

**SUPERINTENDENCIA DE INDUSTRIA Y COMERCIO
DIRECCIÓN DE NUEVAS CREACIONES**

2020
Bogotá, D.C.

Abogado
Juan Pablo Cadena
bricas@bc.com.co

SUPERINTENDENCIA DE INDUSTRIA Y COMERCIO



No. 09-010531- -00007-0000

Fecha: 2012-08-09 09:23:05 Dep. 2020 DIR.NUEVASCR
Tra. 11 PATENTEIYII Eve: 378 FASENACIONALI
Act. 519 TRASLADOINFORME Folios: 1

ASUNTO	Radicación	09-10531
	Trámite	011
	Evento	378
	Actuación	519 E 13982
	Oficio	
	Folios	016

Patente de Invención ☒ Patente de Modelo de Utilidad ☐

Vía Nacional ☐

Vía PCT (fase nacional) ☒ Cap. I ☒ Cap. II ☐

No. Sol. Internacional PCT/US2007/015762
Pub. Internacional WO 2008/008358

Como resultado del examen de fondo, realizado con fundamento en el artículo 45 de la Decisión 486 de la Comisión de la Comunidad Andina, se le comunica:

1. Documentos estudiados

<u>Descripción:</u>	Folios: 135 a 228	04/02/2009
<u>Reivindicaciones:</u>	Folios: 229 a 234	04/02/2009
<u>Figuras:</u>	Folios: 235 a 270	04/02/2009
<u>Secuencias:</u>	Folios: 271 a 431	04/02/2009
<u>Prioridad:</u>	US 60/819,569	10/07/2006

2. Análisis de la Prioridad

Se acepta la reivindicación de Prioridad porque esta solicitud fue presentada dentro del plazo establecido (Art. 9 D 486) y su contenido técnico corresponde con el de la prioridad reivindicada.

3. Objeto de la invención

Título de la solicitud: "Composiciones anti-cocaína y tratamiento".

El objeto de la presente solicitud es suministrar mutantes de cocaína esterasa eficientes, termoestables, que puedan proteger contra los efectos tóxicos y reforzantes de la cocaína.

Expediente 09-10531 1er Art.45

El problema planteado en esta solicitud es la necesidad de obtener un antagonista efectivo de la cocaína que pudiera emplearse en situaciones de abuso y sobredosis.

Para solucionar este problema técnico la solicitud en estudio suministra un polipéptido de cocaína esterasa (CocE) mutante aislado, con termoestabilidad incrementada a 37°C comparado con la CocE tipo silvestre.

El objeto de la solicitud definida anteriormente se clasificó en la IPC: A61K 38/46.

4. Excepción a la Patentabilidad (*Artículo 20 D486 literal d*)

El objeto de las reivindicaciones 21 y 22 es un método de tratamiento y no es patentable de acuerdo con el Art citado.

5. De la solicitud de Patente

5.1. Descripción (*artículo 28 D486*)

El capítulo descriptivo no tiene información suficiente para ser reproducida por un técnico medio, en lo referente a la actividad y termoestabilidad de la enzima CocE con todas las sustituciones reclamadas. Tampoco se presenta información sobre la actividad y termoestabilidad cuando las sustituciones están presentes de manera combinada.

De esta manera, la descripción no cumple con los requisitos establecidos en el artículo 28 de la D486.

5.2. Reivindicaciones (*artículo 30 D 486*)

El capítulo reivindicatorio presentado carece de claridad y concisión por cuanto:

- 1) El polipéptido de cocaína esterasa mutante esta definido partir del resultado a alcanzar con el mismo, no se aportan las características esenciales (estructurales) que permitan establecer como el polipéptido tiene actividad esterasa con termoestabilidad incrementada a 37°C, como logra la inmunogenicidad reducida, y como conserva o reduce la actividad esterasa (reivindicaciones 1, 11-14), lo mismo aplica para el ácido nucleico (reivindicación 8);
- 2) Es necesario definir de manera precisa las sustituciones que presenta la enzima reclamada, de manera que sea evidente su diferencia con la enzima silvestre (y no patentable), su estructura y su actividad. La definición a partir de número de sustituciones no permite establecer la estructura reclamada (reivindicaciones 2-5);
- 3) Tal como se mencionó en el numeral 5.2., hay una inconsistencia entre la materia divulgada y la reclamada puesto que no se aporta información sobre la actividad y termoestabilidad de las múltiples secuencias con sustitución y las combinaciones de estas (reivindicación 6 y 19);
- 4) El ácido nucleico reclamada debe ser definido por su propia secuencia nucleótida, sin hacer referencia a porcentajes de identidad que no permiten establecer la estructura de la molécula ni comprobar su actividad (reivindicación 7, 10, 13);

- 5) Las composiciones reclamadas deben ser caracterizadas a partir de cada uno de los elementos que las componen y las proporciones de los mismos (reivindicación 15, 18);
- 6) No es clara la indicación que el polipéptido se encapsula en un glóbulo rojo, se sugiere definir mejor esta característica y aclarar el sustento de la misma;
- 7) El método de detección selectiva esta definido a partir de etapas generales aplicadas para identificar cualquier CocE que cumpla con una condición experimental a probar. Sin embargo, dichos ensayos se consideran experimentos rutinarios, donde la definición de un parámetro tal como la temperatura no guarda relación con la molécula reclamada;

De esta manera, el capítulo reivindicatorio presentado no cumple con los requisitos establecidos en el art. 30 de la Decisión 486.

6. Determinación del Estado de la Técnica

Para determinar el estado de la técnica se ha tenido en cuenta la fecha de presentación de las solicitudes prioritarias reivindicadas con la solicitud: 10/07/2006.

Teniendo en cuenta la fecha indicada se realizó búsqueda en el estado de la técnica de documentos relacionados con la solicitud, encontrándose los siguientes documentos relevantes a la misma:

Nº	Documento	Título	Fecha de publicación
D1	Bresler M., et.al. Appl. Environ. Microbiol. Vol 66 (3): 904-908	"Gene Cloning and Nucleotide Sequencing and Properties of a Cocaine Esterase from <i>Rhodococcus</i> sp. Strain MB1"	Marzo 2000
D2	Larsen N., et.al. Nature Structural Biology, vol 9(1): 17-21	"Crystal structure of a bacterial cocaine esterase"	Enero 2002
D3	Lehman, et.al. Current Opinion in Biotechnology 2001, 12:371-375	"Engineering proteins for thermostability: the use of sequence alignments versus rational design and directed evolution"	2001
D4	Marshall S. et.al. Drug Discovery today. Vol. 8(5):212-221	"Rational design and engineering of therapeutic proteins"	Marzo 2003

Resumen de los documentos citados:

D1: Divulga la secuencia nucleótida y aminoácida, así como las propiedades físico-químicas de la cocaína esterasa de *Rhodococcus* sp., tipo MB1 (resumen, folio XXX).

D2: Se reporta la primera estructura de una enzima que degrada cocaína, la esterasa bacteriana *cocE*, hidroliza cocaína farmacológicamente activa un metabolito no-psicoactivo, con una tasa más rápida que otras esterases de cocaína conocidas (resumen, folio xxx).

D3: Se divulgan técnicas de evolución dirigida para la ingeniería de proteínas para obtener termoestabilidad y ampliar así el uso industrial de las proteínas recombinantes (resumen, folio XXX).

D4: Se divulgan técnicas empleadas para optimizar proteínas terapéuticas, indicando entre otras técnicas el desarrollo de mutaciones específicas, patrones de fusión o modificaciones químicas que confieren las propiedades deseadas (resumen, folio xxx).

7. Análisis de patentabilidad (artículo 14 D486)

Nivel inventivo (artículo 18 D486)

La reivindicación 1 de la solicitud reclama un polipéptido de cocaína esterasa (*CocE*) mutante aislado que comprende una secuencia de aminoácidos de SEQ N°1, en donde por lo menos un residuo de aminoácido se sustituye de tal modo que el polipéptido *CocE* mutante tenga actividad esterasa con termoestabilidad incrementada a 37°C comparada con la *CocE* tipo silvestre.

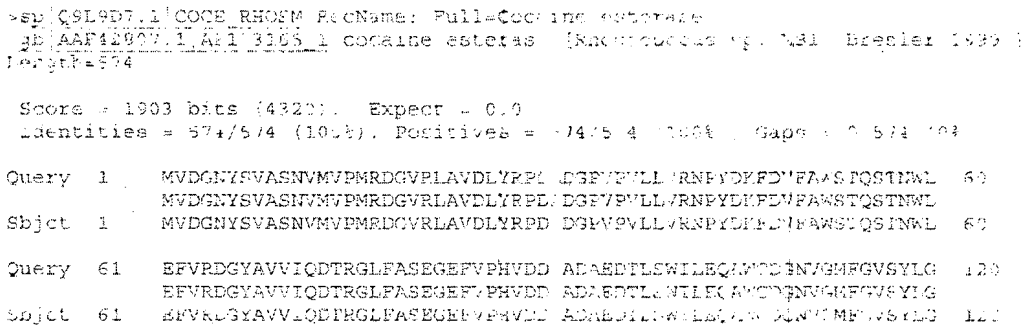
La reivindicación 6 reclama varias opciones de mutantes y la reivindicación 7 reclama el ácido nucleico de la enzima *cocE* mutante.

Comparación de las características técnicas **esenciales**, con las del Estado de la Técnica:

Características esenciales	D1	D3
polipéptido de cocaína esterasa mutante	polipéptido de cocaína esterasa silvestre (resumen)	Polipéptidos recombinantes termoestables
secuencia de aminoácidos SEQ N°1	secuencia de aminoácidos AF173165 100% idéntica a la SEQ N°1 (Pág 905, col der; Fig. 2)	
por lo menos un residuo de aminoácido se sustituye de tal modo que el polipéptido <i>CocE</i> mutante tenga actividad esterasa con termoestabilidad incrementada a 37°C	Polipéptido <i>cocE</i> con actividad a 30°C (Pág. 905 col der párr. 3)	técnicas de diseño computacional/racional (Pág. 371), para obtener proteínas recombinantes termoestables

El documento D1 divulga la secuencia nucleótida y aminoácida de la enzima de la cocaína esterasa de *Rhodococcus sp.* MB1, referenciando su acceso a Genbank como AF173165 (Pág. 905 col. der). Esta proteína es 100% idéntica a la divulgada en la SEQ N°1 de la solicitud (ver Fig.1). Se indica que *Rhodococcus* es un microorganismo aislado de la rizosfera de plantas de coca, e indica la actividad de la enzima en la degradación de cocaína (resumen) y su utilidad en los tratamientos para desintoxicación con cocaína y en la detección de drogas (Pág. 908 col izq). La actividad de la enzima asilada fue ensayada a 30°C, pH 7.5 (Pág. 905 col der, párrafo 3).

Fig. 1. Alineamiento entre las esterases de cocaína *cocE* de la solicitud y D1. Query: secuencia SEQ N°1 solicitud; Subject: Genbank Access: AF173165de D1.



Query	121	VTQWQAAVSCVGGIKAIAPSMASADLYRAPWYGGGALSVEALLGWSALITGTGLITSRSD	180
		VTQWQAAVSCVGGIKAIAPSMASADLYRAPWYGGGALSVEALLGWSALITGTGLITSPSD	
Sbjct	121	VTQWQAAVSCVGGIKAIAPSMASADLYRAPWYGGGALSVEALLGWSALITGTGLITSPSL	180
Query	181	ARPEDAADFVQLAAIILNDVAGAASVTPLAEQPLLGRLLIPWIDQVVDHFDNDESWQSISL	240
		ARPEDAADFVQLAAIILNDVAGAASVTPLAEQPLLGRLLIPWIDQVVDHFDNDESWQSISL	
Sbjct	181	ARPEDAADFVQLAAIILNDVAGAASVTPLAEQPLLGRLLIPWIDQVVDHFDNDESWQSISL	240
Query	241	FERLGGLATPALITAGWYDGFVGESLRTFVAVKDNADARLVVGPWCHSNLTGRNADRKFG	300
		FERLGGLATPALITAGWYDGFVGESLRTFVAVKDNADARLVVGPWCHSNLTGRNADRKFG	
Sbjct	241	FERLGGLATPALITAGWYDGFVGESLRTFVAVKDNADARLVVGPWCHSNLTGRNADRKFG	300
Query	301	IAATYPIQEATTMHKAPFDRHLRGETDALAGVPKVRFLFVMGIDEWRDETDFWFLPDTAYTP	360
		IAATYPIQEATTMHKAPFDRHLRGETDALAGVPKVRFLFVMGIDEWRDETDFWFLPDTAYTP	
Sbjct	301	IAATYPIQEATTMHKAPFDRHLRGETDALAGVPKVRFLFVMGIDEWRDETDFWFLPDTAYTP	360
Query	361	FYLGSGGAANTSTGGGTLSTISGTESADTYLYDPADPVPSLGGTLLFHNGDNGPADQRP	420
		FYLGSGGAANTSTGGGTLSTISGTESADTYLYDPADPVPSLGGTLLFHNGDNGPADQRP	
Sbjct	361	FYLGSGGAANTSTGGGTLSTISGTESADTYLYDPADPVPSLGGTLLFHNGDNGPADQRP	420
Query	421	IHDRDDVLCYSTEVITDPVEVTGTVSARLFVSSSAVDTDFTAKLVDFVFPDGRAIALCDGI	480
		IHDRDDVLCYSTEVITDPVEVTGTVSARLFVSSSAVDTDFTAKLVDFVFPDGRAIALCDGI	
Sbjct	421	IHDRDDVLCYSTEVITDPVEVTGTVSARLFVSSSAVDTDFTAKLVDFVFPDGRAIALCDGI	480
Query	481	VRMRYRETLVNPTLIEAGEIYEVAIDMLATSNVFLGHRIMVQVSSSNFFPYDRNSNTGG	540
		VRMRYRETLVNPTLIEAGEIYEVAIDMLATSNVFLGHRIMVQVSSSNFFPYDRNSNTGG	
Sbjct	481	VRMRYRETLVNPTLIEAGEIYEVAIDMLATSNVFLGHRIMVQVSSSNFFPYDRNSNTGG	540
Query	541	VIAREQLEEMCTAVNRIHRGPEHPHSHIVLPITKR	574
		VIAREQLEEMCTAVNRIHRGPEHPHSHIVLPITKR	
Sbjct	541	VIAREQLEEMCTAVNRIHRGPEHPHSHIVLPITKR	574

La diferencia entre la materia reclamada y D1 es que la enzima reclamada en la solicitud presenta termoestabilidad a 37°C, mientras que la enzima de D1 de acuerdo con lo reportado trabaja a 30°C.

