone; MES, 4-morpholineethanesulfonic acid; bp, base pair(s).

TABLE I
S. cerevisiae strains used in this work

Strain	Relevant genotype	Source
CLA1	*MAT*a *GDH1 GDH3 ura3 leu2*	Ref. 6
CLA4	*MAT*a *GDH1 GDH3 URA3*::Yip5 *LEU2*::Yip351	Ref. 6
CLA6	*MAT*a *gdh1A*::*URA3 GDH3 LEU2*::Yip351	Ref. 6
CLA7	*MAT*a *GDH1 gdh1A*::*LEU2 URA3*::Yip5	Ref. 6
CLA10	*MAT*a *gdh1A*::*URA3 gdh1A*::*LEU2*	Ref. 6
CLA11	*MAT*a *GDH1 GDH3 LEU2*::Yip351 *ura3*	This study
CLA12	*MAT*a *GDH1 gdh1A*::*LEU2 ura3*	This study
CLA13	*MAT*a *gdh1A*::*hamMX4 GDH3 LEU2*::Yip351 *ura3*	This study
CLA14	*MAT*a *gdh1A*::*hamMX4 gdh1A*::*LEU2 ura3*	This study

that Gdh1p and Gdh3p have different allosteric properties and rates of α-ketoglutarate utilization. The construction of chimerical plasmids harboring combinations of the *GDH1* and *GDH3* promoter and coding regions allowed us to determine that expression of these two genes is differentially modulated by the carbon source. Finally, physiological analysis of mutants lacking or overexpressing *GDH1* or *GDH3* showed that expression of both genes is required to achieve wild-type growth on ethanol. Our results indicate that existence of different NADP-GDH isoenzymes allows the functioning of a regulatory system in which the relative abundance of each isoform modulates the rate at which α-ketoglutarate is channeled to glutamate biosynthesis.

EXPERIMENTAL PROCEDURES

Strains

Table I describes the characteristics of the strains used in the present work. All strains constructed for this study were *LEU2* derivatives of CLA1 (*ura3 leu2*) and thus suited for *URA3* selection. To obtain a *gdh1A* mutant, CLA1 was transformed with the *Bgl*II-linearized plasmid pLV8 (6) harboring a 760-bp *GDH3* fragment and the yeast *LEU2* gene, generating strain CLA12 (*GDH1 gdh1A ura3*). A *gdh1A gdh3A ura3* mutant was obtained from CLA12, using the PCR-based gene replacement protocol described by Wach *et al.* (17), with *hamMX4* as a marker. Two deoxyoligonucleotides were designed based on the *GDH1* nucleotide sequence and that of the multiple cloning site present in the pFA6a vector (17). The deoxyoligonucleotide D1 (5′-CAG AAT TTC AAC AAG CTT ACG AAG AAG TTG TCT CCT CTT TGG AAG CGT ACG CTG CAG GTC GAC-3′) comprised 45 bp of the 5′ region of the *GDH1* coding sequence (+11 to +55), and 18 bp (in boldface type) of the pFA6a multiple cloning site. The deoxyoligonucleotide D2 (5′-AAC ACC GAT ATC ACC AGC TGG CAC GTC AGT GTC TTG ACC AAT GTG ATC GAT GAA TTC GAG CTC G-3′) contained 45 bp corresponding to an internal *GDH1* gene fragment (+481 to +495) and 19 bp (in boldface type) from the pFA6a multiple cloning site. Qiagen purified pFA6a DNA was used as template for PCR amplification in a Stratagene Robocycler 40 with the following program: one denaturing cycle for 3-min at 94 °C, followed by 25 cycles of 30-s denaturation at 94 °C, 1-min annealing at 50 °C, and 1-min extension at 72 °C. The 532-bp PCR product obtained was gel-purified and used to transform strain CLA12, generating strain CLA14. A CLA1 *LEU2* derivative was obtained by transforming this strain with plasmid Yip351, generating strain CLA11. To obtain a *gdh1A GDH3 ura3* mutant, the CLA11 strain was transformed with the above mentioned 532-bp PCR product, thus generating CLA13.

Yeast was transformed by the method described by Ito *et al.* (18). Transformants were selected for either leucine prototrophy on minimal medium (MM), or G418 resistance (200 mg/liter) (Life Technologies, Inc.) on yeast extract-peptone-dextrose (YPD)-rich medium.

Growth Conditions

Strains were routinely grown on MM containing salts, trace elements, and vitamins following the formula of yeast nitrogen base (Difco). Filter-sterilized glucose (2%, w/v) or ethanol (2%, v/v) was used as a carbon source, and 40 mM ammonium sulfate was used as a nitrogen source. Supplements needed to satisfy auxotrophic requirements were added at 0.1 mg/ml. Cells were incubated at 30 °C with shaking (250 rpm).

Construction of Low Copy Number and High Copy Number Plasmids Bearing GDH1 or GDH3 Genes

All standard molecular biology techniques were followed as previously described (19). *GDH1* or *GDH3* were PCR-amplified together with their 5′ promoter sequence and cloned into either the pRS316 (*CEN6 ARS4 URA3*) low copy number or pRS426 (2μ ori *URA3*) high copy number yeast shuttle vectors (20, 21). For *GDH1*, the 2256-bp region between −258 from the start codon and +265 from the stop codon was considered to comprise the full *GDH1* promoter and coding sequences (11). For *GDH3*, a 2646-bp fragment was PCR-amplified, containing the putative regulatory region (−1313 from the start codon) plus the full coding sequence and +45 from the end codon, as reported in the nucleotide sequence of chromosome I from *S. cerevisiae* (22). Deoxyoligonucleotides used for this purpose were S1 (5′-CGC GGG ATC CAG TAG TTC AGC GAC AGA AG-3′), S2 (5′-CGC GCG GAT CCC GAG TAA GGT CAT CAA TAA G-3′), S3 (5′-CGC GGG ATC CTG CGG TTA TAT GAT CTT C-3′), and S4 (5′-CGC GCG GAT CCT ACT ACA TAC ACA GAT AG-3′), generating plasmids pLAM1 (*GDH1 CEN URA3*), pLAM11 (*GDH1* 2μ *URA3*), pLAM3 (*GDH3 CEN URA3*), and pLAM33 (*GDH3* 2μ *URA3*). DNA sequencing was carried out, using the T3/T7 priming sites of pRS316 and pRS426, at the Unidad de Biología Molec-

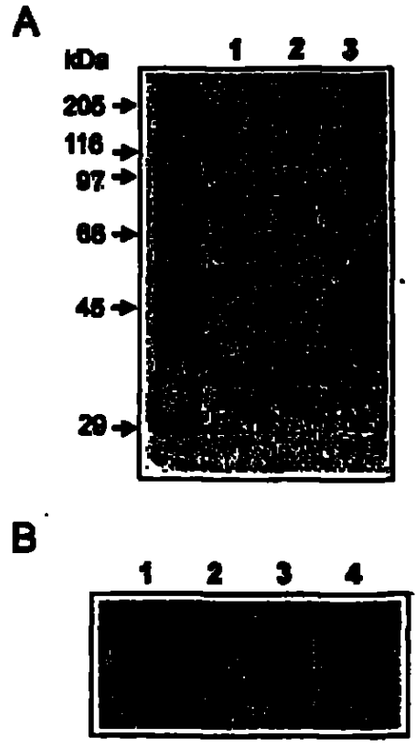


FIG. 1. Purification and electrophoretic characterization of yeast NADP-GDHs. *A*, SDS-PAGE showing proteins purified to electrophoretic homogeneity from ethanol-grown yeast cultures (see "Experimental Procedures" for purification strategy). *Lane 1*, Gdh1p (from CLA14-11); *lane 2*, Gdh3p (from CLA14-22); *lane 3*, wild-type NADP-GDH (from CLA4). *B*, purified proteins (2 μg) were subjected to native gel electrophoresis (6%) and Coomassie-stained. *Lane 1*, Gdh1p (from CLA14-11); *lane 2*, Gdh3p (from CLA14-22); *lane 3*, wild-type NADP-GDH (from CLA4); *lane 4*, Gdh1p plus Gdh3p.

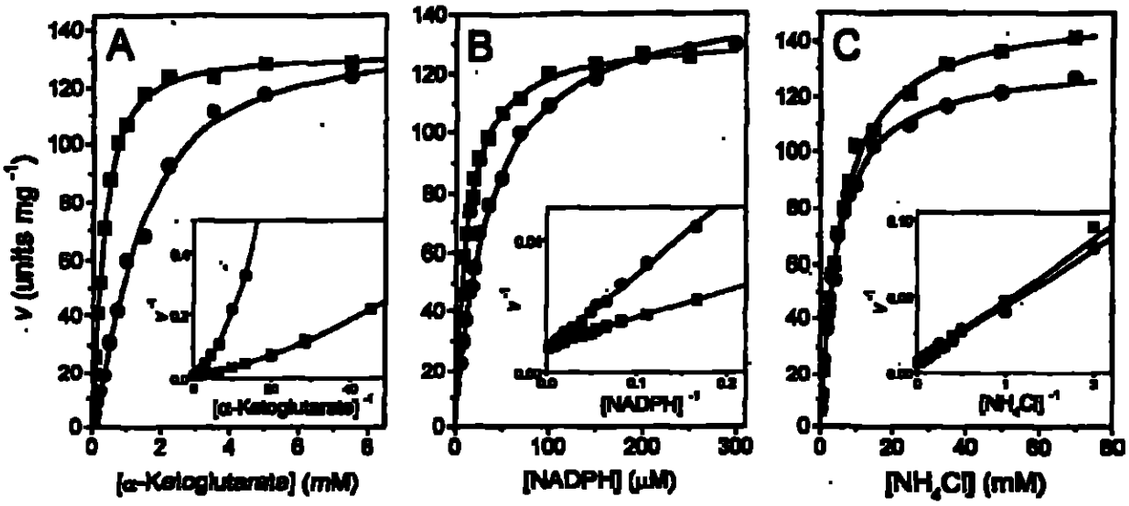


FIG. 2. Comparative kinetic analysis of two NADP-GDH isoenzymes. Plots show the dependence of the relative rate of the reductive amination reaction on the concentration of α-ketoglutarate (*A*), NADPH (*B*), and ammonium (*C*). Reactions were carried out in 100 mM Tris buffer (pH 7.5) at 25 °C (see "Experimental Procedures"). ■, Gdh1p enzyme; ●, Gdh3p enzyme. *Insets* represent double reciprocal plots.

ular, Instituto de Fisiología Celular, Universidad Nacional Autónoma de México (UNAM).