El efecto que produce esta diferencia es que al ser la enzima termoestable a 37°C podría ser empleada en condiciones fisiológicas humanas.

El problema que resuelve la solicitud es producir la enzima de la cocaína esterasa de D1 con actividad a 37°C.

Sin embargo, la forma de producir proteínas termoestables ya había sido divulgada en D3, donde se menciona entre otras técnicas de diseño computacional/racional (Pág. 371), donde se indican tanto algunas mutaciones como los sitios óptimos de acuerdo con la estructura de la proteína, que pueden dar lugar a una mayor termoestabilidad.

De acuerdo con la descripción de la solicitud, este método de diseño computacional/racional fue el empleado para obtener los mutantes a probar para mayor termoestabilidad (ejemplos 4 y 15). Se debe resaltar que a partir de las indicaciones dadas en D3, varias de las opciones de esterasa mutante resultaban evidentes como opción de proteína termoestable, afectando así la reivindicación 1 y 6 de la solicitud).

Iguualmente, habiendo divulgado D1 la secuencia nucleótida de la esterasa cocE y teniendo las opciones de aminoácidos preferentes a modificar, divulgados en D3, la secuencia nucleótida de la esterasa mutante podía ser desarrollada por la persona del oficio normalmente versada en la materia.

De esta manera, frente a las divulgaciones realizadas por D1 y D3 para la persona del oficio normalmente versada en la materia la invención reclamada en la solicitud resultaba obvia y por tanto, incumple el artículo 16 de la D486.

Por otra parte, el documento D2 había divulgado la estructura cristalográfica de la esterasa de cocaína cocE. Identificando los dominios conservados, los sitios de

unión de la enzima con la molécula de cocaína y los sitios de interacción entre ellas. Además realiza la comparación de la enzima cocE con otras esterasas de cocaína. (Pág 17-18; Fig. 3).

La diferencia entre la materia reclamada y D2 es que la enzima reclamada en la solicitud presenta termoestabilidad a 37°C, mientras que la enzima de D2 de acuerdo con lo reportado trabaja a 30°C.

El efecto que produce esta diferencia es que al ser la enzima termoestable a 37°C podría ser empleada en condiciones fisiológicas humanas.

El problema que resuelve la solicitud es producir la enzima de la cocaína esterasa de D1 con actividad a 37°C.

Sin embargo, el mismo documento D2 plantea que debe investigarse en la toxicidad de la enzima y la vida media en plasma de la misma (Pág. 20 col der), sugiriendo así que la enzima debía modificarse.

El documento D4 divulga numerosas técnicas empleadas para mejorar las propiedades de proteínas terapéuticas, indicando las modificaciones físico-químicas realizadas y su efecto en la proteína. Entre estas técnicas resulta relevante las mutaciones puntuales sugeridas para incrementar entre otras la vida media de la proteína (Pág. 214, 217 columna izq) y las modificaciones tales como la pegilación (Fig 1; Pág. 215, 217) y su efecto sobre la farmacocinética de la misma.

De esta manera, frente a las divulgaciones realizadas por D2 y D4 para la persona del oficio normalmente versada en la materia la invención reclamada en la solicitud resultaba obvia y por tanto, incumple el artículo 16 de la D486.

Debe resaltarse que dadas las divulgaciones similares de D1 y D2 sobre la estructura de la esterasa de cocaína cocE y de D3 y D4 sobre las modificaciones para mejorar la actividad las proteínas, estos documentos también podían ser combinados como D1 y D4 ó D2 y D3.

Las reivindicaciones dependientes no realizan un aporte sustancial por lo que no se consideran inventivas.

8. Conclusión

Los requisitos de patentabilidad de la presente solicitud están afectados así:

Ítem	Documento	Reivindicaciones afectadas	Requisito que afecta
D1	Bresler M., <i>et.al.</i> Appl. Environ. Microbiol. Vol 66 (3): 904-908	1-27	Nivel inventivo
D2	Larsen N., <i>et.al.</i> Nature Structural Biology, vol 9(1): 17-21	1-27	Nivel inventivo
D3	Lehman, <i>et.al.</i> Current Opinion in Biotechnology 2001, 12:371-375	1-27	Nivel inventivo
D4	Marshall S. <i>et.al.</i> Drug Discovery today. Vol. 8(5):212-221	1-27	Nivel inventivo

En cumplimiento del artículo 45 de la Decisión 486 de la Comisión de la Comunidad Andina, se le indica que debe dar respuesta dentro del término de 60 días contados a partir de la fecha de notificación de la presente comunicación. En caso de no dar

respuesta al presente requerimiento, o si a pesar de la respuesta subsistieran los impedimentos para la concesión, se denegará la patente.

Se invita al solicitante se sirva indicar, si así lo considera, su renuncia al término faltante para dar respuesta a este requerimiento y presentarla por escrito

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.


JOSE LUIS SALAZAR LÓPEZ
Director de Nuevas Creaciones (C)

El anterior requerimiento fue notificado por fijación en lista, de fecha
13 AGO. 2012

Elaboró: N. Lamprea
Revisó: Consuelo Leguizamón L. 
Aprobó: Jose Luis Salazar

Gene Cloning and Nucleotide Sequencing and Properties of a Cocaine Esterase from *Rhodococcus* sp. Strain MB1

MATTHEW M. BRESLER, SUSAN J. ROSSER, AMRIK BASRAN, AND NEIL C. BRUCE*

Institute of Biotechnology, University of Cambridge, Cambridge CB2 1QT, United Kingdom

Received 6 August 1999/Accepted 1 December 1999

A strain of *Rhodococcus* designated MB1, which was capable of utilizing cocaine as a sole source of carbon and nitrogen for growth, was isolated from rhizosphere soil of the tropane alkaloid-producing plant *Erythroxylum coca*. A cocaine esterase was found to initiate degradation of cocaine, which was hydrolyzed to ecgonine methyl ester and benzoate; both of these esterolytic products were further metabolized by *Rhodococcus* sp. strain MB1. The structural gene encoding a cocaine esterase, designated *cocE*, was cloned from *Rhodococcus* sp. strain MB1 genomic libraries by screening recombinant strains of *Rhodococcus erythropolis* CW25 for growth on cocaine. The nucleotide sequence of *cocE* corresponded to an open reading frame of 1,724 bp that codes for a protein of 574 amino acids. The amino acid sequence of cocaine esterase has a region of similarity with the active serine consensus of X-prolyl dipeptidyl aminopeptidases, suggesting that the cocaine esterase is a serine esterase. The *cocE* coding sequence was subcloned into the pCFX1 expression plasmid and expressed in *Escherichia coli*. The recombinant cocaine esterase was purified to apparent homogeneity and was found to be monomeric, with an M_r of approximately 65,000. The apparent K_m of the enzyme (mean $\pm$ standard deviation) for cocaine was measured as 1.33 ± 0.085 mM. These findings are of potential use in the development of a linked assay for the detection of illicit cocaine.

Tropane alkaloids, which possess a characteristic azabicyclo[3.2.1] octane system, represent one of the most pharmacologically important groups among the alkaloids. In particular, the tropane alkaloids exhibit anticholinergic and anesthetic activities as well as parasympathetic inhibition. Cocaine is the most well known of the tropane alkaloids and is a powerful central nervous system stimulant and adrenergic blocking agent; its hydrochloride salt is also used as a local surface anesthetic in face, eye, nose, and throat operations. Cocaine is naturally located in the leaves of coca plants (*Erythroxylum* spp.) and can be up to 1% of the dry weight content (13). Cocaine is, however, a notorious drug of abuse and is considered to be one of the most powerfully addictive drugs Western society has ever had to confront. Illicit powder cocaine and crack cocaine are generally trafficked as solid particulate matter, and much effort is currently being directed at the development of sensors for drug detection. We envisaged that the isolation of microorganisms from the environment capable of metabolizing drugs, such as cocaine and heroin, as carbon sources for growth might provide a good source of enzymes which could be utilized as recognition components in biosensors for the detection of illicit drugs. Previous work in our laboratory has shown that the enzymes initiating heroin metabolism in bacteria could be successfully used in conjunction with bacterial luciferase for the detection of nanogram quantities of heroin (5, 12, 24). Work has subsequently been directed towards identifying appropriate enzymes active against cocaine. A cocaine esterase was previously identified in a strain of *Pseudomonas maltophilia* termed MB11L that was isolated from a drug processing plant (4). *P. maltophilia* MB11L was capable of utilizing cocaine as a sole source of carbon and energy for growth, and metabolism of cocaine was shown to be initiated by an esterase that hydrolyzed cocaine to benzoate

and ecgonine methyl ester (Fig. 1). Unfortunately, the cocaine esterase was unsuitable for biosensor development, as the enzyme was observed in cell extracts either in an aggregated form or in association with membrane components, and, as such, it proved extremely difficult to purify.

We report here the isolation of a strain of *Rhodococcus* that can utilize cocaine as a sole source of carbon and nitrogen for growth and the cloning, sequencing, and properties of a soluble cytosolic cocaine esterase.

MATERIALS AND METHODS

Organisms, plasmids, and growth conditions. *Rhodococcus* sp. strain MB1 was isolated from soil samples collected from the rhizosphere of coca plants and on the basis of its ability to utilize cocaine as a sole source of carbon. The enrichment conditions and growth conditions were as described by Britt et al. (4). *Escherichia coli* JM109 was obtained from Promega (Southampton, Hampshire, United Kingdom [U.K.]) and was grown according to standard procedures (26). Epicurian Coli XL1-Blue MR was obtained from Stratagene (Cambridge, U.K.). *Rhodococcus erythropolis* CW25 and the *E. coli*-*Rhodococcus* shuttle vector pDA71 were kind gifts from E. Dabbs (University of the Witwatersrand, Johannesburg, South Africa). pCFX1 is an expression vector derived from pBluescript SK(+) by the insertion of the promoter region of pONR (9).

Reagents. Cocaine hydrochloride was purchased from Macfarlan Smith Ltd. (Edinburgh, Scotland, U.K.). Ecgonine methyl ester was synthesized from cocaine as described previously (17). Other reagents were of analytical or higher grade.

DNA manipulation. All restriction enzymes were purchased from New England Biolabs (Hitchin, U.K.) and used according to the manufacturer's protocols. Southern blotting and cloning procedures were performed according to standard methods (26). *Rhodococcus* strains were transformed by electroporation and were rendered electrocompetent via the method used by Kesseler et al. (14). The transformations were conducted using a protocol adapted from those of Desomer et al. (7) and Andersen et al. (2). Cells were grown to exponential phase in a rich Luria-Bertani (LB) medium supplemented with 1.8% (wt/vol) sucrose and 1.5% (wt/vol) glycine. The cells were cooled on ice, harvested by centrifugation, and washed with ice-cold distilled H₂O prior to resuspension in 30% (wt/vol) polyethylene glycol 8000, concentrating the cells 100-fold. Aliquots of cells (400 μ l) were used directly in transformations with the addition of up to 1 μ g of DNA, mixing, and incubation on ice for 10 min in 0.2-cm electroporation cuvettes. Cells were electroporated in a Gene Pulser (Bio-Rad Laboratories, Hercules, Calif.) with a capacitance of 25 μ F, voltage of 2.5 kV, and resistance of 400 Ω to achieve a time constant of more than 7.5 ms. The cells were incubated at 30°C for 3 h in 1 ml of LB medium lacking antibiotic, which allowed pheno-

* Corresponding author. Mailing address: Institute of Biotechnology, University of Cambridge, Tennis Court Rd., Cambridge CB2 1QT, United Kingdom. Phone: 44 (0) 1223 334168. Fax: 44 (0) 1223 334162. E-mail: n.bruce@biotech.cam.ac.uk.

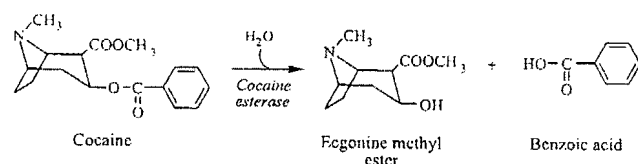


FIG. 1. Cocaine esterase reaction.

typic expression before plating onto LB or minimal medium plates containing 34 µg of chloramphenicol per ml to select for transformants containing pDA71.

Determination of 16S rDNA sequence. Primers ID1 and rD1 were designed around the 5' and 3' ends of bacterial 16S ribosomal DNA (rDNA), respectively, to produce an approximately 1.5-kb fragment of 16S rDNA from bacteria by PCR amplification (29). The annealing temperature was 42°C. The primers had the following sequences: ID1, 5'-CCGAATTCGTCGACAACAGAGTTTGTATCCTGGCTCAG-3'; rD1, 5'-CCCGGGATCCAAGCTTAAGGAGGTGATCCAGCC-3'. The 1.5-kb PCR product amplified from the isolate was gel purified and subjected to direct PCR sequencing using the primer ID1. The DNA sequence obtained was submitted to biological databases for comparison with other 16S rDNAs. The GenBank, EMBL, DDBJ, and PDB databases were searched using the BLASTN search on the National Center for Biological Information website (11).

Construction of genomic libraries. To prepare genomic DNA fragments of approximately 5 to 10 kb, genomic DNA (100 µg) from *Rhodococcus* strain MB1 was subjected to partial digestion with *Sma*3A1 and fractionated on a linear 10 to 40% sucrose density gradient. Fractions containing DNA fragments of the desired size were pooled and ligated into the *E. coli*-*Rhodococcus* shuttle vector pDA71. The plasmid DNA was transformed into the supercompetent *E. coli* (Epicurian Coli XL1-Blue MR) and into *R. erythropolis* CW25 by electroporation. In the case of both the *E. coli* and *Rhodococcus* libraries, colonies were washed off selective plates under sterile conditions using LB medium containing 15% glycerol (vol/vol) and stored at -80°C.

Cloning of *cocE* into an expression host. *cocE* was subcloned by PCR. The primers used were as follows (the bases corresponding to introduced restriction sites *Nde*I and *Hind*III, respectively, are shown in bold): forward, CAG CGA AGG TCG GGA GCA TAT GGT GGA CGG G; reverse, TTT AAG CTT CAG CGT CAG CCA GGC GCG GCT GC. PCR was performed with BioTaq polymerase. The annealing temperature was 67°C. The product was digested with *Nde*I and *Hind*III and ligated into pCFX1 cut with the same enzymes to yield pCOC2.

Protein sequence comparisons. Protein sequence comparisons were performed using the software package MACAW (version 2.05; National Center for Biotechnology Information, National Library of Medicine, Bethesda, MD.) and the BLOSUM62 matrix (11).

Reverse-phase HPLC for analysis of the breakdown of cocaine. Samples were separated by reverse-phase high-pressure liquid chromatography (HPLC) using a model 1050 component system with a multiple wavelength detector (Hewlett-Packard, Waldbronn, Germany) on a Techsphere 5 ODS column (0.46 by 25 cm, 5-µm particles) (HPLC Technology Co. Ltd., Macclesfield, Cheshire, U.K.). The mobile phase consisted of a gradient system comprising an organic phase of methanol and an aqueous phase which contained 1% (wt/vol) glacial acetic acid–50 mM pentane sulfonic acid, adjusted to pH 6.9 with ammonia ($d = 0.88$). Compounds were detected at 228 nm. After application of the sample, the gradient was started with 3% organic phase, which was increased linearly to 84% from 3 to 6 min, where it was held for a further 8 min before being decreased to 3% between 14 and 20 min.

TLC for analysis of the breakdown of cocaine. Thin-layer chromatography (TLC) of cocaine and its metabolites was performed using 200-µm polyester plates precoated with UV-absorbing silica gel (Macherey-Nagel, Duren, Germany). The samples (10 µl) were developed with a mobile phase of ethylacetate-methanol-ammonia (13:7:1 [vol/vol/vol], adapted from the method of Mira et al. [20]). Compounds were detected on the basis of their UV absorbance at 254 nm and color formation when sprayed with Ludy-Tenger reagent (21). Relative migration distances of compounds detected in samples were compared with those of authentic cocaine, benzoate, benzoyl-ecgonine, and ecgonine methyl ester standards.

Purification of cocaine esterase from recombinant *E. coli*. *E. coli* (JM109) carrying pCOC2 was grown until stationary phase on SOB medium (26) containing 100 µg of ampicillin per ml and 0.5 mM isopropyl-β-D-thiogalactopyranoside (IPTG) at 37°C. Cells were pelleted by centrifugation at 10,000 × g and resuspended in 100 mM Tris-HCl at pH 7.5. A cell extract was prepared by sonication of the cells on ice (12 bursts of 15 s at an amplitude of 12 µm), followed by centrifugation of the extract at 20,000 × g for 15 min before ultracentrifugation at 100,000 × g for 60 min. Ammonium sulfate was slowly added to the extract on ice, to a final concentration of 1.25 M. Precipitated protein was removed by centrifugation at 12,000 × g, and the supernatant, which contained the cocaine esterase activity, was further purified by hydrophobic-interaction chromatography. A phenyl-Sepharose CL-4B column (100-ml bed volume; Am-

ersham Pharmacia Biotech Ltd., St. Albans, Hertfordshire, U.K.) was run at 2 ml/min and equilibrated with 1.25 M ammonium sulfate–100 mM Tris-HCl at pH 7.5 before the protein solution was applied. Protein was eluted by decreasing the ammonium sulfate concentration from 0.63 to 0 M over 300 ml, followed by addition of 200 ml of 100 mM Tris-HCl at pH 7.5. Activity-containing fractions from the column were pooled and dialyzed against two 5-liter volumes of 100 mM Tris-HCl at pH 7.5. The dialyzed protein was applied to a Fast Flow Q Sepharose column run at 1 ml/min (80-ml bed volume; Amersham Pharmacia Biotech Ltd.). The column had been previously equilibrated with 100 mM Tris-HCl at pH 7.5 before the protein was applied. A 500-ml NaCl gradient from 0 to 0.6 M was used to elute bound protein. Fractions that contained cocaine esterase activity were pooled and were dialyzed against two 5-liter volumes of 5 mM KH₂PO₄ at pH 7.5. The protein was then loaded onto a Bio-Gel HT hydroxylapatite column run at 0.5 ml/min (12-ml bed volume; Bio-Rad Laboratories). The column had been pre-equilibrated with 5 mM KH₂PO₄ at pH 7.5. After the sample was loaded, the phosphate concentration was increased to 120 mM over 60 ml, during which time the majority of the bound protein was eluted. Activity-containing fractions from the column were pooled and dialyzed against two 2-liter volumes of 100 mM Tris-HCl buffer–100 mM NaCl at pH 7.5. The dialyzed protein was loaded onto a prepacked Resource Q column (6-ml column bed volume; Amersham Pharmacia Biotech Ltd.) at a flow rate of 1 ml/min. The column had been pre-equilibrated with 100 mM Tris-HCl buffer–100 mM NaCl at pH 7.5. A linear sodium chloride gradient from 250 to 400 mM over 132 ml was used to elute the esterase. Fractions containing the cocaine esterase were then pooled.

Estimation of the native molecular weight of the cocaine esterase. The native molecular weight of the purified cocaine esterase was determined by gel filtration chromatography using a prepacked Superose 6 HR column (Amersham Pharmacia Biotech Ltd.). The column was equilibrated with 100 mM Tris-HCl–100 mM NaCl at pH 7.5 and calibrated using a low-molecular-weight calibration kit (Amersham Pharmacia Biotech Ltd.).

Enzyme assays. Cocaine esterase activity was measured by quantifying the amount of benzoate produced from cocaine by reverse-phase HPLC. The standard assay conditions were 100 mM Tris-HCl at pH 7.5, with 10 mM cocaine in a total volume of 100 µl incubated at 30°C. All assay results were compared with those from control assays containing no extract and extract boiled for 5 min to denature the enzyme. For a given cocaine concentration, five assays were set up and incubated at 30°C and stopped at 1-min intervals by the addition of 100 µl of 1 M HCl. The specific activity was expressed in units of cocaine esterase per milligram of total protein, where 1 U of cocaine esterase was defined as the amount required to produce 1 µmol of benzoate per min under these conditions. Assays involving atropine were performed under the same conditions with samples taken after 15, 30, 90, and 180 min. K_m and V_{max} were determined using 0.24 µg (44 U/mg) of cocaine esterase and varying the cocaine concentration between 0.15 and 30 mM. Samples were assayed over a period of 5 min. Kinetic equations were fitted to data using the GraFit 3 software package (Erithacus Software Limited, Middlesex, U.K.).

Nucleotide sequence accession number. The nucleotide sequence data reported have been submitted to GenBank and assigned the accession number AF173155.

RESULTS AND DISCUSSION

Isolation of *Rhodococcus* sp. strain MB1. As cocaine is a natural compound, we considered it likely that microorganisms that had evolved the ability to metabolize cocaine as a carbon source would exist in close proximity to cocaine-producing coca plants. We therefore carried out selective enrichments with soil samples taken from the rhizosphere of coca plants in defined liquid medium, which resulted in the isolation of a bacterium, designated strain MB1, that could utilize cocaine as the sole carbon and nitrogen source for growth. Analysis of the culture broth by TLC and HPLC showed the disappearance of cocaine and the transient accumulation of ecgonine methyl ester and benzoate (data not shown); both of these metabolites supported growth of the bacterium when supplied in minimal medium as a sole carbon source, while ecgonine methyl ester could also be utilized as a source of nitrogen.

Strain MB1 was found to be a gram-positive, rod-forming bacterium. A partial (500-bp) 16S rDNA sequence was obtained for strain MB1, and comparative phylogenetic analysis of the sequences revealed that strain MB1 clustered closely with species entirely from the mycelic acid-containing nocardioform actinomycetes. The highest scoring match was to the 16S rDNA of *Rhodococcus equi* strain DSM 20307T (accession number X80614), to which the DNA from MB1 displayed

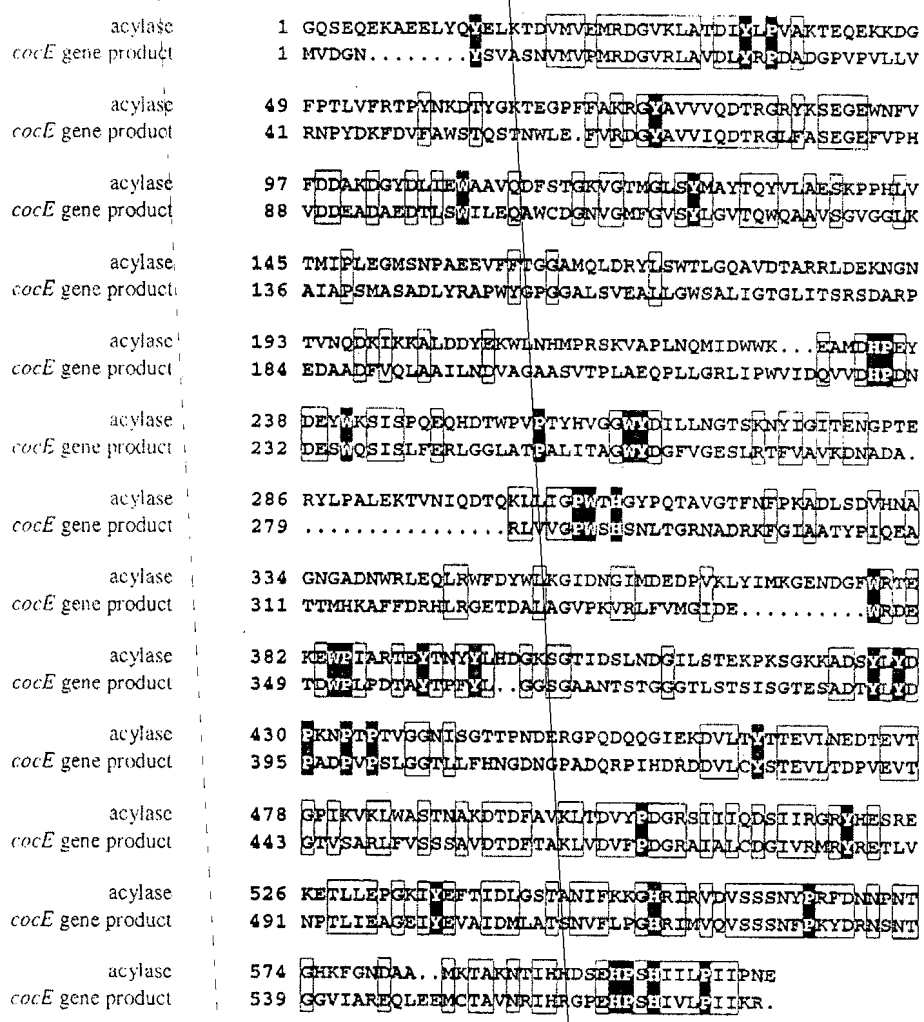


FIG. 2. Alignment of protein encoded by *cocE* with glutaryl 7-aminocapthosporanic acid acylase. The deduced amino acid sequence of the putative cocaine esterase encoded by *cocE* was aligned with the amino acid sequence of the glutaryl 7-aminocapthosporanic acid acylase from *B. laterosporus* J1 (3). The alignment was constructed using the software package MACAW (version 2.05; National Center for Biotechnology Information, National Library of Medicine) and the BLOSUM62 matrix. Dots indicate gaps introduced to maximize the alignment. Pairwise scores from 0 to 33% are not shaded, scores from 33 to 67% are boxed, and scores from 67 to 100% are shaded black, where 100% is the maximum score for the BLOSUM62 matrix.

95.8% sequence similarity. On the basis of the phylogenetic analysis, organism MB1 was identified as a species of the genus *Rhodococcus*.

Cocaine esterase. Extracts from cells grown on 5 mM cocaine were found to possess a cytosolic cocaine esterase with a specific activity of 0.15 U/mg that hydrolyzed cocaine to benzoate and ecgonine methyl ester (Fig. 1). This activity was found to be inducible in *Rhodococcus* sp. strain MB1, since no cocaine esterase activity was observed in cells grown on 15 mM succinate as a sole source of carbon.

Cloning and sequencing of *cocE*. In order to obtain significant quantities of biomass for enzyme purification and characterization studies, prohibitive quantities of cocaine were required as a growth substrate; therefore, it was decided to clone the gene encoding the cocaine esterase activity and express it in *E. coli*. The cloning strategy was to construct genomic libraries of MB1 using the *E. coli*-*Rhodococcus* shuttle vector pDA71. The libraries were then transformed into a rhodococcal host strain, *R. erythropolis* CW25, which is capable of growth on benzoate but not on cocaine. Expression of a cocaine esterase gene was selected for by screening recombinant

strains of *R. erythropolis* CW25 for growth on cocaine as the sole carbon source. This procedure yielded eight clones that were capable of slowly hydrolyzing cocaine to benzoate and ecgonine methyl ester; the benzoate was further metabolized as a growth substrate by the clones. All the cocaine-degrading clones contained an identical 2.7-kb fragment in the recombinant plasmid designated pCOC1. Sequencing of the inserted fragment indicated the presence of an open reading frame encoding a polypeptide of 574 amino acids from which the theoretical molecular mass of the protein was calculated to be 62,128 Da. Cell extracts of *R. erythropolis* CW25 as well as extracts prepared from *E. coli* JM109 harboring the construct pCOC1 displayed low levels of cocaine esterase activity (4.8×10^{-3} and 1.5×10^{-3} U/mg, respectively). This evidence, taken together with the sequence analysis (see below), supports the suggestion that *cocE* encoded the cocaine esterase.

The low rate of growth on cocaine of *R. erythropolis* CW25 transformants harboring pCOC1 can, certainly in part, be explained by the fact that only 23 bp of the native *Rhodococcus* sp. strain MB1 DNA upstream of the *cocE* start codon was cloned. The regions controlling transcription and protein

cocE	87	HVDDDEADAEDTL	SWILEQAMCDGNVGMFGVSYLGVTQWQA	AVSGVGGLKATAPSMASADLY	147
P40334	333	TGDRVAYTDRTRRFETKAS	MCSCGNVGMTGRSYLGLTLQIAI	IATTGVKGLKTVVSEAAISSWY	393
Q59485	333	HGDRVAYTDRTRTVQTTAD	MCNNGNIGMTGRSYLGLTLQIAI	IATTGVKGLKTVVSEAAISSWY	393
Q48571	333	HGDRVAYTDRTRTVQTTAD	MCNNGNIGMTGRSYLGLTLQIAI	IATTGVKGLKTVVSEAAISSWY	393
P22093	318	NGRARAYTSRKKTHEIKAS	WANGKVAMTGKSYLGLTMAYGA	ATTGVEGLEVLAEAGISSWY	378
P22346	318	NGRARAYTSRKKTHEIKAS	WANGKVAMTGKSYLGLTMAYGA	ATTGVEGLEVLAEAGISSWY	378
			GxSYxG		

FIG. 3. Alignment of cocaine esterase with X-prolyl dipeptidyl aminopeptidases from lactic acid bacteria. The amino acid sequence of a region of the cocaine esterase encoded by *cocE* (*cocE*) was aligned with the homologous regions of five cloned X-prolyl dipeptidyl aminopeptidases from *Lactobacillus* and *Lactococcus* spp. (labeled by accession no.) (6, 18, 19, 23, 27, 29, 30). The alignment was constructed using the software package MACAW (version 2.05; National Center for Biotechnology Information, National Library of Medicine) and the BLOSUM62 matrix. Mean scores from 0 to 33% are not shaded, mean scores from 33 to 67% are boxed, and mean scores from 67 to 100% are shaded black, where 100% is the maximum mean score for the BLOSUM62 matrix. The G-X-S-Y-X-G consensus sequence surrounding the active-site serine (*) in the three known X-prolyl dipeptidyl aminopeptidases (mammalian, yeast, and bacterial) is indicated (X is a nonconserved amino acid).

translation in *Rhodococcus* are not well characterized, and the position at which transcription is initiated in *Rhodococcus* genes whose promoter regions have been identified varies considerably. For example, the mRNA transcript of the *nitA* (nitrilase) gene from *Rhodococcus rhodochrous* J1 is initiated only 26 bases upstream of the ATG start site (15), while the transcription initiation site of the *cmr* (chloramphenicol resistance) gene from *Rhodococcus fascians* NCPPB 1675 was determined to be 279 bases upstream of the start codon (8). In the other rhodococcal genes whose transcriptional regulation has been investigated, transcription was shown to be initiated at distances of 46 bases (16), 78 bases (22), and 168 bases (10) upstream of the translational start site, and promoter regions further upstream of these sites were shown to be essential for gene expression.