Plasmids were subsequently transformed into CLA14 double mutant, and uracil prototrophs were selected, thus generating strains CLA14-1, CLA14-11, CLA14-2, and CLA14-22. Control strains harboring the 2μ pRS426 plasmid were constructed by transforming CLA11, CLA12, CLA13, and CLA14, generating strains CLA11-00, CLA12-00, CLA13-00, and CLA14-00, respectively.

Construction of GDH1 and GDH3 Chimerical Fusion Plasmids

Fusions containing either the *GDH1* promoter and the *GDH3* coding sequence or the *GDH3* promoter and the *GDH1* coding sequence were generated by overlapping PCR amplification. For this purpose, primers S1 and S6 (5′-CTC TGG TTC GCT TGT CAT TTC TTT TTC TTT TTG G-3′) were used to obtain a 980-bp product corresponding to the *GDH1* 5′ cognate sequence and the first 19 bp of the *GDH3* coding sequence (in boldface type); this was overlapped with the 1491-bp product of primers S4 and S5 (5′-GAC AAG CGA ACC AGA GTT TC-3′), which included the complete *GDH3* coding sequence. Similarly, primers S2 and S3 (5′-GAA ATT CTG GCT CTG AGA TTT TTA CTT TTT ACC-3′) were used to obtain a 1344-bp product corresponding to the *GDH3* 5′ cognate sequence, together with the first 17 bp of the *GDH1* coding sequence (in boldface type), and overlapped with the 1632-bp product of primers S3 and S10 (5′-GTC AGA GCC AGA ATT TCA AC-3′), which included the complete *GDH1* coding sequence. The whole procedure led to the generation of the following plasmids: pLAM3 (5′*GDH3-GDH1* CEN *URA3*), pLAM33 (5′*GDH3-GDH1* 2μ *URA3*), pLAM44 (5′*GDH1-GDH3* CEN *URA3*), and pLAM444 (5′*GDH1-GDH3* 2μ *URA3*). Constructs were verified by DNA sequencing as described above.

Plasmids were subsequently transformed into the CLA14 double mutant, and uracil prototrophs were selected, generating strains CLA14-3, CLA14-33, CLA14-4, and CLA14-44.

NADP-GDH Purification

NADP-GDH activity was purified from ethanol-grown cultures of CLA 14-11 (*gdh1Δ gdh3Δ*/pLAM11 (*GDH1* 2μ *URA3*)), CLA 14-33 (*gdh1Δ gdh3Δ*/pLAM33 (*GDH3* 2μ *URA3*)), and the CLA4 wild-type strain. Strains were grown in 10 liters of MM supplemented with ethanol and ammonium sulfate, in a fermentor at the Unidad de Bioprocesos, Instituto de Investigaciones Biomédicas, UNAM. Cultures were incubated at 30 °C and 300 rpm and aerated with 7 liters of oxygen/min. Cells were harvested at an optical density of 0.8–1.0 at 600 nm and stored at −70 °C until used. NADP-GDH was purified by a modified version of the method of Doherty (33). All steps were carried out at 5 °C.

Step 1: Whole Cell Soluble Protein Extract—Cells were thawed and resuspended in 1 ml of buffer A (100 mM Tris (pH 7.5), 1 mM EDTA, 1 mM dithiothreitol, 1 mM phenylmethylsulfonyl fluoride, and 50 μg of *N*-*p*-tosyl-L-lysine chloromethyl ketone (TLCK)/ml/g of cells. Crude extracts were obtained after mechanical disruption of cells with a Bead-Beater (8 cycles of 1 min). After centrifugation at 20,000 × *g* for 30 min, protein extracts were resuspended in buffer A and diluted to ~25 mg/ml.

Step 2: Ammonium Sulfate Fractionation—Proteins that precipitated between 40 and 60% saturation of ammonium sulfate were resuspended in buffer A. Mixtures were dialyzed twice against 4 liters of buffer B (30 mM Tris (pH 7.5), 1 mM EDTA).

Step 3: DEAE Bio-Gel A Chromatography—Dialyzed fractions were applied to a DEAE Bio-Gel A column (35 by 2.6 cm) equilibrated with buffer C (30 mM Tris (pH 7.5), 1 mM EDTA, 1 mM phenylmethylsulfonyl fluoride, and 50 mg of TLCK/liter). After sample application, the column was washed with 5 column volumes of buffer B. NADP-GDH was subsequently eluted with a linear NaCl gradient of 10 column volumes (0–0.5 M). Fractions with NADP-GDH activity were pooled and dialyzed against 4 liters of buffer B.

Step 4: Affinity Chromatography—A Reactive Red-agarose column (15 by 1.8 cm) was equilibrated with buffer A. After application of the sample from the previous step, the column was washed with 10 volumes of buffer A. NADP-GDH was eluted with buffer A containing 0.1 mM NADPH. Fractions with NADP-GDH activity were pooled, dialyzed against buffer B, concentrated by ultrafiltration to ~1 mg/ml with an Amicon YM30 membrane, and stored at −70 °C until used.

Enzyme Assay and Protein Determination

Whole cell soluble protein extracts were prepared by glass bead lysis of cell pellets harvested during exponential growth, as described (34). NADP-GDH and NAD-GDH were assayed by the method of Doherty (33). One unit of activity is defined as the oxidation of 1.0 μmol of NADPH or NADH/min. Protein was measured by the method of Lowry *et al.* (35), using bovine serum albumin as a standard.

Preparation of Anti-NADP-GDH Antibodies

Antibodies were raised in rabbits injected with purified yeast *GDH1*-encoded NADP-GDH and partially purified by ammonium sulfate precipitation according to the method of Gonzáles-Halphen *et al.* (36).

Electrophoresis and Immunoblotting

SDS-polyacrylamide gel electrophoresis (PAGE) and native PAGE were performed with 10 and 6% slab gels, respectively. Proteins on polyacrylamide gels were visualized with Coomassie Blue. Immunoblot analysis of SDS-electrophoresed crude extract or pure NADP-GDH was carried out as described by Towbin *et al.* (37). Immunoblot signaling was optimized by analyzing a number of combinations of antigen and antibody concentrations in the linear range of detectability. Scanned blots were subjected to densitometric analysis using the program ImageQuaNT 4.2 (Molecular Dynamics, Inc., Sunnyvale, CA). Data were

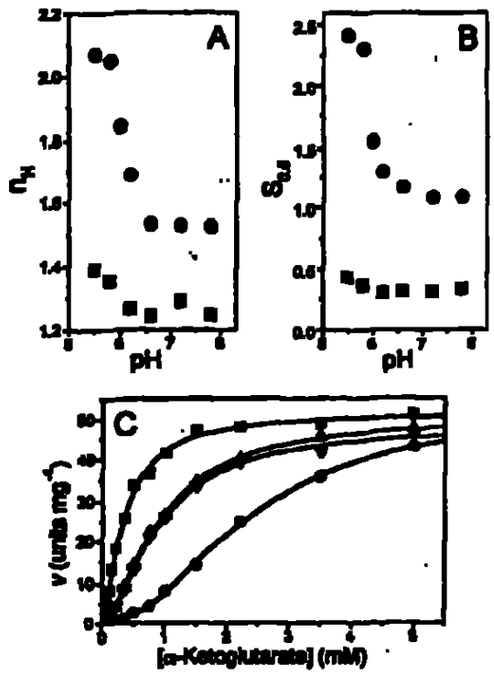


FIG. 3. NADP-GDH isoenzymes show different allosteric properties depending on the concentration of α-ketoglutarate. Plots show the dependence of n_H (*A*) and $S_{0.5}$ (*B*) on the pH of the reaction. *C*, at pH 8.5, the relative rates of the NADP-GDH purified from the wild-type strain depend on the relative abundance of each isoenzyme. Assays were carried out in 25 mM acetic acid, 25 mM MES, 50 mM Tris buffer, at 25 °C. Purified samples used in this experiment were the same as those shown in Fig. 1*B*. ■, Gdh1p enzyme; ●, Gdh3p enzyme; ▲, wild-type NADP-GDH; ▼, Gdh1p plus Gdh3p (2:1 mixture).

normalized to the immunoblot signals of the corresponding purified protein.

Molecular Mass Determination

Native molecular mass was determined on a Sephacryl S-300 gel filtration column (2.6 by 90 cm) equilibrated with 50 mM Tris (pH 7.5), 150 mM NaCl, and 1 mM dithiothreitol. The column was calibrated with molecular mass standards (29–700 kDa) from Sigma. Purified NADP-GDH was diluted in the same buffer, loaded into the column, and eluted at a rate of 6 ml/h. Molecular mass was determined from a plot of the log molecular mass against elution volume per void volume.