A region to which ribosomes are capable of binding to initiate translation in the rhodococcal host harboring *cocE* appears to be located within a few bases upstream of the ATG start codon (5'GGGAG3'); however, it is likely that transcription is initiated from fortuitous regions of the *EcoRI* endonuclease gene in pCOC1 that the *Rhodococcus* transcription machinery is capable of recognizing. Since there appears to be little homology between the *Rhodococcus* promoter regions identified to date, such regions could not be identified with confidence. However, regions of the *EcoRI* sequence upstream of *cocE* in pCOC1 with homology to known *E. coli* -10 and -35 promoter regions were detected in the upstream *EcoRI* DNA (data not shown) that may be responsible for the low levels of expression in *E. coli* JM109.

Sequence comparisons of CocE with other related sequences. The deduced amino acid sequence was used for homology searches with a range of biological databases, and it was found that CocE was related to a number of prokaryotic enzymes. The closest match was to a glutaryl 7-aminocephalosporanic acid acylase enzyme from *Bacillus laterosporus* J1, which converts glutaryl 7-aminocephalosporanic acid to 7-aminocephalosporanic acid through the cleavage of an acyl linkage (3). The two proteins exhibit 37.2% sequence identity and 46.7% sequence similarity. Their similarity is illustrated by an alignment of the amino acid sequences (Fig. 2). The other high-scoring matches on the protein level were three hypothetical proteins of unknown function from *Mycobacterium* and five matches to X-prolyl dipeptidyl aminopeptidase enzymes from *Lactococcus lactis* and *Lactobacillus* spp. (Fig. 3), but the homology to these protein sequences was over a significantly smaller region (approximately 60 residues [18, 19, 23, 27, 30]). On average, a sequence identity of 39% and a sequence similarity of 50% with the dipeptidyl aminopeptidases were displayed over this region. The X-prolyl dipeptidyl aminopeptidases have amidase and esterase activities in addition to peptidase activities (29). These enzymes are serine proteases,

and the consensus sequence surrounding the active-site serine has been identified as G-X-S-Y-X-G, where X is a nonconserved amino acid (6). The sequence of CocE shows conservation of sequence with the X-prolyl dipeptidyl aminopeptidases from *L. lactis* and *Lactobacillus* spp. around the active-site serine, suggesting that *cocE* encodes a serine esterase (Fig. 3).

Expression of *cocE* in *E. coli*. High levels of expression of *cocE* were achieved by subcloning the *cocE* coding region, obtained by PCR, into pCFX1 to yield pCOC2. Cell extracts of recombinant *E. coli* were found to contain cocaine esterase activity with a specific activity of 0.13 U/mg of protein.

In order to characterize the recombinant cocaine esterase, the enzyme was purified 350-fold to homogeneity, as judged by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) (Fig. 4); it possessed a specific activity of 44 U/mg (Table 1). The native molecular mass of the purified esterase was determined by gel filtration to be approximately 65,000 Da. This value is consistent with the subunit molecular weight of the cocaine esterase, as determined by SDS-PAGE and through translation of *cocE*, which suggests that active cocaine esterase exists as a monomer. The apparent K_m (mean $\pm$ standard deviation) of the enzyme was found to be 1.33 ± 0.085 mM. The cocaine esterase displayed low levels of activity (2.1 U/mg) with 20 mM atropine, a structurally related tropane alkaloid. Interestingly, bacterial atropine esterases do not display any activity against cocaine (25), while an estero-lytic activity from a strain of *P. maltophilia* that was isolated

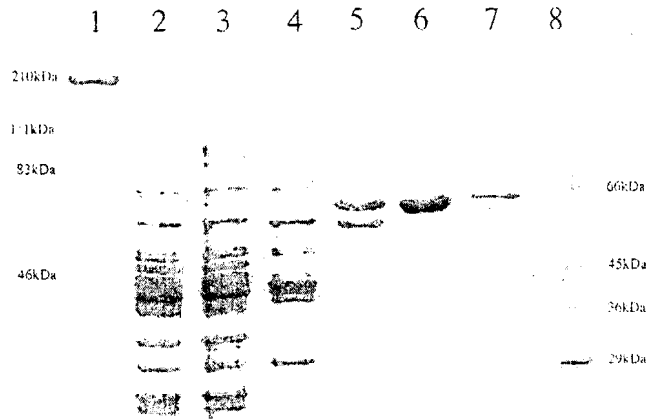


FIG. 4. SDS-PAGE analysis of the purification of cocaine esterase from recombinant *E. coli*. Lanes 1 and 8, high- and low-range molecular mass markers; lane 2, cell extract from *E. coli* JM109 pCOC2; lane 3, product after ammonium sulfate fractionation; lanes 4 to 6, cocaine esterase activity obtained after hydrophobic-interaction chromatography, anion-exchange chromatography, and hydroxyapatite chromatography, respectively; lane 7, purified cocaine esterase (2 µg).

TABLE 1. Purification of cocaine esterase activity

Purification step	Amt of protein (mg)	Sp act (U/mg)	Total activity (U)	Purification (fold)	Recovery (%)
Cell extract	4,884	0.124	605.6	1.00	100
Ammonium sulfate	4,091	0.144	589.1	1.16	97.3
Phenyl-Sepharose	1,034	0.329	340.7	2.66	56.3
Q Sepharose	80.4	2.543	204.5	20.5	33.8
Hydroxylapatite	12.9	12.13	156.7	97.9	25.9
Resource Q	2.75	43.92	120.8	354	20.0

from industrial waste liquors was shown to display activity against cocaine but not atropine (4). The N-terminal sequence of the first 11 amino acid residues of the purified protein was determined, and it was found to match the deduced amino acid sequence of cocaine esterase (CocE).

Finally, the discovery of this soluble cocaine esterase is of particular interest for its use in a potential illicit-drug sensor. Enzymatic detection of illicit heroin has already been described (12). In conjunction with a suitable NADP⁺-dependent dehydrogenase active against ecgonine methyl ester, the cocaine esterase could enable simultaneous detection of heroin and cocaine, using established bioluminescence technology (12).

ACKNOWLEDGMENTS

This work was supported by the Leverhulme Trust and the BBSRC.

REFERENCES

1. Altschul, S. F., W. Gish, W. Miller, E. W. Meyers, and D. J. Lipman. 1990. Basic local alignment search tool. *J. Mol. Biol.* **215**:403–410.
2. Andersen, S. J., S. Quan, B. Gowan, and E. R. Dabbs. 1997. Monooxygenase-like sequence of a *Rhodococcus equi* gene conferring increased resistance to rifampin by inactivating this antibiotic. *Antimicrob. Agents Chemother.* **41**: 218–221.
3. Aramori, I., M. Fukagawa, M. Tsumura, M. Iwami, H. Ono, H. Kojo, M. Kohsaka, Y. Ueda, and H. Imanaka. 1991. Cloning and nucleotide sequencing of a novel 7β-(4-carboxybutanamido)cephalosporanic acid acylase gene of *Bacillus laterosporus* and its expression in *Escherichia coli* and *Bacillus subtilis*. *J. Bacteriol.* **173**:7848–7855.
4. Britt, A. J., N. C. Bruce, and C. R. Lowe. 1992. Identification of a cocaine esterase in a strain of *Pseudomonas maltophilia*. *J. Bacteriol.* **174**:2087–2094.
5. Cameron, G. W. W., K. N. Jordan, P.-J. Holt, P. B. Baker, C. R. Lowe, and N. C. Bruce. 1994. Identification of a heroin esterase in *Rhodococcus* sp. strain H1. *Appl. Environ. Microbiol.* **60**:3881–3883.
6. Chieh, J. F., M. P. Chapot-Chartier, B. Ribadeau-Dumas, and J. C. Gripon. 1992. Identification of the active-site serine of the X-prolyl dipeptidyl aminopeptidase from *Lactococcus lactis*. *FEBS Lett.* **314**:139–142.
7. Desomer, J., P. Dhaese, and M. Van Montagu. 1990. Transformation of *Rhodococcus fascians* by high-voltage electroporation and development of *R. fascians* cloning vectors. *Appl. Environ. Microbiol.* **56**:2818–2825.
8. Desomer, J., D. Vereecke, M. Crespi, and M. van Montagu. 1992. The plasmid-encoded chloramphenicol-resistance protein of *Rhodococcus fascians* is homologous to the transmembrane tetracycline efflux proteins. *Mol. Microbiol.* **6**:2377–2385.
9. French, C. E., S. Nicklin, and N. C. Bruce. 1996. Sequence and properties of pentacythritol tetranitrate reductase from *Enterobacter cloacae* PB2. *J. Bacteriol.* **178**:6623–6627.
10. Grzeszik, C., M. Lubbers, M. Reh, and H. G. Schlegel. 1997. Genes encoding

the NAD-reducing hydrogenase of *Rhodococcus opacus* MR11. *Microbiology* **143**:1271–1286.
11. Heikoff, S., and J. G. Heikoff. 1992. Amino acid substitution matrices from protein blocks. *Proc. Natl. Acad. Sci. USA* **89**:10915–10919.
12. Holt, P. J., N. C. Bruce, and C. R. Lowe. 1996. Bioluminescent assay for heroin and its metabolites. *Anal. Chem.* **68**:1877–1882.
13. Johnson, E. L. 1996. Alkaloid content in *Erythroxylum coca* tissue during reproductive development. *Phytochemistry* **42**:35–38.
14. Kessler, M., E. R. Dabbs, B. Averhoff, and G. Gottschalk. 1996. Studies on the isopropylbenzene 2,3-dioxygenase and the 3-isopropylcatechol 2,3-dioxygenase genes encoded by the linear plasmid of *Rhodococcus erythropolis* BD2. *Microbiology* **142**:3241–3251.
15. Komeda, H., Y. Hori, M. Kobayashi, and S. Shimizu. 1996. Transcriptional regulation of the *Rhodococcus rhodochrous* J1 *nit1* gene encoding a nitrilase. *Proc. Natl. Acad. Sci. USA* **93**:10572–10577.
16. Li, M. Z., C. H. Squires, D. J. Monticello, and J. D. Childs. 1996. Genetic analysis of the *dsz* promoter and associated regulatory regions of *Rhodococcus erythropolis* IGT58. *J. Bacteriol.* **178**:6409–6418.
17. Lister, D. L., R. F. Sproule, A. J. Britt, C. R. Lowe, and N. C. Bruce. 1996. Degradation of cocaine by a mixed culture of *Pseudomonas fluorescens* MBER and *Comamonas acidovorans* MBLF. *Appl. Environ. Microbiol.* **62**: 94–99.
18. Mayo, B., J. Kok, K. Venema, W. Bockelmann, M. Teuber, H. Reinke, and G. Venema. 1991. Molecular cloning and sequence analysis of the X-prolyl dipeptidyl aminopeptidase gene from *Lactococcus lactis* subsp. *cremoris*. *Appl. Environ. Microbiol.* **57**:38–44.
19. Meyer-Barton, E. C., J. R. Klein, M. Imam, and R. Plapp. 1993. Cloning and sequence analysis of the X-prolyl dipeptidyl aminopeptidase gene (Pepx) from *Lactobacillus delbrueckii* ssp. *lactis* DSM 7290. *Appl. Microbiol. Biotechnol.* **40**:82–89.
20. Mira, A. L., R. B. Pontani, and S. J. Mule. 1973. Separation of cocaine, some of its metabolites and congeners on glass fibre sheets. *J. Chromatogr.* **81**: 167–169.
21. Moffat, A. C. 1986. Clarke's isolation and identification of drugs, 2nd ed. The Pharmaceutical Press, London, England.
22. Nagy, I., G. Schoofs, F. Compernelle, P. Proost, J. Vanderleyden, and R. De Mot. 1995. Degradation of the thiocarbamate herbicide EPTC (S-ethyl dipropylcarbamothioate) and bioassaying by *Rhodococcus* sp. strain N18/21 involve an inducible cytochrome P-450 system and aldehyde dehydrogenase. *J. Bacteriol.* **177**:676–687.
23. Nardi, M., M.-C. Chopin, A. Chopin, M.-M. Cals, and J.-C. Gripon. 1991. Cloning and DNA sequence analysis of an X-prolyl dipeptidyl aminopeptidase gene from *Lactococcus lactis* subsp. *lactis* NCDO 763. *Appl. Environ. Microbiol.* **57**:45–50.
24. Rathbone, D. A., P. J. Holt, C. R. Lowe, and N. C. Bruce. 1996. The use of a novel recombinant heroin esterase in the development of an illicit drugs biosensor. *Ann. N. Y. Acad. Sci.* **799**:90–96.
25. Rorsch, A., F. Berends, H. C. Bartlema, and W. F. Stevens. 1971. The isolation and properties of *Pseudomonas* strains growing on atropine and producing an atropine esterase. *Proc. K. Ned. Akad. Wet. Ser. C* **74**:132–147.
26. Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
27. Vesanto, E., K. Savijoki, T. Rantanen, J. L. Steele, and A. Palva. 1995. An X-prolyl dipeptidyl aminopeptidase (Pepx) gene from *Lactobacillus helveticus*. *Microbiology* **141**:3067–3075.
28. Weisburg, W. G., S. M. Barns, D. A. Pelletier, and D. J. Lane. 1991. 16S ribosomal DNA amplification for phylogenetic study. *J. Bacteriol.* **173**:697–703.
29. Yoshpe-Besancon, I., J. C. Gripon, and B. Ribadeau-Dumas. 1994. Xaa-Pro dipeptidyl aminopeptidase from *Lactococcus lactis* catalyzes kinetically controlled synthesis of peptide bonds involving proline. *Biotechnol. Appl. Biochem.* **20**:131–140.
30. Yüksel, G. U., and J. L. Steele. 1996. DNA sequence analysis, expression, distribution, and physiological role of the Xaa-prolyl dipeptidyl aminopeptidase gene from *Lactobacillus helveticus* CNRZ32. *Appl. Microbiol. Biotechnol.* **44**:766–773.