The apparent molecular masses of denatured subunits were determined by SDS-PAGE with molecular mass standards (29–205 kDa) from Sigma.

Amino-terminal Sequencing

The isolation of polypeptides for amino-terminal sequencing was carried out as described previously (32). Edman degradation was carried on an Applied Biosystems Sequencer at the Laboratoire de Microséquençage des Protéines (Institut Pasteur, Paris, France).

Enzyme Kinetics and Analysis of Kinetic Data

NADP-GDH activity was assayed for the reductive amination reaction at different concentrations of α-ketoglutarate, NADPH, or ammonium chloride and at saturating concentrations of the remaining substrates (5 mM α-ketoglutarate, 200 μM NADPH, and 50 mM ammonium chloride). For the oxidative deamination reaction, different concentrations of glutamate or $NADP^+$ and saturating concentration of the remaining substrate (100 mM glutamate and 200 μM $NADP^+$) were used. The progress of the reaction was always kept below 5% conversion of the initial substrate. Measurements were made at 25 °C in 100 mM Tris at pH 7.2 or 8.0 for the reductive amination or oxidative deamination reaction, respectively. For experiments in which pH was varied, 25 mM acetic acid, 25 mM MES, 50 mM Tris was used as buffer. This buffer

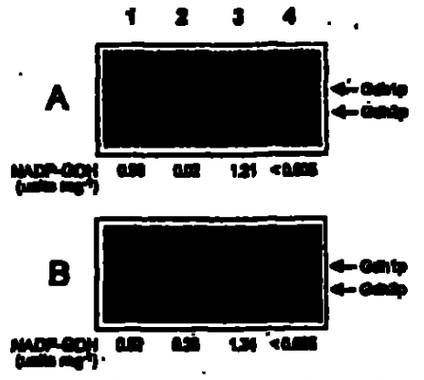


FIG. 4. Levels of Gdh1p and Gdh3p are differentially regulated by carbon source. Cells were grown on MM supplemented with glucose (*A*) or ethanol (*B*) and subjected to immunoblot analysis using Gdh1p antiserum. Cells were harvested during logarithmic growth, and protein extracts were assayed for NADP-GDH activity and electrophoresed (SDS-10% PAGE, 20 μg of protein/lane). *Lane 1*, CLA4 (wild type); *lane 2*, CLA8 (*gdh1Δ*); *lane 3*, CLA7 (*gdh3Δ*); *lane 4*, CLA10 (*gdh1Δ gdh3Δ*).

minimizes the change of ionic strength with pH (26). Kinetic data were analyzed by nonlinear regression using the program Origin 4.1 (MicroCal Software, Inc.).

Extraction and Determination of Intracellular α-Ketoglutarate

Protein-free cell extracts were prepared as described by Kang *et al.* (20). The intracellular concentration of α-ketoglutarate relative to protein concentration was determined with beef glutamate dehydrogenase (Sigma) by following NADH oxidation (21).

Determination of Extracellular Glucose Concentration

Cells were filtered through 0.22-μm Millipore membranes. Extracellular glucose concentration was determined in the filtrate with the Glucose [HK] kit from Sigma.

RESULTS

NADP-GDH Purification from Mutant and Wild-type Strains—*S. cerevisiae* is the first microorganism in which the existence of two NADP-GDH isoenzymes has been reported (6). Although yeast NADP-GDH has been previously purified and characterized (32), the properties described could be ascribed to either or both isoenzymes. Therefore, we purified the Gdh1p and Gdh3p enzymes to electrophoretic homogeneity to study their individual biochemical properties. Gdh1p was 36-fold purified from the CLA14-11 mutant strain harboring plasmid pLAM11, whereas Gdh3p was 49-fold purified from strain CLA14-22 bearing plasmid pLAM22. Additionally, NADP-GDH was 252-fold purified from the wild-type strain CLA4 grown on ethanol, a condition in which both isoenzymes are readily expressed (see below). Apparent molecular masses of the monomers were 51 and 48 kDa for Gdh1p and Gdh3p, respectively (Fig. 1*A*). The observed molecular mass of the latter was at variance with the expected value deduced from its amino acid sequence, which is 49.6 kDa. This suggested the existence of a post-translational modification of Gdh3p, which remains to be identified. Amino-terminal sequencing was not possible, because both Gdh1p and Gdh3p purified polypeptides were blocked.

The active oligomeric structures of the purified samples obtained from the wild-type and mutant strains were hexameric, as revealed by gel filtration experiments (data not shown). This is in agreement with results obtained for all members of the small glutamate dehydrogenase subfamily, which show an α_6 50-kDa oligomeric structure (33) and whose three-dimensional crystal structure has been reported (34, 35). Native PAGE

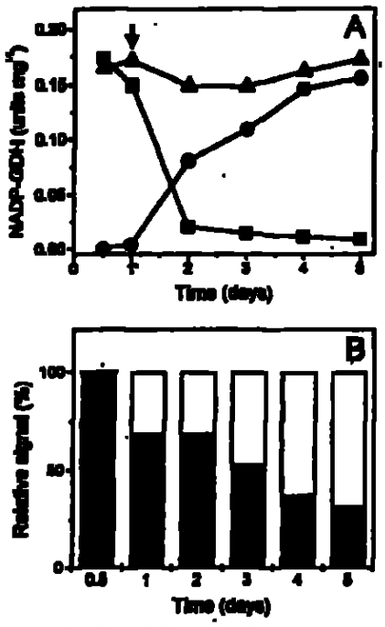


FIG. 5. Relative levels of Gdh1p and Gdh3p change depending on growth phase. *A*, yeast were cultured by extended growth in YPD-rich medium. Aliquots were withdrawn at different times, and protein extracts were assayed for NADP-GDH activity. Extracellular glucose concentration was determined in parallel; a *dark arrow* indicates the time at which glucose was exhausted from the medium. ■, CLA1 (*gdh3Δ*); ●, CLA5 (*gdh1Δ*); ▲, CLA4 (wild type). *B*, the abundance of each isoenzyme relative to the total NADP-GDH of the wild-type strain was calculated from the normalized densitometric analysis of immunoblot signals obtained for electrophoresed protein extracts of strain CLA4. *Black bars*, Gdh1p; *white bars*, Gdh3p.

analysis of the protein purified from the wild-type strain showed a smeared pattern (Fig. 1*B*, *lane 3*), compared with the sharp bands observed when a mixture of equivalent amounts of purified homomeric proteins was electrophoresed (Fig. 1*B*, *lane 4*). Hence, the enzyme purified from the wild-type strain was most probably a natural mixture of several isoforms built up by the oligomerization of the two different monomers encoded by *GDH1* and *GDH3*. In SDS-PAGE electrophoresis, the NADP-GDH purified from the wild-type strain grown on ethanol showed two bands corresponding to Gdh1p and Gdh3p monomers (Fig. 1*A*). Densitometric analysis of Coomassie-stained gels revealed that 73% of the total NADP-GDH was composed of Gdh1p; the remaining 27% corresponded to Gdh3p.

Kinetic Analysis of NADP-GDH Isoenzymes—Enzymological properties were separately determined for the Gdh1p and Gdh3p homomeric NADP-GDHs. Activities were measured in Tris-MES-acetic acid buffer at pH values ranging from 4.5 to 9.5 (data not shown); maximum activity was obtained at pH 6.8 for both enzymes. We examined the dependence of NADP-GDH activity on α-ketoglutarate, NADPH, or ammonium, using saturating concentrations of the two remaining substrates (Fig. 2). Both isoenzymes showed hyperbolic behavior at increasing NADPH and ammonium concentrations but sigmoidal responses to increasing α-ketoglutarate concentrations (Fig. 2*A*).

NADP-GDH isoenzymes showed V_{max} values that were similar in all experiments (120–150 units mg^{-1}). All substrates caused inhibition of enzyme activity above a given threshold concentration (data not shown). NADPH began to inhibit the activity of both enzymes at a concentration of 300 μM (10% inhibition); with 100 mM ammonium chloride, we observed a similar effect. A 5% inhibition of the maximal activity was observed with 10 mM α-ketoglutarate for the Gdh1p enzyme, whereas a 25 mM substrate concentration was needed to generate the same inhibition of the Gdh3p enzyme.

For the Gdh1p enzyme assayed in both directions of the NADP-GDH reaction, the K_m values for NADPH, ammonium, $NADP^+$, and glutamate were 11.3 μM, 5.96 mM, 14.1 μM, and 9.79 mM, respectively. Values of 58.1 μM, 5.00 mM, 10.5 μM, and 6.36 mM, respectively, were obtained for the Gdh3p isoenzyme. Phosphate competitive inhibition on NADPH binding has been previously described for yeast NADP-GDH (36). We confirmed that with respect to NADPH concentration, phosphate competitively inhibited both isoenzymes at various concentrations (0–250 mM sodium phosphate) (data not shown). However, Gdh3p was more sensitive to this effect, with a K_i value for phosphate of 9.2 mM, compared with 72.5 mM for the Gdh1p enzyme.

Differences were also found between the two isoenzymes in their kinetics for α-ketoglutarate. At pH 7.2, substrate concentrations at which rates were equal to half the V_{max} ($S_{0.5}$) were 0.29 and 1.27 mM for the Gdh1p and Gdh3p enzyme, respectively. Hill coefficients (n_H) in the same experiments were 1.3 and 1.5 for the Gdh1p and Gdh3p enzyme, respectively. In this regard, homomeric glutamate dehydrogenases from other organisms are known to be allosteric enzymes activated by different molecules (AMP, ADP, GTP, ATP, $NADP^+$, succinate, aspartate, and asparagine) (33, 37–39). The effect of these compounds was assayed for the yeast NADP-GDH isoenzymes, but none of them behaved as an allosteric effector (data not shown). However, sigmoidal kinetics could most likely reflect a phenomenon of cooperativity, since n_H values strictly depended on the pH at which the kinetics for α-ketoglutarate was assayed (Fig. 2*A*). The n_H plot for the Gdh3p isoenzyme against pH showed an inflection point at pH 6.8. Near optimum pH, Gdh3p exhibited a higher $S_{0.5}$ value compared with its homologue; this difference was higher at low pH (Fig. 2*B*). Conversely, the Gdh1p isoenzyme showed no considerable changes in sigmoidicity and had higher affinity for α-ketoglutarate in terms of $S_{0.5}$. Thus, the overall data indicate that the NADP-GDH isoenzymes differ in their allosteric properties and rates at which they use α-ketoglutarate.

We determined α-ketoglutarate kinetics for the NADP-GDH purified from the wild-type strain and compared them with those of the homomeric Gdh1p and Gdh3p isoenzymes. Since the maximum kinetic differences between the two isoenzymes were observed at pH 5.8, we analyzed the behavior of the wild-type enzyme at this pH. The wild-type enzyme exhibited kinetic parameters ($S_{0.5}$, 0.80 mM; n_H, 1.6) similar to those of a preparation containing 75% Gdh1p and 25% Gdh3p homomeric isoenzymes (Fig. 3*C*). This indicates that kinetics toward α-ketoglutarate depends on the relative abundance of the *GDH1*- and *GDH3*-encoded monomers, whether or not these proteins associate in heteromeric structures.

Relative Levels of Gdh1p and Gdh3p Are Carbon-dependent—To compare the relative levels of the two NADP-GDHs under different conditions, extracts were prepared from the wild-type or the pertinent null mutant strains grown on glucose or ethanol as carbon sources. The specific activities and immunochemically detected levels of Gdh1p were similar in extracts obtained from the *gdh3Δ* strain grown on glucose or ethanol (Fig. 4, *lane 3*). For Gdh3p, low levels of NADP-GDH activity were observed, and no signal in immunoblots could be detected when glucose was the carbon source. However, when extracts were prepared from ethanol-grown cells, Gdh3p enzymatic activity increased 20-fold, and an immunoblot signal was clearly observed (Fig. 4, *lane 2*). Normalized densitometric

TABLE II
NADP-GDH activity of MM-grown strains harboring recombinant plasmids constructed in this study

Strain	Specific activity[a]	
	Glucose	Ethanol
	pmol min^{-1} mg^{-1}	
CLA14-0 (*gdh1Δ gdh3Δ*/pRS316 (CEN *URA3*))	<0.005	<0.005
CLA14-1 (*gdh1Δ gdh3Δ*/pLAM1 (*GDH1* CEN *URA3*))	3.28 ± 0.220	1.26 ± 0.089
CLA14-2 (*gdh1Δ gdh3Δ*/pLAM2 (*GDH3* CEN *URA3*))	0.05 ± 0.007	0.51 ± 0.042
CLA14-3 (*gdh1Δ gdh3Δ*/pLAM3 (5′*GDH3-GDH1* CEN *URA3*))	0.13 ± 0.015	0.97 ± 0.070
CLA14-4 (*gdh1Δ gdh3Δ*/pLAM4 (5′*GDH1-GDH3* CEN *URA3*))	0.60 ± 0.066	0.62 ± 0.021
CLA14-00 (*gdh1Δ gdh3Δ*/pRS426 (2μ *URA3*))	<0.005	<0.005
CLA14-11 (*gdh1Δ gdh3Δ*/pLAM11 (*GDH1* 2μ *URA3*))	8.01 ± 0.85	10.6 ± 1.02
CLA14-22 (*gdh1Δ gdh3Δ*/pLAM22 (*GDH3* 2μ *URA3*))	0.17 ± 0.013	6.05 ± 0.77
CLA14-33 (*gdh1Δ gdh3Δ*/pLAM33 (5′*GDH3-GDH1* 2μ *URA3*))	0.80 ± 0.089	11.4 ± 0.94
CLA14-44 (*gdh1Δ gdh3Δ*/pLAM44 (5′*GDH1-GDH3* 2μ *URA3*))	2.41 ± 0.211	9.90 ± 0.86

[a] Values are presented as means from three independent experiments ± S.D.

TABLE III
Growth phenotypes and glutamate dehydrogenase activities of strains lacking or overexpressing GDH1 or GDH3 in MM

Strain	Relative growth[a]		NADP-GDH Specific activity[b]		NAD-GDH Specific activity[b]	
	Glucose	Ethanol	Glucose	Ethanol	Glucose	Ethanol
	%	%				
CLA11-00 (*GDH1 GDH3*/pRS426 (2μ *URA3*))	100	100	0.884	0.718	0.036	0.055
CLA12-00 (*gdh1Δ GDH3*/pRS426 (2μ *URA3*))	62	95	0.519	0.459	0.033	0.062
CLA13-00 (*GDH1 gdh3Δ*/pRS426 (2μ *URA3*))	105	67	1.84	0.940	0.042	0.158
CLA14-00 (*gdh1Δ gdh3Δ*/pRS426 (2μ *URA3*))	53	45	<0.005	<0.005	0.033	0.041
CLA14-11 (*gdh1Δ gdh3Δ*/pLAM11 (*GDH1* 2μ *URA3*))	92	81	10.5	8.88	0.038	0.160
CLA14-12 (*gdh1Δ gdh3Δ*/pLAM12 (*GDH3* 2μ *URA3*))	75	89	0.894	7.97	0.035	0.283

[a] Values are shown relative to doubling time of the wild-type strain (2.3 and 5.5 h on glucose and ethanol, respectively) and are presented as means from three independent experiments (variation was always ≤10%).
[b] Values are given in pmol/min/mg of protein and are presented as means from three independent experiments (variation was always ≤10%).

analysis of immunoblots showed that 25% of the wild-type NADP-GDH from ethanol-grown yeasts corresponded to Gdh3p.

In light of the previous results, it was relevant to determine whether NADP-GDHs containing different Gdh1p/Gdh3p ratios could be found in long term yeast cultures. In YPD-rich medium, *S. cerevisiae* grows by fermentation; diauxic shift occurs after glucose is exhausted from the medium and cells adapt to respiratory metabolism using the ethanol produced during glucose fermentation (40). In fermentative growth, with glucose as the only carbon source, NADP-GDH activity was solely due to Gdh1p (Fig. 5, *A* and *B*). However, as cells proceeded through postdiauxic growth, different Gdh1p/Gdh3p ratios were observed, and after 5 days of incubation, 70% of the total NADP-GDH activity in the wild-type strain corresponded to Gdh3p (Fig. 5*B*). Within this context, it is relevant that NADP-GDH proteolysis has been observed after glucose starvation (41); this could account for the specific inactivation of Gdh1p after glucose was exhausted from the medium. Taken together, these results indicate that the relative abundance of Gdh1p or Gdh3p depends on the carbon source.

GDH3 Expression Is Transcriptionally Regulated by the Nature of the Carbon Source—To determine whether *GDH3* carbon-dependent regulation was exerted at the transcriptional level, several recombinant plasmids were constructed (see "Experimental Procedures"). NADP-GDH activities were determined for strains derived from the CLA14 mutant strain transformed with these plasmids (Table II). Cells bearing low copy number constructs showed differences in enzymatic activity, which could be mainly attributed to the different levels of expression allowed by the cognate 5′ promoter sequences of either *GDH1* or *GDH3*. In glucose-grown cells, Gdh1p-dependent NADP-GDH activity was 27-fold higher when expressed from its own promoter as compared with that fostered by the 5′*GDH3-GDH1* fusion. Likewise, Gdh3p-dependent activity was 20-fold higher when this gene was under the regulation of the *GDH1* promoter sequences (5′*GDH1-GDH3*) than when expressed from its cognate promoter. Similar results were obtained using cells harboring high copy number plasmids. When NADP-GDH activity was monitored in extracts obtained from ethanol-grown cells, high levels were observed for either Gdh1p or Gdh3p, regardless of which promoter fostered expression. These results confirmed that *GDH3* expression was repressed by glucose at the transcriptional level.

It is worth mentioning that in extracts prepared from glucose-grown cultures, Gdh1p activity was at least 5-fold higher than that of Gdh3p when the genes were expressed from either promoter; this effect was barely observed in ethanol (Table II). Considering that V_{max} values are similar for both isoenzymes, this differential level of expression could be attributed to a post-transcriptional level of regulation. In fact, it could be considered that the codon bias difference of these genes (0.75 and 0.19 for *GDH1* and *GDH3*, respectively) may account for different translation rates of their transcripts.

NADP-GDH Isoenzymes Modulate α-Ketoglutarate Utilization for Glutamate Biosynthesis—Gdh1p enzyme is the primary pathway for glutamate biosynthesis in glucose-grown cells (6). Double *gdh1Δ gdh3Δ* mutants lacking NADP-GDH activity are not full glutamate auxotrophs; this strain grows with a 2-fold higher doubling time than that observed in the wild-type strain. This growth is achieved through the action of the *GLT1*-encoded glutamate synthase, which constitutes an ancillary pathway for glutamate biosynthesis (42). *GDH3* expressed from a high copy number plasmid conferred only a partial recovery of the slow growth phenotype of a *gdh1Δ gdh3Δ* strain (Table III), as expected from the observed repression of the *GDH3* gene by glucose. Conversely, *GDH1* expressed from a high copy number plasmid completely restored wild-type growth to a *gdh1Δ gdh3Δ* strain.