Crystal structure of a bacterial cocaine esterase

Nicholas A. Larsen¹, James M. Turner¹, James Stevens¹, Susan J. Rosser², Amrik Basran², Richard A. Lerner^{1,3}, Neil C. Bruce² and Ian A. Wilson^{1,3}

¹Department of Molecular Biology, The Scripps Research Institute, 10550 North Torrey Pines Rd., La Jolla, California 92037, USA. ²Institute of Biotechnology, Cambridge University, Tennis Court Rd., Cambridge CB2 1QT, UK. ³The Skaggs Institute for Chemical Biology, The Scripps Research Institute, 10550 North Torrey Pines Rd., La Jolla, California 92037, USA.

Published online: 17 December 2001, DOI: 10.1038/nsb742

Here we report the first structure of a cocaine-degrading enzyme. The bacterial esterase, cocE, hydrolyzes pharmacologically active (–)-cocaine to a nonpsychoactive metabolite with a rate faster than any other reported cocaine esterase ($k_{\text{cat}} = 7.8 \text{ s}^{-1}$ and $K_M = 640 \text{ nM}$). Because of the high catalytic proficiency of cocE, it is an attractive candidate for novel protein-based therapies for cocaine overdose. The crystal structure of cocE, solved by multiple anomalous dispersion (MAD) methods, reveals that cocE is a serine esterase composed of three domains: (i) a canonical α/β hydrolase fold (ii) an α -helical domain that caps the active site and (iii) a jelly-roll-like β -domain that interacts extensively with the other two domains. The active site was identified within the interface of all three domains by analysis of the crystal structures of transition state analog adduct and product complexes, which were refined at 1.58 Å and 1.63 Å resolution, respectively. These structural studies suggest that substrate recognition arises partly from interactions between the benzoyl moiety of cocaine and a highly evolved specificity pocket.

The cocaine esterase cocE was identified in the bacteria *Rhodococcus* sp. strain MB1 that thrives in the rhizosphere soil of the cocaine-producing plant *Erythroxylum coca*¹. This *Rhodococcus* strain is capable of using cocaine as a sole carbon and nitrogen source; cocE is the first enzyme in the metabolic pathway leading to cocaine catabolism¹. This esterase hydrolyzes the benzoyl ester of cocaine to yield a nonpsychoactive metabolite, ecgonine methyl ester (Fig. 1)¹. Although other natural enzymes (serum butyrylcholinesterase (BchE)², and human liver carboxylesterase (CE) enzymes hCE-1 (ref. 3) and hCE-2 (ref. 4) accept cocaine as part of a wide repertoire of substrates, cocE and a corresponding esterase from *Pseudomonas maltophilia*⁵ are the only known enzymes that have evolved under direct selection pressure for cocaine hydrolysis.

Despite many advances in the understanding of cocaine, no effective pharmacotherapy currently exists for cocaine addiction and overdose⁶. An alternative approach to the development of antagonists, agonists and antidepressants for psychotherapeutic intervention is an antibody-based strategy, in which actively derived or passively administered antibodies bind cocaine before it reaches the brain, thus blunting its effect⁷. Although the long

plasma lifetime of actively or passively derived antibodies is suitable for the extended periods of time required for rehabilitation and recovery from addiction, a cocaine-degrading catalyst would be more effective for rapid detoxification in an emergency setting. Assuming that no adverse effects occur upon administration of cocE, the kinetic parameters of cocE ($k_{\text{cat}} = 7.8 \text{ s}^{-1}$ and $K_M = 640 \text{ nM}$) are commensurate with such a detoxification strategy⁸.

Considerable interest has arisen in generating tailor-made catalytic antibodies as therapeutics for cocaine addiction⁹. A thorough understanding of cocE will provide substantial insight for the generation of more efficient cocaine antibody catalysts. Because of the therapeutic potential of cocE and cocaine catalytic antibodies, we initiated these crystallographic studies of the cocE enzyme to determine the structural basis for the specificity and catalytic mechanism for cocaine degradation.

Overview of cocE structure

The cocE enzyme is a 574-amino acid protein with a molecular weight of ~65,000 Da (GenBank AF173165)¹. The overall secondary structure is a mixture of β -sheet (30%) and α -helix (23%) (Fig. 2). The globular structure (~80 Å × ~60 Å × ~40 Å) can be subdivided into three domains¹⁰, DOM1 (residues 1–144 and 241–354), DOM2 (145–240), and DOM3 (355–574), whose structures are a three-layered α/β sandwich, α -helical domain and jelly-roll β -barrel, respectively (Fig. 2).

Domain structure and association

DOM1 is assembled from two linear sequence segments, residues 1–144 and 241–354, due to insertion of a helical domain, DOM2 (Fig. 2a,b). DOM1 has the canonical hydrolase fold, which is composed of repeating β - α motifs that form a central 10-stranded β -sheet flanked on either side by α -helices (Fig. 2a,b)^{11,12}. The central β -sheet is predominantly parallel, with topology +1, +2, –1 $\times$, +2 $\times$, (+1 $\times$)₅ (strand order β 1, β 2, β 4, β 3, β 5, β 6, β 7, β 8, β 9 and β 10) and is capped on the C-terminal edge by DOM2, which is inserted between strands β 6 and β 7 (Fig. 2b). Strands β 9 and β 10 are C-terminal additions to the canonical eight-stranded α/β hydrolase central sheet, which is similar to C-terminal additions observed in members of the large esterase family of the α/β hydrolase superfamily¹³.

For members of the α/β hydrolase superfamily, insertions are common at the junction between β 6 and β 7. The length of DOM2 (95 amino acids) sets a new upper limit on the insertion size tolerated by the α/β hydrolase fold, further underscoring the remarkable diversity already noted in this family¹³. DOM2 is an α -helical structure composed of seven helices. Helices H1 and H7 pack together at the base of the domain, while helices H2–H6 pack together to form a five-helix core (Fig. 2c). The longest helices are H2 and H3 (Fig. 2c), which pack antiparallel to each other with the helical axes offset by ~25°.

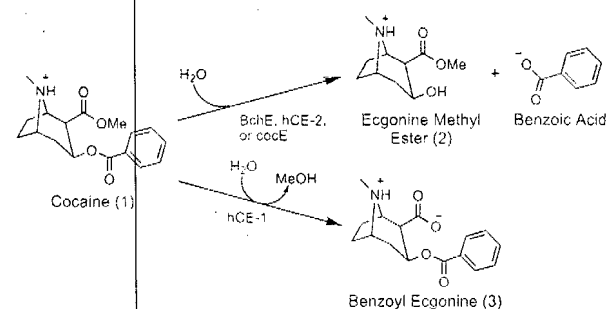


Fig. 1 Cocaine hydrolysis. Both the benzoyl ester, and the methyl ester of cocaine (1) can be hydrolyzed by different enzymes to give nonpsychoactive metabolites ecgonine methyl ester (2) and benzoyl ecgonine (3). Human liver esterases hCE-1 and hCE-2, and the serum esterase butyrylcholinesterase BchE account for the ~30 min *in vivo* half-life of cocaine. CocE hydrolyzes the benzoyl ester of cocaine to form ecgonine methyl ester (2) and benzoic acid.

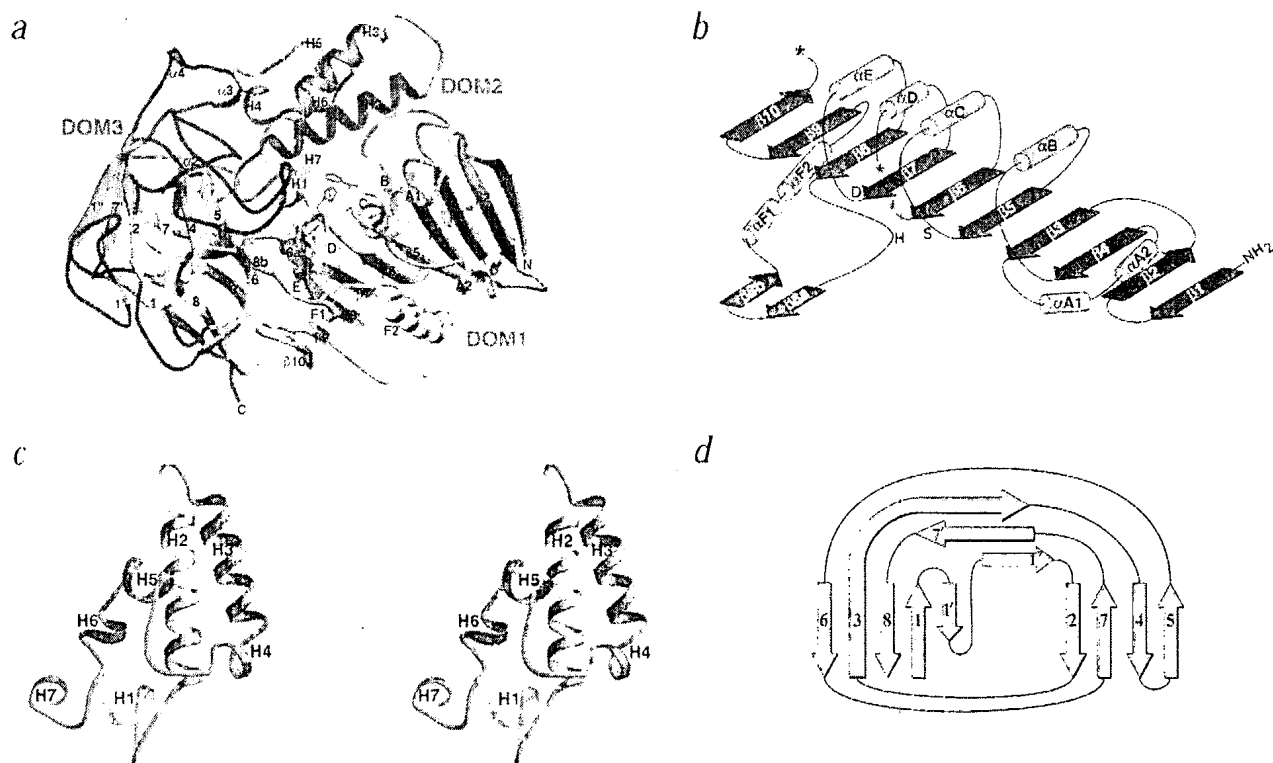


Fig. 2 CocE structure and topology. **a**, Ribbon diagram of the cocE structure. The enzyme is composed of three domains: DOM1 (red sheet, orange helices and gray loops), DOM2 (green) and DOM3 (blue). Note that this coloring scheme will be used in all subsequent figures. The active site is located at the interface of all three domains, and the residues of the catalytic triad are shown. A molecule of benzoic acid was located in the specificity pocket of the enzyme. The figure was generated in BOBSCRIPT²² and rendered in RASTER3D²⁴. **b**, Topology diagram for DOM1. DOM1 has the canonical α/β hydrolase fold. The β -sheet and β - α - β motifs are all right handed, except for the crossover between strand $\beta 8$ and $\beta 9$ which is left handed and contains the active site His, consistent with other hydrolase structures¹². S, H and D denote the relative position of the catalytic triad Ser 117, His 287 and Asp 259, respectively. DOM2 is inserted between strands $\beta 6$ and $\beta 7$ (green stars). DOM3 is located at the C terminal end of DOM1 (blue star). **c**, Stereo view of the structure of DOM2. The orientation of DOM2 has been altered relative to (a) to show the relative orientation of the seven helices comprising this domain. The core is composed of five helices, H2–H6. **d**, Topology diagram for DOM3. The overall fold has a jelly-roll-like topology. Strands in the two sheets are connected by crossover loops, several of which form additional secondary elements, including an additional three-stranded β -sheet that caps the jelly-roll core. The long loops form extensive interactions with DOM1 and DOM2. The 1', 1'' and 7' strands are labeled to maintain the topological homology with strands 1–8 of other jelly-roll proteins.

DOM3 is mostly a β -structure, with a β -barrel-like core formed by two antiparallel β -sheets ($\beta 2$, $\beta 7$, $\beta 4$ and $\beta 5$) and ($\beta 1'$, $\beta 1$, $\beta 8$, $\beta 3$ and $\beta 6$) that pack against each other with an overall jelly-roll topology (Fig. 2a,d). The edge strand $\beta 1'$ is inserted after strand $\beta 1$ in this fold, and two hydrogen bonds are formed between edge strands $\beta 1'$ and $\beta 2$. The jelly-roll fold, however, is compressed into a β -sandwich like structure; the lack of curvature in the sheets prevents hydrogen bonding between $\beta 5$ and $\beta 6$ despite their spatial proximity. The strands between the sheets are connected by six crossover loops, several of which form additional secondary elements that surround the β -barrel-like core, including an additional three stranded antiparallel β -sheet ($\beta 1''$, $\beta 7'$ and $\beta 3$) that caps the core (Fig. 2d).

The association of DOM1 and DOM2 buries a total surface area¹⁴ of 1,342 Å², whereas the DOM3 interaction with DOM1 and DOM2 buries 3,940 Å². A total of 377 contacts¹⁵ (27 hydrogen bonds, 10 salt bridges and 340 van der Waals) are formed between DOM3 and DOM1/2, of which 195 are with DOM1 and 182 with DOM2. All of the contacts between DOM1 and DOM3 occur after the DOM2 insertion. The crossover loops and short helices in DOM3 account for 271 (~72%) of the total contacts, with most of the remaining contacts (94) derived from the β -barrel like core (strands $\beta 3$, $\beta 5$ and $\beta 6$). Hence, DOM3 appears

to play an important role as a scaffold governing the arrangement of the overall tertiary structure.

Architecture of the active site

Two ligand structures (Table 1) were obtained: crystals soaked in cocaine or phenyl boronic acid yielded the product (benzoic acid) and tetrahedral-analog adduct complexes, respectively (Fig. 3). The active site is located in a cleft formed at the interface of the three domains. The catalytic triad, located towards the bottom of this cleft in DOM1, comprises residues Ser 117, Asp 259 and His 287 at the C-terminal end of strands $\beta 5$, $\beta 7$ and $\beta 8$, respectively (Fig. 3). The sequence surrounding Ser 117 is GVSYLG, a slight variant of the canonical GX₁SX₂G motif (where X is any amino acid) found in most members of the α/β hydrolase superfamily. Ser 117 is located in the strained nucleophilic elbow ($\phi_{S117} = 55^\circ$ and $\psi_{S117} = -112^\circ$), which is a hallmark of this fold¹².

A binding pocket is located at the very base of the cleft just below the catalytic triad. In the product complex, excellent density is visible in this binding pocket for benzoic acid (Fig. 3a). The benzoic acid is 95% buried within this pocket (124 Å² out of 130 Å²)¹⁴ and forms van der Waals (vdw) contacts with residues from all three domains (Fig. 3a). The benzoic acid interacts predominantly with DOM1 (23 vdw contacts and two hydrogen

bonds) and DOM2 (18 vdw contacts), with fewer contacts to DOM3 (four vdw contacts)¹⁵. One face of the benzoate forms a π -stacking interaction with the five-membered pyrrole ring of Trp 166. The DOM3 residues contacting benzoate are Leu 407 and Phe 408 (Fig. 3a), which belong to a long loop that connects strand $\beta 1'$ and $\beta 2$. This loop inserts between DOM1 and DOM2 in the quaternary structure to form part of the active site (Fig. 2a).

In addition to the specificity pocket, we have identified a two-component oxyanion hole. The backbone amide of Tyr 118 and the hydroxyl group of Tyr 44 both form hydrogen bonds to the benzoate carboxyl (Fig. 3a). These hydrogen bonds are precisely positioned to activate the benzoyl carbonyl carbon of cocaine for serine attack and subsequent hydrolysis of the acyl intermediate (Fig. 4). Therefore, Tyr 44 appears to play a decisive role in activating the substrate carbonyl and the acyl enzyme for hydrolysis. Because the oxyanion hole is also located at the N-terminus of helix αC , the partial positive charge from the helix dipole may play a subtle role in stabilizing the oxyanion transition state.

Structural changes along the reaction coordinate

Crystals soaked with an inhibitor, phenyl boronic acid, gave rise to a covalently bound tetrahedral-intermediate analog (Table 1; Fig. 3b). The pro-*S* hydroxyl of the boronic acid is positioned in the oxyanion hole, whereas the pro-*R* hydroxyl is oriented toward the putative secondary specificity pocket for the tropane ring of cocaine (Figs 3b, 4). The Ser residue binds covalently to boron through attack at the corresponding *si* face of the benzoyl ester linkage of cocaine, consistent with the enantio-recognition observed in other members of the α/β hydrolase superfamily¹⁶.

The protonated N ϵ H of His 287 forms a bifurcated hydrogen bond to the Ser 117 O γ and the pro-*R* hydroxyl (Figs 3b, 4). In this structure, the pro-*R* hydroxyl is positioned where the attacking water molecule would be during nucleophilic attack of the acyl intermediate (Fig. 4). The bifurcated hydrogen bond in this stable complex probably resembles that which is formed during the transition state of this reaction. Hence, this structure cap-

tures the catalytic triad in a productive configuration that reproduces the classic geometry seen in other esterase and protease structures. Comparison of the adduct and product structures shows that hydrolysis of the acyl enzyme causes Ser 117 to rotate 110° about χ_1 to face away from the benzoic acid specificity pocket. Additionally, His 287 swivels 57° about χ_1 , leading to an unproductive configuration of the triad in which the interaction between Asp 259 and His 287 is mediated by a water molecule (Figs 3a, 4).

Modeling cocaine in the active site by superimposing the benzoyl ester of cocaine on the pro-*R* oxygen of the phenyl boronic acid adduct shows that recognition of the tropane ring would occur primarily in a putative secondary specificity pocket composed of residues from DOM1. Recognition of the tropane nitrogen may occur *via* cation- π interaction with Tyr 44 or a hydrogen bond with Gln 55.

Comparison to other hydrolase structures

In general, members of the hydrolase family of enzymes have very low sequence homology and are thought to represent the prototypic example of divergent evolution¹². Indeed, DALI¹⁷ identified *Streptomyces* lipase (PDB 1JFR)¹⁸ as having the highest structural homology to DOM1 (Z-score = 19.1) but with only 17% sequence identity. Prolyl oligopeptidase from porcine muscle (PDB 1QFS)¹⁹ also has high structural homology to cocE. Remarkably, prolyl oligopeptidase and cocE use a Tyr hydroxyl to stabilize the oxyanion, differing from the chymotrypsin-type and α/β hydrolase enzymes, which invoke two main chain NH hydrogen bonds¹⁹, and subtilisin-type enzymes that use a side chain Asn residue. Superposition of this structure onto cocE shows that the Tyr hydroxyls superimpose within 0.5 Å of each other. The unusual role of this Tyr in both the bacterial esterase and the mammalian oligopeptidase provides an elegant example of convergent evolution in the formation of the oxyanion hole.

Comparison to other α/β -hydrolase structures reveals that a wide variety of domains are typically inserted at the N- or C-terminus, and/or the junction between strands $\beta 6$ and $\beta 7$ (DOM2).

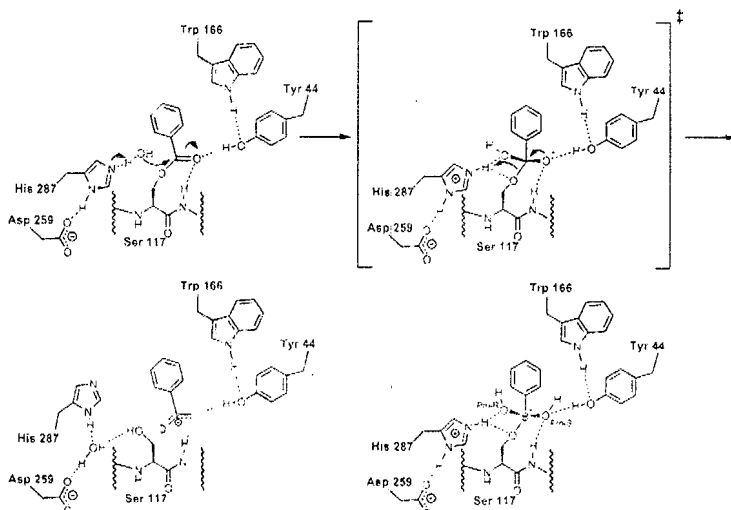
These additional domains are generally believed to tailor the specificity of the enzymes^{12,13,20}. This apparent structural diversity prompted us to search for structural homology to DOM2 and DOM3 alone. DOM2 proved to be quite novel: the highest DALI Z-score was only 3.9, where a Z-score <2 represents a totally dissimilar structure¹⁷. The structural homology for domain DOM3 is also quite low, with the highest DALI Z-score of only 6.



Fig. 3 Stereo views of the active site of cocE. The color scheme and secondary structure is the same as in Fig. 2a. The side chains that contact the bound ligands are shown in brown with red oxygens and blue nitrogens. **a**, The product complex. Hydrolysis of the acyl enzyme gives the product complex with benzoic acid. After hydrolysis, Ser 117 swings away from the specificity pocket, and His 287 rotates towards DOM3, obliterating the productive catalytic triad configuration. Hydrogen bonds within the triad are mediated by a water molecule. The σ_A -weighted³⁵ $2F_o - F_c$ electron density map contoured at 2.0σ is shown around the Ser residue and benzoic acid. **b**, The tetrahedral analog adduct complex. This complex was formed by soaking crystals with phenyl boronic acid. The compound forms a covalent adduct with Ser 117, and the residues of the catalytic triad align themselves in a productive configuration. The σ_A -weighted³⁵ $2F_o - F_c$ electron density map contoured at 1.5σ is shown around the Ser adduct.

letters

Fig. 4 Proposed mechanism for acyl intermediate hydrolysis. The hydrogens have been assigned according to the hydrogen bond donor-acceptors in the active site. The covalent acyl intermediate (top left) is proposed to go through a tetrahedral transition state (top right) on its way to product (bottom left). The boron inhibitor (bottom right) mimics the tetrahedral transition-state of the acyl enzyme. The pro-*R* hydroxyl of the boron-adduct is probably positioned where the attacking water would be in the tetrahedral intermediate. In the product complex, whether the water molecule HOH donates or accepts a hydrogen from Ser 117 is unclear.



Conclusions

We have structurally characterized an efficient cocaine-degrading enzyme, which is a novel variant of the α/β hydrolase fold, and obtained snapshots at two positions along the reaction coordinate. Structure-based mutagenesis is underway to characterize the relative contributions of active site residues towards the overall mechanism of the enzyme.

The structural and biochemical characterization of cocE also has important ramifications for further development of catalytic antibodies against cocaine. Specifically, catalytic antibodies are generally raised against stable transition state analogs. Inhibition studies with cocE will show whether or not such analogs mimic the transition-state of the reaction sufficiently to warrant immunization trials. The trapping of phenyl-boronic acid as a covalent adduct in the crystal structure would suggest that substitution of the benzoyl ester of cocaine with either a boronic ester or acid may yield useful haptens for a reactive immunization-like strategy²¹. Hence, the enzyme structure provides a basis for further antibody engineering.

The cocE enzyme itself has therapeutic potential as an enzyme-based treatment for cocaine overdose. Intravenous administration of this enzyme into a patient experiencing cocaine overdose could result in cocaine clearance within seconds. The precedent for bacterial enzymes as therapeutics has already been established — for example, in the antibody-directed enzyme prodrug therapy (ADEPT), where use of foreign enzymes is requisite for specific prodrug activation at the targeted tumor location²². Two of the most important *in vivo* parameters that need to be addressed are toxicity and plasma lifetime of the enzyme. Short plasma lifetimes would suffice to clear cocaine from the system of a patient experiencing an overdose while avoiding adverse humoral and cellular immune response.

Table 1 Data collection and refinement statistics for cocE

Data collection		Pt λ 1 (f')	Pt λ 2 (f')	Pt λ 3	Product	Adduct
Wavelength (Å)		1.07167	1.07198	0.86095	1.1	0.965
Resolution range (Å) ^a		20.0–2.1 (2.14–2.10)	20.0–2.18 (2.22–2.18)	20.0–2.13 (2.22–2.18)	20.0–1.63 (1.66–1.63)	20.0–1.58 (1.61–1.58)
Number of unique refs.		75,311	77,884	82,174	95,669	90,349
Completeness (%) ^b		91.7 (55.8)	94.3 (77.3)	99.8 (96.7)	97.9 (90.7)	90.0 (71.2)
Redundancy		3.3	3.4	3.6	4.3	5.4
R _{merge} ^{1,2}		6.4 (38.7)	6.5 (46.0)	8.0 (51.1)	5.4 (40.6)	6.4 (51.4)
I/ σ ^c		17.2 (1.9)	17.9 (1.9)	14.9 (2.1)	20.7 (2.0)	24.5 (2.2)
Refinement statistics		Product		Adduct		
Refined residues		570	570	Ramachandran statistics (%)		
Refined waters		436	544	Most favored ^d	88.2	87.8
R _{cryst} ³ (%)		19.0	19.9	Additional allowed	10.1	10.5
R _{free} ⁴ (%)		21.9	21.4	Generously allowed	1.0	0.4
Average B-values (Å ²)				Disallowed	0.6	1.2
Enzyme		21.6	20.6	R.m.s. deviations from ideal geometry		
Waters		30.1	30.8	Bond lengths (Å)	0.008	0.011
Ligand		19.2	19.4	Bond angles (°)	2.1	1.6
				Dihedrals angle (°)	26.4	25.7
				Improper angle (°)	1.4	1.6

¹Numbers in parentheses refer to the highest resolution shell.

²R_{merge} = $(\sum_h \sum_l |I(h) - \langle I(h) \rangle| / \sum_h \sum_l I(h)) \times 100$, where $\langle I(h) \rangle$ is the average intensity of symmetry-related observations of reflections with Bragg index h .

³R_{cryst} = $(\sum_h |F_o - F_c| / \sum_h |F_o|) \times 100$, where F_o and F_c are the observed and calculated structure factors, respectively, for all data.

⁴R_{free} was calculated as for R_{cryst}, but on 5% of data excluded before refinement.

⁵A significant fraction of the residues are located within 5° of the most favored regions of ϕ/ψ space. We felt that 'massaging' these parameters to increase the number of residues in the most favored region would be inappropriate due to the exceptional quality of the electron density.

letters

The plasma lifetime can be lengthened, if necessary, by covalent modification of the enzyme with polyethylene glycol (PEG), which would mask the enzyme from the immune system²³. This structure will help identify and/or predict where such modifications would take place.

Methods

Cloning, expression and crystallization. The cocE enzyme was subcloned into pET-22b(+) (Novagen) and transformed into BL21-Gold (DE3) cells (Stratagene). Soluble enzyme (20 mg l⁻¹) was obtained by overnight induction with 1 mM isopropyl- β -D-thiogalactopyranoside (IPTG) at -23 °C. CocE was ~99% pure after two consecutive rounds of Ni-NTA agarose chromatography (Qiagen). The purified enzyme was assayed for cocaine hydrolysis by monitoring a decrease in absorbance at 240 nm (ref. 24), which is more sensitive than the HPLC method¹. The HPLC method failed to measure the initial rates, leading to an artificially inflated measurement for K_M (ref. 1). Therefore, the kinetic parameters reported here supersede the previous values¹. Subsequently, cocE was dialyzed into 10 mM Tris, pH 7.5, 25 mM NaCl and 10 mM dithiothreitol (DTT) and then concentrated to 40 mg ml⁻¹ for crystallization trials. High quality crystals were grown in 1.5–1.8 M ammonium sulfate in hexagonal space group P6₃22, with $a = b = 106$ Å and $c = 221$ Å, one molecule per asymmetric unit and a Matthew's coefficient of 2.63 Å³ Da⁻¹ (solvent content of 53%).

Data collection, structure solution and refinement. A three-wavelength platinum multiple anomalous dispersion (MAD) data set was collected at SSRL beamline 9-2 from crystals soaked in 0.5 mM platinum (II) (2,2',6',2''-terpyridine). The data were integrated and scaled with HKL2000 (ref. 25), and five platinum sites were identified, using SOLVE²⁶, that were sufficient to phase the entire molecule at 2.5 Å (figure of merit (FOM) = 0.43 for all reflections). Density modification was performed using RESOLVE²⁷ (FOM = 0.69 for all reflections), and the platinum derivative phases were extended to 2.1 Å using DM²⁸. The phases were further extended to 1.63 Å using a data set obtained from a crystal soaked in 1 mM cocaine. ARP/WARP²⁹ then successfully traced and built 97% of the structure. The remaining model was built manually in O³⁰, and the refinement was started in CNS³¹ and finished using SHELXL³², with final $R_{\text{cryst}} = 19.0\%$ and $R_{\text{free}} = 21.9\%$ for all data ($R_{\text{cryst}} = 17.2\%$ and $R_{\text{free}} = 20.0\%$ for $F_o > 4 \sigma$).