It is relevant that in cells grown on ethanol, single disruptions of either *GDH1* or *GDH3* resulted in a slower growth with respect to the wild-type strain. Furthermore, the *gdh1Δ gdh3Δ*

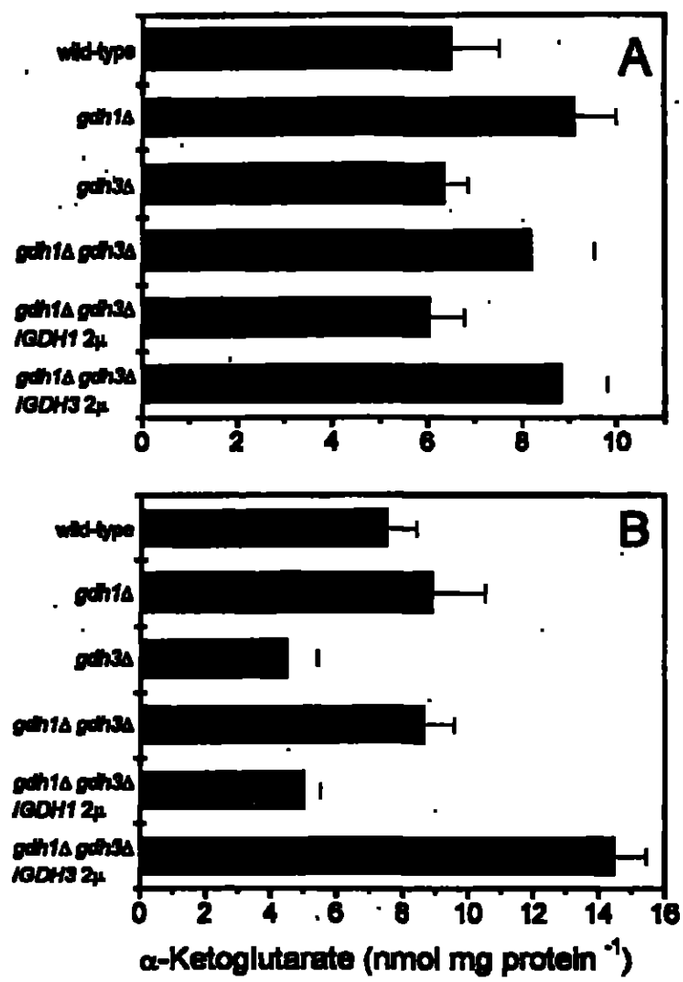


FIG. 6. *In vivo*, NADP-GDHs consume α-ketoglutarate at different relative velocities. Yeast cells were grown on MM supplemented with glucose (*A*) or ethanol (*B*) and harvested during logarithmic growth. Protein-free extracts and soluble protein extracts were prepared as described under "Experimental Procedures." Values of intracellular α-ketoglutarate relative to protein concentration are presented as means from three independent experiments ± S.D. Strains used were CLA11-00 (wild-type), CLA13-00 (*gdh1Δ*), CLA12-00 (*gdh3Δ*), CLA14-00 (*gdh1Δ gdh3Δ*), CLA14-11 (*gdh1Δ gdh3Δ*/pLAM11 (*GDH1* 2μ)), and CLA14-22 (*gdh1Δ gdh3Δ*/pLAM22 (*GDH3* 2μ)).

double mutant strain overexpressing *GDH1* from a plasmid grew considerably slower on ethanol than the one bearing the *GDH3* high copy number construct. Thus, it can be concluded that wild-type growth on ethanol depends on both Gdh1p and Gdh3p and that overexpression of *GDH1* could result in a deleterious effect.

Because of the differences in the rates of α-ketoglutarate *in vitro* utilization by the NADP-GDH isoenzymes, we explored if these differences could be observed *in vivo*. To this end, we measured α-ketoglutarate intracellular pools in cells lacking or overexpressing *GDH1* or *GDH3*. We also determined the NAD-GDH-specific activities in the various strains, since this catabolic enzyme would be expected to increase α-ketoglutarate concentration. In yeast cells grown on glucose, the only evident phenotype was due to the lack of *GDH1*; either single (*gdh1Δ*) or double (*gdh1Δ gdh3Δ*) mutants exhibited a significant accumulation of α-ketoglutarate. A lack of *GDH3* did not affect either the intracellular concentration of this intermediate or NAD-GDH activity (Fig. 6*A*, Table III). These results are in consonance with the growth phenotypes observed for the same strains on glucose. When *GDH1* was overexpressed, NAD-GDH activity exhibited a 2-fold increase, suggesting that this activity increased as a result of glutamate accumulation (3, 43). As expected, *GDH3* overexpression did not result in increased NAD-GDH activity (Table III).

When grown on ethanol, the single *gdh1Δ* mutant did not show a net increase in α-ketoglutarate concentration, whereas the *gdh3Δ* single mutant exhibited a 2-fold lower α-ketoglutarate pool size as compared with that of the wild-type strain. However, the *gdh1Δ gdh3Δ* strain had α-ketoglutarate levels similar to those found in the *gdh1Δ* mutant; this suggested that the α-ketoglutarate depletion observed in a *gdh3Δ* mutant was due to a Gdh1p-dependent consumption of this compound in the absence of the Gdh3p enzyme (Fig. 6*B*). These results are in agreement with the fact that Gdh1p enzyme has a higher rate of α-ketoglutarate utilization than the heteromeric enzyme that exists in ethanol-grown cells. Moreover, NAD-GDH specific activity was induced 3-fold in ethanol-grown cells lacking the Gdh3p enzyme (Table III), indicating that under this condition glutamate accumulated, resulting in induced *GDH2* expression (43).

Overexpression of either *GDH1* or *GDH3* in ethanol-grown cells caused an increase in the specific activity of NAD-GDH. However, the effect on α-ketoglutarate concentration was contrasting. Cells overexpressing *GDH1* showed a reduced α-ketoglutarate pool size, whereas *GDH3* high copy number expres-

sion caused its accumulation. Thus, it can be concluded that *GDH1* overexpression causes a drain of the intracellular α-ketoglutarate pool, suggesting that *in vivo* Gdh1p uses this compound at a higher rate than Gdh3p.

DISCUSSION

This study addresses the question of whether *GDH1* and *GDH3* play overlapping or distinct roles and whether these roles are involved in the inherent capacity of *S. cerevisiae* to grow under fermentative or respiratory conditions. The results presented in this paper indicate that the existence of different NADP-GDH isoforms results in glutamate biosynthesis and balanced α-ketoglutarate utilization. The main observations that support this assertion are the following: (*a*) NADP-GDHs showed differences in their allosteric properties and rates of α-ketoglutarate utilization; (*b*) the relative abundance of both isoenzymes depended on the nature of the carbon source; (*c*) a *gdh3*Δ mutant grew slowly on ethanol, although it had wild-type NADP-GDH activity levels (this mutant showed reduced α-ketoglutarate pools and high activity levels of the catabolic NAD-GDH, indicating an abnormal high glutamate production rate); and (*d*) *GDH1* overexpression from a plasmid did not suppress slow growth or the reduced α-ketoglutarate pool phenotypes of a *gdh1*Δ *gdh3*Δ strain; in contrast, overexpression of *GDH3* resulted in faster growth and α-ketoglutarate accumulation.

It has been recently shown that the regulated expression of yeast tricarboxylic acid cycle genes is governed by two transcriptional complexes that function alternatively, depending on the integrity of the respiratory function (44). The *HAP* system regulates the expression of genes that lead to the synthesis of α-ketoglutarate during respiratory metabolism (12), whereas expression of these genes is controlled by the *RTG* system when respiratory function is dampened or lost. This model considers that glutamate plays a central role by repressing *RTG*-dependent expression of genes leading to α-ketoglutarate (44), thus indicating that NADP-GDH activity should be controlled accordingly. A yeast NADP-GDH activity was previously purified (32) at a time when the existence of two isoenzymes was not yet recognized; thus, the kinetic properties and regulation of each isoenzyme could not possibly be discerned. In this study, purification and independent characterization of Gdh1p and Gdh3p enzymes shows that yeast possesses NADP-GDH isoforms that differ in their biochemical properties.

Even after the two NADP-GDHs were recognized, induction of *GDH3* could not be observed in genome-wide transcription analysis of ethanol-grown yeast, probably because of detectability limitations (45, 46). The results presented here differ from those mentioned above and show unequivocally that *GDH3* expression is ethanol-induced and glucose-repressed and that *GDH1* expression is high on both carbon sources. This brings into accountability the role of the different NADP-GDH isoenzymes in either glucose or ethanol-grown cells. Our results also consider the allosteric regulation of the *GDH3*-encoded enzyme, which suggests particular regulatory properties for this activity *in vivo*. This would mediate a more relaxed distribution of α-ketoglutarate to either glutamate biosynthesis or energy-yielding metabolism when cells grow on a nonfermentable or limiting carbon source. During fermentative growth, glutamate biosynthesis would be afforded by the Gdh1p isoenzyme that uses α-ketoglutarate at a faster rate. Accordingly, the existence of multiple isoforms of NADP-GDH activity would provide the pacemaker mechanism that assures optimum glutamate biosynthesis in either fermentative or respiratory conditions without compromising the energy-yielding metabolism. Within this context, it is relevant that the nonfacultative yeast *Kluyveromyces lactis*, closely related to *S. cerevisiae*, bears a single homomeric NADP-GDH enzyme (47).