Crystals soaked for 5 min in 1 mM phenyl boronic acid resulted in a covalent tetrahedral analog complex (Table 1). The complex was refined using the MAD structure as a starting model, and the covalent serine phenyl-boronic-acid adduct was built in Insight2000 (ACCELRYs). We subsequently modified the protein parameter and topology files in CNS³¹ to include our Ser adduct as a new type of residue. This modification gave stable refinement and maintained acceptable geometry with final $R_{\text{cryst}} = 19.9\%$ and $R_{\text{free}} = 21.4\%$.

Coordinates. The atomic coordinates and structure factors for the tetrahedral intermediate analog adduct and product complexes have been deposited in the Protein Data Bank (accession codes 1JU3 and 1JU4, respectively).

Acknowledgments

The authors thank the Advanced Light Source at Berkeley and the Stanford Synchrotron Radiation Laboratory (SSRL), in particular A. Gonzalez, for advice on MAD data collection. We are also indebted to R. Stanfield for assistance on synchrotron trips and M. Elsiger for helpful discussions. Support was provided by the National Institutes of Health (I.A.W.), the Biotechnology and Biological Sciences Research Council (N.C.B.), the Howard Hughes Medical Institute (J.M.T.), and the Bernie Gilula Foundation (N.A.L.).

Correspondence should be addressed to I.A.W. email: wilson@scripps.edu

Received 27 August, 2001; accepted 9 November, 2001.

1. Bresler, M.M., Rosser, S.J., Basran, A. & Bruce, N.C. *Appl. Environ. Microbiol.* **66**, 904–908 (2000).
2. Inaba, T., Stewart, D.J. & Kalow, W. *Clin. Pharmacol. Ther.* **23**, 547–552 (1978).
3. Brzezinski, M.R., Abraham, T.L., Stone, C.L., Dean, R.A. & Bosron, W.F. *Biochem. Pharmacol.* **48**, 1747–1755 (1994).
4. Pindel, E.V. *et al.* *J. Biol. Chem.* **272**, 14769–14775 (1997).
5. Britt, A.J., Bruce, N.C. & Lowe, C.R. *J. Bacteriol.* **174**, 2087–2094 (1992).
6. Carroll, F.L., Howell, L.L. & Kuhar, M.J. *J. Med. Chem.* **42**, 2721–2736 (1999).
7. Carrera, M.R.A. *et al.* *Nature* **378**, 121–130 (1995).
8. Meis, B. *et al.* *Proc. Natl. Acad. Sci. USA* **95**, 10176–10181 (1998).
9. Landry, D.W., Zhao, K., Yang, G.X.Q., Glickman, M. & Georgiadis, T.M. *Science* **259**, 1899–1901 (1993).
10. Lu, G. *DOMID* <http://biolinfol.mbfys.lu.se/Domid/> (1999).
11. Ollis, D.L. *et al.* *Protein Eng.* **5**, 197–211 (1992).
12. Nardini, M. & Dijkstra, B.W. *Curr. Opin. Struct. Biol.* **9**, 732–737 (1999).
13. Holkinheimo, P., Goldman, A., Jeffries, C. & Ollis, D.L. *Structure* **7**, R141–R146 (1999).
14. Connolly, M.L. *J. Appl. Crystallogr.* **16**, 548–558 (1983).
15. Sheriff, S., Hendrickson, W.A. & Smith, J.L. *J. Mol. Biol.* **197**, 273–296 (1987).
16. Derewenda, Z.S. & Wei, Y. *J. Am. Chem. Soc.* **117**, 2104–2105 (1995).
17. Holm, L. & Sander, C. *J. Mol. Biol.* **233**, 123–138 (1993).
18. Wei, Y. *et al.* *Structure* **6**, 511–519 (1997).
19. Fulop, V., Booskei, Z. & Polgar, L. *Cell* **94**, 161–170 (1998).
20. Egloff, M.P. *et al.* *Biochemistry* **34**, 2751–2762 (1995).
21. Wirsching, P., Ashley, J.A., Lo, C.-H.L., Janda, K.D. & Lerner, R.A. *Science* **270**, 1775–1782 (1995).
22. Melton, R.G. & Sherwood, R.F. *J. Natl. Cancer Inst.* **88**, 153–165 (1996).
23. Abuchowski, A., Davis, F. & Davis, S. *Cancer Treat. Rep.* **65**, 1077–1081 (1981).
24. Xie, W. *et al.* *Mol. Pharmacol.* **55**, 83–91 (1999).
25. Otwinowski, Z. & Minor, W. *Methods Enzymol.* **276**, 307–326 (1997).
26. Terwilliger, T.C. & Berendzen, J. *Acta Crystallogr. D* **53**, 571–579 (1997).
27. Terwilliger, T.C. *Acta Crystallogr. D* **56**, 965–972 (2000).
28. Cowtan, K.D. & Main, P. *Acta Crystallogr. D* **52**, 43–48 (1996).
29. Lamzin, V.S. & Wilson, K.S. *Acta Crystallogr. D* **49**, 129–149 (1993).
30. Jones, T.A., Zou, J.Y., Cowan, S.W. & Kjeldgaard, M. *Acta Crystallogr. A* **47**, 110–119 (1991).
31. Brunger, A.T. *et al.* *Acta Crystallogr. D* **54**, 905–921 (1998).
32. Sheldrick, G.M. & Schneider, T.R. *Methods Enzymol.* **277B**, 319–343 (1997).
33. Esnouf, R.M. *Acta Crystallogr. D* **55**, 938–940 (1999).
34. Merritt, E.A. & Murphy, M.E.P. *Acta Crystallogr. D* **50**, 869–873 (1994).
35. Read, R.J. *Acta Crystallogr. A* **42**, 140–149 (1986).

Engineering proteins for thermostability: the use of sequence alignments versus rational design and directed evolution

Martin Lehmann* and Markus Wyss†

With the advent of directed evolution techniques, protein engineering has received a fresh impetus. Engineering proteins for thermostability is a particularly exciting and challenging field, as it is crucial for broadening the industrial use of recombinant proteins. In addition to directed evolution, a variety of partially successful rational concepts for engineering thermostability have been developed in the past. Recent results suggest that amino acid sequence comparisons of mesophilic proteins alone can be used efficiently to engineer thermostable proteins. The potential benefits of the underlying, semirational 'consensus concept' are compared with those of rational design and directed evolution approaches.

Addresses

F Hoffmann-La Roche Ltd., Vitamins and Fine Chemicals Division, Department VFB, Building 203, CH-4070 Basel, Switzerland

*e-mail: martin.lehmann@roche.com

†e-mail: markus.wyss@roche.com

Current Opinion in Biotechnology 2001, 12:371–375

0958-1669/01/\$ – see front matter

© 2001 Elsevier Science Ltd. All rights reserved.

Introduction

Because natural enzymes are adapted to their particular function in a living cell, in most instances they are poorly suited for industrial applications that often encounter extremes of pH, temperature and/or salinity. Both for broadening the industrial applicability of enzymes and for furthering our understanding of protein structure/function relationships, protein engineering is currently a very active area of research. Interest in this subject area has further increased following the emergence of high-throughput directed evolution techniques, which can be seen as test tube models of natural evolution. This strong interest in protein engineering is reflected in two recent publications dedicated entirely to this topic [1,2]. The present review focuses on engineering proteins for thermostability and compares the advantages and benefits of classical rational design principles, directed evolution, and a 'consensus approach' that extracts valuable information from sequence comparisons of homologous (mesophilic) proteins alone.

Rational design principles

The stability of a protein is determined by a multitude of both local and long-range interactions. In order to achieve pronounced thermostabilisation, several substitutions each with a relatively small effect usually need to be combined in a multiple mutant. This endeavour is facilitated by the fact that in many cases the thermostabilisation effects of individual mutations are independent and nearly additive (e.g., [3–8]).

One strategy for identifying thermostabilising mutations involves the comparison of more stable proteins with less stable ones, with the goal of identifying amino acid sequence patterns that correlate with thermostability [9]. In a systematic study, Perl *et al.* [10**] compared the cold shock proteins from the thermophile *Bacillus caldolyticus* and the mesophile *Bacillus subtilis*, which differ in only 12 out of 67 residues but display a considerable difference in stability (15.8 kJ/mol at 70°C). Site-directed mutagenesis of all 12 residues in the *Bacillus caldolyticus* enzyme revealed that the difference in thermostability can be fully accounted for by only two amino acid substitutions (Glu3Arg, Glu66Leu) on the surface of the molecule. In this illustrative example, with small proteins that exhibit high homology and a pronounced difference in thermostability, it is noted that less than 20% of the amino acid substitutions actually contribute to the difference in stability. This observation highlights the problem of identifying the relevant thermostabilising mutations in larger and less homologous sets of proteins.

In other studies published last year, higher thermostability was suggested to correlate with more proline and less asparagine and glutamate residues [11], more arginines and tyrosines but less cysteine and serine residues, and with increased numbers of salt bridges and sidechain–sidechain hydrogen bonds [12]. In addition, higher thermostability correlated with a larger fraction of residues in α helices and with more arginine and less proline, cysteine and histidine residues in those α helices [12]. The information accumulated to date indicates that nature relies on no single strategy for stabilisation (see [13]), and as a result many publications in this area arrive at different and even conflicting conclusions. The availability of more complete genome sequences may cause the reliability of the correlations to improve, but sequence statistics of this kind, when taken alone, are still unlikely to provide useful information for the accurate prediction of thermostabilising mutations.

Thermostabilising mutations either increase the thermodynamic stability of a protein (i.e., they increase the free-energy difference between the unfolded and the folded state) or they decrease the rate of unfolding by increasing the free-energy difference between the folded state and the transition state of unfolding (see [14]). To achieve either of these, various rational concepts have been proposed: to decrease the entropy of the unfolded state by introducing additional disulfide bridges or by X→Pro mutations; to increase α -helix propensity by Gly→Ala substitutions or by stabilisation of α -helix macrodipoles; to improve electrostatic interactions between charged surface residues by introducing additional salt bridges or even salt-bridge networks, or by predicting

thermostabilising mutations on the basis of calculations of electrostatic potentials. Again, because the predictive power of these rational concepts is rather limited [15], the supposed thermostabilising mutations need to be tested individually by site-directed mutagenesis. This reduces speed and throughput and thereby limits the sequence space amenable to testing.

Directed evolution

The term 'directed evolution' encompasses a series of experimental techniques that reproduce, on an accelerated timescale in the test tube, the evolution of natural diversity and environmental adaptation. This is achieved through mutation and recombination and by giving the process a 'direction' towards the optimisation of one or more properties of interest. Either a selective pressure is applied, or in each round of mutagenesis and/or recombination the library of variants obtained is screened for the desired trait.

One frequently applied strategy comprises repeated rounds of random mutagenesis, starting with a given parent gene of interest. After each round, the best mutant(s) is (are) selected and used as parent sequence(s) in the following round of random mutagenesis. Typically, rather low mutation frequencies are employed to suppress accumulation of neutral or even deleterious mutations [13,14]. In a recent example of this type, two cycles of random mutagenesis and screening yielded two mutants of phospholipase A1 with six and seven amino acid substitutions each; the mutants displayed increases in their temperature of half-inactivation of 7 and 11°C, respectively [16].

Since the introduction of DNA shuffling in 1994 [17,18], most directed evolution experiments encompass one or more cycles of recombination between a set of homologous sequences. These experiments may use a set of improved variants of a given enzyme, obtained by random mutagenesis, or a set of naturally occurring, homologous genes isolated from wild-type organisms. DNA shuffling was shown to be very powerful in recombining favourable mutations and in eliminating deleterious or neutral mutations.

Combinations of random mutagenesis and/or DNA (family) shuffling have been used successfully to engineer protein thermostability [19–21,22*,23–27]. In these examples, increases in 'thermostability' of 14–20°C were attained with 7–19 amino acid substitutions (relative to the parent protein), which were accumulated over up to eight generations of directed evolution. Two important conclusions can be drawn from these results. Firstly, the increases in thermostability obtained by directed evolution have, as yet, been no more impressive than the best examples of rational design (e.g., [28,29]), although they may have required less time and/or less effort. Secondly, despite improved equipment that allows higher throughput and increased automation, the sequence space amenable to testing is still rather limited. Key prerequisites for larger steps and longer walks in sequence space include the efficient elimination

of neutral and deleterious mutations, higher frequencies of recombination between homologous genes, recombination events in stretches of low amino acid sequence identity (e.g., [30**]), and powerful selection tools for improved traits. Although longer walks in sequence space might not actually be required for increases in thermostability of 20–30°C, they may be desirable for more dramatic increases in stability and will definitely be advantageous for optimising other enzyme properties, such as reaction mechanism or substrate specificity.

The consensus concept

A third, semirational approach for engineering thermostability is based on the hypothesis that at a given position in an amino acid sequence alignment of homologous proteins, the respective consensus amino acid contributes more than average to the stability of the protein than the nonconsensus amino acids. Consequently, substitution of nonconsensus by consensus amino acids may be a feasible approach for improving the thermostability of a protein.

Although Pantoliano *et al.* [3] were the first to apply the 'consensus concept', showing that the consensus-type mutation Met50Phe increased the unfolding temperature (T_m) of subtilisin BPN' by 1.8°C, it was five years later that Steipe *et al.* [31] offered a possible theoretical explanation, based on statistical thermodynamics, for the feasibility of this approach. Steipe *et al.* [31] predicted ten individual, stabilising mutations for the immunoglobulin variable V_L domain, of which six were indeed stabilising. Further experiments on the immunoglobulin V_L and V_H domains as well as on a catalytic single-chain Fv fragment confirmed these results and showed that the combination of a set of stabilising mutations in a multiple variant provided additive thermostabilising effects [5,32–34].

In addition to applying the consensus concept to sets of homologous amino acid sequences, Steipe and coworkers also tried to apply the concept to structural motifs. The hypothesis that the most frequently occurring residues in specific positions of β -turn motifs increase the folding stability of a protein was experimentally confirmed for an immunoglobulin V_L domain [32]. In contrast, when the concept was applied to surface-exposed loops and turns, which were assumed to be stabilised primarily by local interactions, only less stable variants were observed, suggesting that this particular application of the consensus concept does not allow reliable prediction of thermostabilising mutations.

A set of systematic studies of the consensus concept has also been provided by Fersht's group. In one study [35], p53 homologues from 23 species were aligned, and 20 nonconsensus residues of human p53 were mutated individually to the respective consensus residue. The changes in stability ranged from +1.27 to -1.49 kcal/mol, and the theoretical sum of the stability changes of the 20 individual mutations was no more than -0.48 kcal/mol.