It has been recognized that the expression of the NAD-GDH catabolic enzyme is induced in the presence of ethanol (48). However, the physiological significance of this observation has remained obscure, since *gdh2*Δ mutants show no evident phenotype in ethanol-grown cultures. Considering the results presented in this paper, it can be suggested that the coordinated action of *GDH1*-, *GDH3*-, and *GDH2*-encoded enzymes allows growth on ethanol, equilibrating the production and utilization of α-ketoglutarate. This study further confirms that nitrogen and carbon metabolisms are coordinately modulated for ammonium assimilation (11, 48) and that the genetic and metabolic regulation of genes involved in nitrogen metabolism can be influenced by the nature of the carbon source.

Finally and worth mentioning is the existence of other duplicated yeast genes, such as *COX5A/COX5B*, *HYP2/ANB1*, *CYC1/CYC7*, and *AAC2/AAC3*, whose regulation has diverged and which are differentially expressed under aerobic or anaerobic conditions (49). Thus, the described duplication and further diversification of an NADP-GDH gene may be representative of a general mechanism through which *S. cerevisiae* acquired facultative metabolic properties (50).

Acknowledgments—We thank C. Aranda and M. Vázquez-Acevedo for skillful technical assistance and X. Aguirre, who worked on the construction of recombinant plasmids. We are grateful to A. Jimenez (Unidad de Escalamiento, Instituto de Investigaciones Biomédicas, UNAM) for assistance in growing yeast cultures in the fermentor; L. Ongay, G. Codiz, and M. Sosa (Unidad de Biología Molecular, Instituto de Fisiología Celular, UNAM) for DNA sequencing and synthesis of oligonucleotides; and J. d'Alayer (Institut Pasteur) for amino-terminal peptide sequencing. We are indebted to M. M. Altamirano, F. Bastarrachea, D. G. Fraenkel, D. González-Halphen, and L. Valenzuela for helpful discussions and critical review of the manuscript. We appreciate the encouraging and insightful comments of A. Gómez-Puyou during this work.

REFERENCES

1. Magasanik, B. (1992) in *The Molecular and Cellular Biology of the Yeast Saccharomyces cerevisiae: Gene Expression* (Jones, E. W., Pringle, J. R., and Broach, J. R., ed) pp. 283–317, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
2. Tempest, D. W., Meers, J. L., Brown, C. M. (1970) *Biochem. J.* 117, 405–407
3. Roon, R. J., Even, H. L., and Larimore, F. (1974) *J. Bacteriol.* 118, 89–95
4. Holzer, H., and Schneider, S. (1957) *Biochem. Z.* 329, 361–369
5. Benachenhou-Lahfa, N., Forterre, P., and Labedan, B. (1993) *J. Mol. Evol.* 36, 335–346
6. Avendaño, A., DeLuna, A., Olivera, H., Valenzuela, L., and Gonzalez, A. (1997) *J. Bacteriol.* 179, 5594–5597
7. Smith, E. L. (1975) in *The Enzymes*, Vol. XI (Boyer, P. D., ed) pp. 293–367, Academic Press, Inc., New York
8. Miller, S. M., and Magasanik, B. (1990) *J. Bacteriol.* 172, 4927–4935
9. Dubois, E., and Lacroute, F. (1973) *J. Bacteriol.* 109, [illegible]
10. Grenson, M., Dubois, E., Piotrowska, M., Drillien, R., and Aigle, M. (1974) *Mol. Gen. Genet.* 128, 73–85
11. Dang, V. D., Bohn, C., Bolotin-Fukuhara, M., and Daignan-Fornier, B. (1996) *J. Bacteriol.* 178, 1842–1849
12. Forsburg, S. L., and Guarente, L. (1989) *Annu. Rev. Cell Biol.* 5, 153–180
13. Cox, K. H., Pinchak, A. B., and Cooper, T. G. (1999) *Yeast* 15, 703–713
14. James, B. K., Pompeiano, P. J., Perez-Matos, A., Majeskas, D. J., Gass, T. J., and Bender, R. A. (2001) *J. Bacteriol.* 183, 5709–5714
15. Marks, A., Jiang, P., Atkinson, M. R., and Pelletier, J. A. (2000) *Curr. Top. Cell Reg.* 36, 31–75
16. Roy-Mercier, E., Poirot, M., Bussereau, F., Boucherie, H., and Jacquet, M. (1996) *J. Bacteriol.* 180, 1044–1052
17. Wach, A., Brachat, A., Pöhlmann, R., and Philippsen, P. (1994) *Yeast* 10, 1793–1808
18. Ito, H., Fukuda, Y., Murata, K., and Kimura, A. (1983) *J. Bacteriol.* 153, 163–168
19. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd Ed., pp. 1.1–7.87, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
20. Sikorski, R. S., and Hieter, P. (1989) *Genetics* 122, 19–27
21. Christianson, T. W., Sikorski, R. S., Dante, M., Shero, J. H., Hieter, P. (1992) *Gene (Amst.)* 110, 119–122
22. Bussey, H., Kaback, D. B., Zhong, W., Vo, D. T., Clark, M. W., Fortin, N., Hall, J., Ouellette, B. F., Keng, T., Barton, A. B., Su, Y., Davies, C. J., and Storms, R. K. (1995) *Proc. Natl. Acad. Sci. U. S. A.* 92, 3809–3813
23. Doherty, D. (1970) *Methods Enzymol.* 17, 850–856
24. Coschigano, C., Valenzuela, L., González-Halphen, D., Olivera, H., Mendoza, G., Ballario, P., and González, A. (1995) *J. Bacteriol.* 177, 783–786

25. Lowry, O. H., Rosebrough, N. J., Farr, A. L., and Randall, R. J. (1951) *J. Biol. Chem.* 193, 265–275
26. González-Halphen, D., Lindorfer, M. A., Capaldi, R. A. (1988) *Biochemistry* 27, 7021–7031
27. Towbin, H., Staehelin, T., Gordon, J. (1979) *Bio/Technology* 24, 145–149
28. Gutiérrez-Cirlos, E. B., Antaramian, A., Vázquez-Acevedo, M., Coria, R., and González-Halphen, D. (1994) *J. Biol. Chem.* 269, 9147–9154
29. Ellis, K. J., Morrison, J. F. (1982) *Methods Enzymol.* 87, 405–426
30. Kang, L., Keeler, M. L., Dunlap, P. C., and Rose, R. J. (1988) *J. Bacteriol.* 151, [illegible]–[illegible]
31. Dubois, E., Grenson, M., and Wiame, J. M. (1974) *Eur. J. Biochem.* 48, [illegible]–[illegible]
32. Camardella, L., Di Prisco, G., Garofano, F., and Guerrini, A. M. (1976) *Biochim. Biophys. Acta* 429, 324–330
33. Mittendorf, R., Olivera, E. R., Jensen, R. A., and Luengo, J. M. (2000) *J. Biol. Chem.* 275, [illegible]–[illegible]
34. Baker, P. J., Britton, K. L., Engel, P. C., Farrants, G. W., Lilley, K. S., Rice, D. W., and Stillman, T. J. (1992) *Proteins* 12, 75–86
35. Yip, K. S., Stillman, T. J., Britton, K. L., Artymiuk, P. J., Baker, P. J., Sedelnikova, S. E., Engel, P. C., Pasquo, A., Chiaraluce, R., Consalvi, V., Scandurra, R., and Rice, D. W. (1995) *Structure* 3, 1147–1158
36. Veronese, F., Julien, J. M., Fernandez, A., Iwatsubo, M. (1975) *Eur. J. Biochem.* 57, 373–378
37. Veronese, F. M., Boccu, E., and Conventi, L. (1975) *Biochim. Biophys. Acta* 377, 217–228
38. Loeza, L., Baron, T., Ty, N. G., Nguyen, L. D., Baron, O., and Mantsch, H. H. (1977) *Biochim. Biophys. Acta* [illegible], [illegible]–[illegible]
39. Bonete, M. J., Pérez-Pomares, F., Ferrer, J., and Camacho, M. L. (1996) *Biochim. Biophys. Acta* 1289, 14–24
40. Page, R. K., Brown R. L., Werner-Washburne, M. (1994) *J. Bacteriol.* 176, [illegible]–[illegible]
41. Menge, M. J., and Hemmings, B. A. (1978) *J. Bacteriol.* 133, [illegible]–[illegible]
42. Valenzuela, L., Ballario, P., Aranda, C., Filetici, P., and González, A. (1998) *J. Bacteriol.* 180, 3533–3540
43. Cunningham, P. V., Miller, S. M., and Magasanik, B. (1991) *Mol. Cell. Biol.* 11, 4455–4465
44. Liu, Z., and Butow, R. A. (1999) *Mol. Cell. Biol.* 19, 6720–6728
45. Shamji, A., Kuruvilla, F. G., and Schreiber, S. L. (2000) *Curr. Biol.* 10, 1574–1581
46. Kuruvilla, F. G., Shamji, A. F., and Schreiber, S. L. (2001) *Proc. Natl. Acad. Sci. U. S. A.* 98, 7283–7288
47. Romero, M., Guzmán-León, S., Aranda, C., González-Halphen, D., Valenzuela, L., and González, A. (2000) *Microbiology* 146, 239–245
48. ter Schure, E. G., van Riel, N. A., and Verrips, T. (2000) *FEMS Microbiol. Rev.* 24, 67–83
49. Kwast, K. E., Burke, P. V., and Poyton, R. O. (1998) *J. Exp. Biol.* 201, 1177–1195
50. Piškur, J. (2001) *Trends Genet.* 17, 302–303

PATENT COOPERATION TREATY

From the INTERNATIONAL BUREAU

PCT

NOTIFICATION CONCERNING SUBMISSION OR TRANSMITTAL OF PRIORITY DOCUMENT

(PCT Administrative Instructions, Section 411)

To:

SKELTON, Jeffrey, J.
Cargill, Incorporated
P.O. Box 5624
Minneapolis, MN 55440-5624
United States of America

Date of mailing (day/month/year) 28 August 2003 (28.08.03)	
Applicant's or agent's file reference 6682-64349	IMPORTANT NOTIFICATION
International application No. PCT/US03/12588	International filing date (day/month/year) 23 April 2003 (23.04.03)
International publication date (day/month/year) Not yet published	Priority date (day/month/year) 23 April 2002 (23.04.02)
Applicant CARGILL, INCORPORATED	

1. The applicant is hereby notified of the date of receipt (except where the letters "NR" appear in the right-hand column) by the International Bureau of the priority document(s) relating to the earlier application(s) indicated below. Unless otherwise indicated by an asterisk appearing next to a date of receipt, or by the letters "NR", in the right-hand column, the priority document concerned was submitted or transmitted to the International Bureau in compliance with Rule 17.1(a) or (b).

2. This updates and replaces any previously issued notification concerning submission or transmittal of priority documents.

3. An asterisk(*) appearing next to a date of receipt, in the right-hand column, denotes a priority document submitted or transmitted to the International Bureau but not in compliance with Rule 17.1(a) or (b). In such a case, the attention of the applicant is directed to Rule 17.1(c) which provides that no designated Office may disregard the priority claim concerned before giving the applicant an opportunity, upon entry into the national phase, to furnish the priority document within a time limit which is reasonable under the circumstances.

4. The letters "NR" appearing in the right-hand column denote a priority document which was not received by the International Bureau or which the applicant did not request the receiving Office to prepare and transmit to the International Bureau, as provided by Rule 17.1(a) or (b), respectively. In such a case, the attention of the applicant is directed to Rule 17.1(c) which provides that no designated Office may disregard the priority claim concerned before giving the applicant an opportunity, upon entry into the national phase, to furnish the priority document within a time limit which is reasonable under the circumstances.

Priority date	Priority application No.	Country or regional Office or PCT receiving Office	Date of receipt of priority document
23 April 2002 (23.04.02)	60/374,831	US	05 June 2003 (05.06.03)

The International Bureau of WIPO 34, chemin des Colombettes 1211 Geneva 20, Switzerland	Authorized officer Ghislaine BORNET
Facsimile No. (41-22) 338.71.40	Telephone No. (41-22) 338 9525

Form PCT/IB/304 (July 1998) 000020765

383

PCT/US03/12588

PATENT COOPERATION TREATY

From the INTERNATIONAL BUREAU

PCT

NOTICE INFORMING THE APPLICANT OF THE COMMUNICATION OF THE INTERNATIONAL APPLICATION TO THE DESIGNATED OFFICES

(PCT Rule 47.1(c), first sentence)

To:

NOV 25 2003

SKELTON, Jeffrey, J.
Cargill, Incorporated
P.O. Box 5624
Minneapolis, MN 55440-5624
ETATS-UNIS D'AMERIQUE

DTP
MSE
CZH
MAP

Date of mailing (day/month/year)	
29 October 2003 (29.10.03)	
Applicant's or agent's file reference: 6582-64348 10790-040001	IMPORTANT NOTICE

International application No.	International filing date (day/month/year)	Priority date (day/month/year)
PCT/US03/12588	23 April 2003 (23.04.03)	23 April 2002 (23.04.02)

Applicant: CARGILL, INCORPORATED

1. Notice is hereby given that the International Bureau has communicated, as provided in Article 20, the international application to the following designated Offices on the date indicated above as the date of mailing of this notice:

   AU, AZ, BY, CH, CN, CO, DE, DZ, HU, JP, KG, KP, KR, MD, MK, MZ, RU, TM

   In accordance with Rule 47.1(c), third sentence, these Offices will accept the present notice as conclusive evidence that the communication of the international application has duly taken place on the date of mailing indicated above and no copy of the international application is required to be furnished by the applicant to the designated Office(s).

2. The following designated Offices have waived the requirement for such a communication at this time:

   AE, AG, AL, AM, AP, AT, BA, BB, BG, BR, BZ, CA, CR, CU, CZ, DK, DM, EA, EC, EE, EP, ES, FI, GB, GD, GE, GH, GM, HR, ID, IL, IN, IS, KE, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MG, MN, MW, MX, NI, NO, NZ, OA, OM, PH, PL, PT, RO, SC, SD, SE, SG, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VC, VN, YU, ZA, ZM, ZW

   The communication will be made in these Offices only upon their request. Furthermore, these Offices do not require the applicant to furnish a copy of the international application (Rule 49.1(a-bis)).

3. Enclosed with this notice is a copy of the international application as published by the International Bureau on 06 November 2003 (06.11.03) under No. WO 03/091 105

4. TIME LIMITS for filing a demand for international preliminary examination and for entry into the national phase

   The applicable time limit for entering the national phase will, subject to what is said in the following paragraph, be 30 MONTHS from the priority date, not only in respect of any elected Office if a demand for international preliminary examination is filed before the expiration of 19 months from the priority date, but also in respect of any designated Office, in the absence of filing of such demand, where Article 22(1) as modified with effect from 1 April 2002 applies in respect of that designated Office. For further details, see *PCT Gazette* No. 44/2001 of 1 November 2001, pages 19905, 19932 and 19934, as well as the *PCT Newsletter*, October and November 2001 and February 2002 issues.

   In practice, time limits other than the 30-month time limit will continue to apply, for various periods of time, in respect of certain designated or elected Offices. For regular updates on the applicable time limits (20, 21, 30 or 31 months, or other time limit), Offices by Office, refer to the *PCT Gazette*, the *PCT Newsletter* and the *PCT Applicant's Guide*, Volume II, National Chapters, all available from WIPO's Internet site, at http://www.wipo.int/pct/en/texts/index.html.

   For filing a demand for international preliminary examination, see the *PCT Applicant's Guide*, Volume I/A, Chapter IX. Only an applicant who is a national or resident of a PCT Contracting State which is bound by Chapter II has the right to file a demand for international preliminary examination (at present, all PCT Contracting States are bound by Chapter II).

   It is the applicant's sole responsibility to monitor all these time limits.

* No Docketing Required *
Reviewed By Practice Systems
Initials:
Reviewed By Billing Secretary
Initials:

The International Bureau of WIPO 34, chemin des Colombettes 1211 Geneva 20, Switzerland	Authorized officer: Judith Zahra
Facsimile No.(41-22) 740.14.35	Telephone No.(41-22) 338.91.11

Form PCT/IB/308 (April 2002)

389

PCT/US2003/012588

DDP
MSE
C24
MMG

PATENT COOPERATION TREATY

From the INTERNATIONAL BUREAU

PCT JAN 29 2004

INFORMATION CONCERNING ELECTED OFFICES NOTIFIED OF THEIR ELECTION

(PCT Rule 61.3)

To:

FISH & RICHARDSON P.C., P.A.
60 South Sixth Street
Suite 3300
Minneapolis, MN 55402
United States of America

Date of mailing (day/month/year)	
14 January 2004 (14.01.2004)	
Applicant's or agent's file reference CGL01/0145WO1 10790-040WO1	**IMPORTANT INFORMATION**
International application No. PCT/US2003/012588	**International filing date (day/month/year)** 23 April 2003 (23.04.2003)
Priority date (day/month/year) 23 April 2002 (23.04.2002)	
Applicant CARGILL, INCORPORATED	

1. The applicant is hereby informed that the International Bureau has, according to Article 31(7), notified each of the following Offices of its election:

   EP :AT,BE,BG,CH,CY,CZ,DE,DK,EE,ES,FI,FR,GB,GR,HU,IE,IT,LU,MC,NL,PT,RO,SE,
   SI,SK,TR
   National :BG,CA,CN,DE,GB,IL,JP,KP,KR,MN,NO,PL,RO,RU,SK

2. The following Offices have waived the requirement for the notification of their election; the notification will be sent to them by the International Bureau only upon their request:

   AP :GH,GM,KE,LS,MW,MZ,SD,SL,SZ,TZ,UG,ZM,ZW
   EA :AM,AZ,BY,KG,KZ,MD,RU,TJ,TM
   OA :BF,BJ,CF,CG,CI,CM,GA,GN,GQ,GW,ML,MR,NE,SN,TD,TG
   National :AE,AG,AL,AM,AT,AU,AZ,BA,BB,BR,BY,BZ,CH,CO,CR,CU,CZ,DK,DM,DZ,EC,
   EE,ES,FI,GD,GE,GH,GM,HR,HU,ID,IN,IS,KE,KG,KZ,LC,LK,LR,LS,LT,LU,LV,MA,MD,
   MG,MK,MW,MX,MZ,NI,NZ,OM,PH,PT,SC,SD,SE,SG,SL,TJ,TM,TN,TR,TT,TZ,UA,UG,UZ,
   VC,VN,YU,ZA,ZM,ZW

3. The applicant is reminded that he must enter the "national phase" before the expiration of 30 months from the priority date before each of the Offices listed above. This must be done by paying the national fee(s) and furnishing , if prescribed, a translation of the international application (Article 39(1)(a)), as well as, where applicable, by furnishing a translation of any annexes of the international preliminary examination report (Article 36(3)(b) and Rule 74.1).

   Some offices have fixed time limits expiring later than the above-mentioned time limit. For detailed information about the applicable time limits and the acts to be performed upon entry into the national phase before a particular Office, see Volume II of the PCT Applicant's Guide.

   The entry into the European regional phase is postponed until 31 months from the priority date for all States designated for the purposes of obtaining a European patent.

"No Docketing Required"
Reviewed By Practice Systems
Initials:
Reviewed By Billing Secretary
Initials:

The International Bureau of WIPO 34, chemin des Colombettes 1211 Geneva 20, Switzerland	Authorized officer: Nadine GANDY (Fax 338-87-20)
Facsimile No. (41-22) 338.71.40	Telephone No. (41-22) 338 8679

Form PCT/IB/332 (September 1997) 6048437

PATENT COOPERATION TREATY

From the
INTERNATIONAL PRELIMINARY EXAMINING AUTHORITY

To:
EDWARD L. LEVINE
CARGILL, INCORPORATED
P.O.BOX 5624
MINNEAPOLIS, MINNESOTA 55440-5624

PCT

NOTIFICATION OF RECEIPT
OF DEMAND BY COMPETENT INTERNATIONAL
PRELIMINARY EXAMINING AUTHORITY

(PCT Rules 59.3(e) and 61.1(b), first sentence
and Administrative Instructions, Section 601(a))

Date of mailing (day/month/year)	16 DEC 2003
Applicant's or agent's file reference: 6882-64349	IMPORTANT NOTIFICATION

International application No.	International filing date (day/month/year)	Priority date (day/month/year)
PCT/US03/12588	23 Apr 2003	23 Apr 2002

Applicant
CARGILL, INCORPORATED

1. The applicant is hereby notified that this International Preliminary Examining Authority considers the following date as the date of receipt of the demand for international preliminary examination of the international application:

21 NOV 2003

2. That date of receipt is:

- [x] the actual date of receipt of the demand by this Authority (Rule 61.1(b)).
- [ ] the actual date of receipt of the demand on behalf of this Authority (Rule 59.3(c)).
- [ ] the date on which this Authority has, in response to the invitation to correct defects in the demand (Form PCT/IPEA/404), received the required corrections.

3. [ ] ATTENTION: That date of receipt is AFTER the expiration of 19 months from the priority date. Consequently, the election(s) made in the demand does (do) not have the effect of postponing the entry into the national phase until 30 months from the priority date (or later in some Offices) (Article 39(1)). Therefore, the acts for entry into the national phase must be performed within 20 months from the priority date (or later in some Offices) (Article 22). For details, see the *PCT Applicant's Guide*, Volume II.

[ ] *(If applicable)* This notification confirms the information given by telephone, facsimile transmission or in person on:

4. Only where paragraph 3 applies, a copy of this notification has been sent to the International Bureau.

Name and mailing address of the IPEA/	Authorized officer
Mail Stop PCT, Commissioner for Patents P.O. Box 1450, Alexandria, VA 22313-1450	Sonya Barnes
Facsimile No. 703-305-3230	Telephone No. 703-305-3665

Form PCT/IPEA/402 (July 1998)

PCT/US03/12588

PATENT COOPERATION TREATY

From the INTERNATIONAL BUREAU

PCT

NOTIFICATION OF THE RECORDING OF A CHANGE

(PCT Rule 92bis.1 and Administrative Instructions, Section 422)

To:

FISH & RICHARDSON P.C., P.A.
60 South Sixth Street
Suite 3300
Minneapolis, MN 55402
United States of America

Date of mailing (day/month/year): 03 December 2003 (03.12.03)	
Applicant's or agent's file reference: CGL01/0145WO1 10790-040W01	IMPORTANT NOTIFICATION
International application No.: PCT/US03/12588	International filing date (day/month/year): 23 April 2003 (23.04.03)

1. The following indications appeared on record concerning:

☐ the applicant ☐ the inventor ☒ the agent ☐ the common representative

Name and Address	State of Nationality	State of Residence
SKELTON, Jeffrey, J. Cargill, Incorporated P.O. Box 5624 Minneapolis, MN 55440-5624 United States of America		
	Telephone No.: 952.742.5402	
	Facsimile No.: 952.742.6349	
	Teleprinter No.:	

2. The International Bureau hereby notifies the applicant that the following change has been recorded concerning:

☒ the person ☐ the name ☐ the address ☐ the nationality ☐ the residence

Name and Address	State of Nationality	State of Residence
FISH & RICHARDSON P.C., P.A. 60 South Sixth Street Suite 3300 Minneapolis, MN 55402 United States of America		
	Telephone No.: (612) 335-5070	
	Facsimile No.: (612) 288-9696	
	Teleprinter No.:	

3. Further observations, if necessary:
Please also note that the agent's file reference has been changed accordingly.

4. A copy of this notification has been sent to:

☒ the receiving Office
☐ the International Searching Authority
☐ the International Preliminary Examining Authority
☒ the designated Offices concerned
☐ the elected Offices concerned
☒ other: SKELTON, Jeffrey, J.

The International Bureau of WIPO 34, chemin des Colombettes 1211 Geneva 20, Switzerland	Authorized officer: François BAECHLER
Facsimile No. (41-22) 338.71.40	Telephone No. (41-22) 338 9544

Form PCT/IB/306 (March 1994) 005997253

387

SUPERINDUSTRIA Y COMERCIO
Radicacion : 04117741 [illegible]
Fecha (dia): 2024-11-23 17:12:51
Tramite : 011 PCT-PATENT 379 FASE NACI 411 PRESENTA
Dependencia: ES20 DIVISION DE NUEVAS CREACIONES
Folios: 384

REQUISITOS NECESARIOS PARA PRESENTACION Y AGILIZACION DEL TRAMITE

	USUARIO	RADICADOR
PATENTE DE INVENCIÓN ✓ MODELO DE UTILIDAD	( )	(✓)
- Indicación de que se solicita la concesión de una patente.	( )	(✓)
- Datos de identificación del solicitante o de la persona que se presenta la solicitud.	( )	(✓)
- Descripción de la invención.	( )	(✓)
- Dibujos o demás datos pertinentes ✓	( )	(✓)
- Comprobante de Pago de las tasas establecidas.		
COMPLETA ( ) INCOMPLETA	( )	( )
DISEÑO INDUSTRIAL ( )		
- Indicación de que se solicita el registro de Diseño Industrial	( )	( )
- Datos de identificación del solicitante o de la persona que presenta	( )	( )
- la solicitud.		
- Representación gráfica o fotográfica del Diseño Industrial o muestra del Material que incorpora el diseño.	( )	( )
Comprobante de pago de las tasas establecidas	( )	( )
COMPLETA ( ) INCOMPLETA	( )	( )
ESQUEMA DE TRAZADO ( )		
- Indicación de que solicita el registro de un Esquema de trazado.	( )	( )
- Datos de identificacion del solicitante o de la persona que presenta La solicitud.	( )	( )
Representación gráfica del esquema de trazado	( )	( )
Comprobante de pago de las tasas establecidas.	( )	( )
COMPLETA ( ) INCOMPLETA	( )	( )

Firma Responsable: Claudia M Neva

Fecha:

388

Industria y Comercio
SUPERINTENDENCIA

2020

Bogotá D C

Doctor(a) A
JUAN PABLO CADENA SARMIENTO
Apoderado(a)
CARGILL, INCORPORATED

11481

REFERENCIA:	Radicación N°	**04-117741**
	Solicitud internacional	**PCT/US03/12588**
	Fecha inicio fase nacional	**2004-11-23**
	Trámite	011
	Evento	379
	Actuación	430
	Folios	001

Como resultado del examen de forma, realizado con fundamento en el artículo 38 de la Decisión 486 de la Comisión de la Comunidad Andina, artículo 27 del Tratado de Cooperación en Materia de Patentes (PCT), regla 51 bis del Reglamento y Circular Única de la Superintendencia de Industria y Comercio, se le comunica que:

- La solicitud no contiene el poder debidamente otorgado, con el lleno de requisitos exigidos por los artículos 63 y subsiguientes del Código de Procedimiento Civil. Adicionalmente deberá presentar el documento que acredita la existencia y representación legal de la sociedad solicitante, cuando éste fuere una persona jurídica.
- La solicitud no contiene copia del documento en que consta la cesión del derecho de patente del inventor al solicitante o a su causante (Art. 26-k) Decisión 486).
- En caso que el examen preliminar contenga anexos estos deben ser adjuntados junto con su traducción.

En cumplimiento de lo dispuesto por el artículo 39 de la Decisión 486, se le requiere para que dentro del término de dos meses siguientes a la fecha de notificación del presente oficio, complete los requisitos anteriormente enunciados. En caso de no dar cumplimiento al presente requerimiento, la solicitud se considerará abandonada y perderá su prelación.

Notifíquese el presente oficio conforme a lo dispuesto en el numeral 5.2, literal e), del capítulo quinto del título primero de la Circular Única N° 10 de 2001.

ALIX CARMENZA CESPEDES DE VERGEL
Jefe de la División de Nuevas Creaciones.

El anterior requerimiento fue notificado por fijación en lista de fecha,
202063 17 DIC. 2004