Four stabilising mutations were combined in a quadruple mutant (Met133Leu, Val203Ala, Asn239Tyr, Asn268Asp), which was stabilised by 2.65 kcal/mol and displayed an increase in T_m of 5.6°C.

In two other studies [36*,37], the consensus concept was applied to GroEL minichaperones (i.e., fragments of GroEL capable of facilitating protein folding). A sequence alignment of 130 sequences of homologous chaperonin 60 proteins revealed that 31 (out of approximately 150) amino acids of the *Escherichia coli* minichaperone GroEL(193–345) occur with a frequency of <35% at their respective positions in the alignment. Each of these residues was replaced individually with the most frequently occurring residue(s), yielding a total of 34 single mutants which displayed differences in stability relative to wild type ranging from +1.55 to –1.78 kcal/mol. Summing up the stability effects of all the individual mutations yields a theoretical overall difference in stability of –0.30 kcal/mol, which is unlikely to be significantly different from zero. Among the 34 amino acid substitutions analysed, 13 were stabilising (38% success rate), five were neutral, and 16 were destabilising. However, when considering for substitution only those residues that occur with a frequency of <20% (rather than <35%) at the respective position in the alignment, 13 out of 18 mutations were stabilising (72% success rate), two were neutral, and only three were destabilising. Two multiple variants each combining a set of six stabilising mutations were stabilised by 6.99 and 6.15 kcal/mol, and displayed T_m increases of 18.6 and 14.2°C relative to wild-type GroEL(193–345).

No attempt has been made by either Steipe's or Fersht's group to apply the consensus concept over the entire sequence of a protein. In fact, on the basis of the data of Nikolova *et al.* [35] and Wang *et al.* [36*] (see above), one might be inclined to assume that such an attempt will fail, because stabilising and destabilising mutations may counterbalance each other. Very much to the contrary are the results of Lehmann *et al.* [38**] describing the design of a consensus phytase. In this study an appropriate computer program was used to calculate an entire consensus sequence from 13 homologous amino acid sequences of wild-type phytases from mesophilic fungi. A synthetic gene was constructed from the consensus sequence, and recombinant expression of this gene gave rise to a consensus phytase (consensus phytase-1) that was 15–26°C more thermostable than all of its parents [38**]. Subsequently, incorporation of additional wild-type sequences in the alignment yielded consensus phytase-10, which displayed 32 amino acid differences relative to consensus phytase-1 and a further 7.4°C increase in T_m (M Lehmann *et al.*, unpublished data). Both consensus phytases differ in at least 80 amino acids from any of their parents. This clearly shows that the consensus concept allows multiple amino acid exchanges to be combined in a single step in order to provide a significantly improved variant of the enzyme.

Remarkably, the overall correlation often observed between increases in thermostability and decreases in catalytic turnover rate at ambient temperature does not apply to the consensus phytases: instead, the opposite is true (M Lehmann *et al.*, unpublished data).

In summary, the consensus approach allows much larger steps in sequence space than most directed evolution techniques reported to date. In addition, at present, increases in thermostability of up to 30°C seem possible in a single shot.

Conclusions

In the case of engineering proteins for thermostability, researchers are in the enviable situation of being able to choose between three different, apparently equally successful, strategies: rational design, directed evolution, and the construction of (semirational) synthetic consensus genes. Rather than playing off one approach against the others, future efforts should focus on how to best combine these alternative approaches in order to access a larger fraction of the total sequence space and to approach the global optimum for a given trait in a consistent and direct way. A first successful study of this type has been reported by Cherry *et al.* [39*]. In their endeavour to improve the stability of a haem peroxidase for laundry applications, four mutations were rationally designed: one to increase the enzyme's thermostability and three to increase resistance to oxidative damage. The combination of these mutations with favourable amino acid exchanges identified in directed evolution experiments yielded a final mutant with 174 times the thermal stability and 100 times the oxidative stability of the wild-type haem peroxidase.

We expect that the consensus concept will prove to be, and will thus be generally recognised as, an additional valuable instrument in our tool box for designing more thermostable enzymes for all types of industrial applications. Construction of a single consensus protein has the potential to provide a similar increase in thermostability as many months or years of rational design or as directed evolution campaigns involving multiple generations and analysis of thousands of mutants.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- of special interest
- ** of outstanding interest

1. Dalboge H, Borchert IV (Guest Eds): Protein engineering of enzymes. *Biochim Biophys Acta* 2000; 1543:203–455
2. Arnold FH (Guest Ed): Evolutionary protein design. *Adv Protein Chem* 2001; 55:1–438
3. Partoliano MW, Whillow M, Wood JF, Dodd SW, Hardman KD, Rollence ML, Bryan PN: Large increases in general stability for subtilisin BPN' through incremental changes in the free energy of unfolding. *Biochemistry* 1989; 28:7205–7213
4. Serrano L, Day AG, Fersht AR: Step-wise mutation of barnase to binase. A procedure for engineering increased stability of proteins and an experimental analysis of the evolution of protein stability. *J Mol Biol* 1993; 233:305–312

Rational design and engineering of therapeutic proteins

Shannon A. Marshall, Greg A. Lazar, Arthur J. Chirino
and John R. Desjarlais

An increasing number of engineered protein therapeutics are currently being developed, tested in clinical trials and marketed for use. Many of these proteins arose out of hit-and-miss efforts to discover specific mutations, fusion partners or chemical modifications that confer desired properties. Through these efforts, several useful strategies have emerged for rational optimization of therapeutic candidates. The controlled manipulation of the physical, chemical and biological properties of proteins enabled by structure-based simulation is now being used to refine established rational engineering approaches and to advance new strategies. These methods provide clear, hypothesis-driven routes to solve problems that plague many proteins and to create novel mechanisms of action. We anticipate that rational protein engineering will shape the field of protein therapeutics dramatically by improving existing products and enabling the development of novel therapeutic agents.

Shannon A. Marshall
Greg A. Lazar
Arthur J. Chirino
John R. Desjarlais*
Xencor
111 W. Lemon Ave.
Monrovia
CA 91016, USA
tel: +1 626 305 5900
fax: +1 626 305 0350
*e-mail: jrd@xencor.com

▼ The exquisite specificity of biological therapeutics for their clinical targets has led to their continued development and application as medicines, despite competition from small molecule drugs. Several engineered protein therapeutics are currently being marketed (Table 1), and the annual sales of protein therapeutics are projected to exceed US\$59 billion in 2010, which is twice the revenue generated in 2001 (<http://www.pharmafile.com/Pharmafocus/Features/feature.asp?fid=281>) [1]. For well-validated targets, naturally occurring protein interaction partners constitute preselected 'lead' compounds with high affinity and specificity. However, because natural proteins are not evolved for utilization as drugs, lead optimization is frequently beneficial for development of a protein therapeutic. Modifications can influence the

mechanism of action, side effects and efficacy, and satisfy practical constraints such as production costs, intellectual property and dosing frequency.

A variety of strategies have emerged for modulating protein properties, such as efficacy, stability, specificity, immunogenicity and pharmacokinetics (PK). Mechanisms for altering these properties include manipulation of primary structure, incorporation of chemical and post-translation modifications and utilization of fusion partners. The most common route to optimization is site-directed mutagenesis, which is often performed in a brute force or trial-and-error manner. A smaller number of examples exist whereby semirational application of diversity methods, such as phage display, has been used to optimize a therapeutic candidate. Important recent developments are the creation and successful application of rational protein design methods and the determination of an increasing number of high-resolution protein structures.

For the purposes of this review, we define rational protein engineering as the hypothesis-driven manipulation of protein sequence and/or composition. The controlled modification of specific biophysical properties of proteins can potentially impact a variety of therapeutic features (Table 2). An important subset of rational engineering methods consists of approaches that utilize high-resolution, 3D structure information. The most sophisticated of these methods offers an extraordinary level of control over protein sequence and structure, a mechanism to explore sequence combinations that extends far beyond natural diversity, and the ability to couple multiple constraints algorithmically for

Table 1. Engineered protein therapeutics on the market^a

Name	Family	Company	Indication	Modification	Property
Proleukin [®] (aldesleukin)	IL-2	Chiron	Cancer	Mutated free cysteine	Decreased aggregation; improved bioavailability
Betaseron [®] (interferon beta-1b)	IFN- β	Berlex/Chiron	Multiple sclerosis	Mutated free cysteine	Decreased aggregation
Humalog [®] (insulin lispro)	Insulin	Eli Lilly	Diabetes	Monomer not hexamer	Fast acting
NovoLog [®] (insulin aspart)	Insulin	Novo Nordisk	Diabetes	Monomer not hexamer	Fast acting
Lantus [®] (insulin glargine)	Insulin	Aventis	Diabetes	Precipitates in dermis	Sustained release
Enbrel [®] (etanercept)	TNF receptor	Immunex/ Amgen/Wyeth	Rheumatoid arthritis	Fc fusion	Longer serum half-life; increased avidity
Ontak [®] (denileukin diftitox)	Diphtheria toxin-IL-2	Seragen/Ligand	Cancer	Fusion	Targets cancer cells
PEG-Intron [®] (peginterferon alfa-2b)	IFN- α	Schering-Plough	Hepatitis	PEGylation	Increased serum half-life; weaker receptor binding
PEGasys [®] (peginterferon alfa-2a)	IFN- α	Roche	Hepatitis	PEGylation	Increased serum half-life; weaker receptor binding
Neulasta [™] (pegfilgrastim)	G-CSF	Amgen	Leukopenia	PEGylation	Increased serum half-life
Oncaspar [®] (pegaspargase)	Asparaginase	Enzon	Cancer	PEGylation	Decreased immunogenicity; increased serum half-life
Aranesp [®] (darbepoetin alfa)	Epo	Amgen	Anemia	Additional glycosylation sites	Increased serum half-life; weaker receptor binding
Somavert [®] (pegvisomant)	Growth hormone	Genentech/ Seragen/ Pharmacia	Acromegaly	PEGylation; binding site mutations	Novel mode of action; increased serum half-life

Chiron (<http://www.chiron.com>); Berlex (<http://berlex.com>); Eli Lilly (<http://www.lilly.com>); Novo Nordisk (<http://www.novonordisk.com>); Aventis (<http://www.aventis.com>); Immunex/Amgen (<http://www.amgen.com>); Wyeth (<http://www.wyeth.com>); Seragen/Ligand (<http://www.ligand.com>); Schering-Plough (<http://www.sch-plough.com>); Roche (<http://www.roche.com>); Enzon (<http://www.enzon.com>); Genentech (<http://www.genentech.com>); Pharmacia (<http://www.pharmacia.com>).

^aAbbreviations: G-CSF, granulocyte-colony stimulating factor; IFN- α , interferon α ; IL-2, Interleukin 2; PEG, polyethylene glycol; TNF, tumor necrosis factor.

simultaneous optimization of several protein properties. Furthermore, proven hypotheses can be reapplied to additional protein systems, thus saving discovery cost and time. Rational methods can be distinguished from those that rely on random sequence perturbations or combinations, such as the class of optimization techniques referred to as directed evolution, although some implementations of these methods have a rational component [2,3].

Physicochemical properties

The physical and chemical properties of protein therapeutics significantly determine their performance during development, manufacturing and clinical use. Many therapeutically

interesting proteins are naturally expressed at low concentrations and are degraded rapidly. By contrast, fully developed protein therapeutics require high levels of solubility as well as retention of activity through purification, formulation, storage and administration. Several rational design and engineering strategies, such as those highlighted in Figure 1, have been developed to improve properties such as solubility and stability while maintaining desired biological activity.

Stability

Protein therapeutics are exposed to a variety of stresses that can cause protein unfolding or degradation. Using rational optimization methods, proteins can be re-engineered

Table 2. The biophysical properties of proteins that can be optimized to obtain desired therapeutic outcomes^a

	Enable discovery	Mechanism of action	Pharmaco-kinetics	Immuno-genicity	Route of administration	Cost of goods	Shelf life	Intellectual property
Stability	x				x	x	x	
Solubility	x		x	x	x	x	x	
Receptor binding affinity and specificity		x	x					
MHC binding affinity			x	x				
Oligomerization state		x	x		x			
Chemical modifications		x	x	x	x		x	x
Posttranslational modifications		x	x	x			x	x
Sequence diversity								x
Conformational state		x						

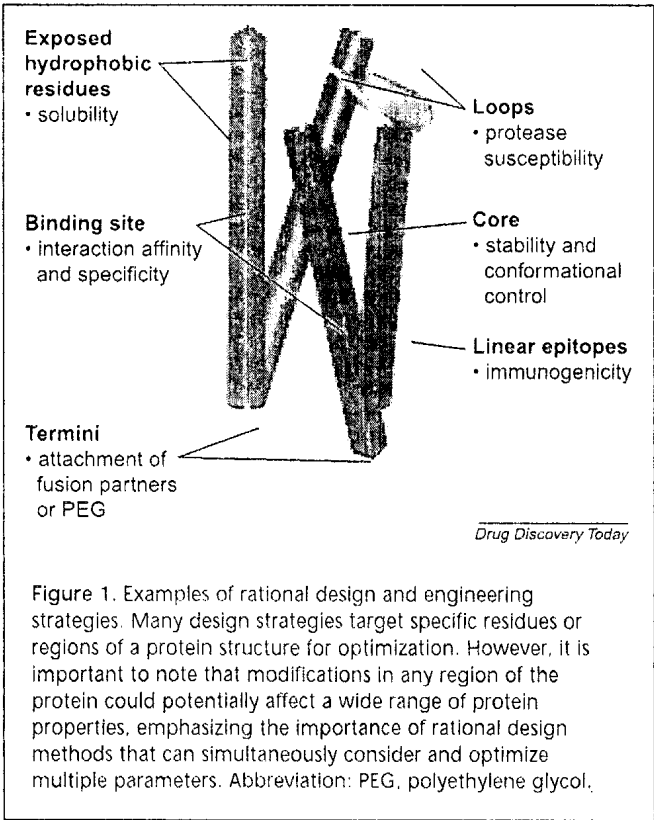
^aAbbreviation: MHC, major histocompatibility complex

such that their structure and activity are substantially more robust with respect to protease exposure, oxidative stress and changes in temperature, pH and solution conditions. One simple stabilization strategy is to replace free cysteines, thereby preventing the formation of unwanted intermolecular or intramolecular disulfide bonds. Cysteine to serine mutations have been introduced successfully into several therapeutic proteins, including granulocyte

colony-stimulating factor (G-CSF) and interferon (IFN) β 1b, resulting in a longer shelf life [4,5]. Cysteine to serine mutations have also been shown to increase the half-life of human fibroblast growth factor (FGF) [6]. Interestingly, each of the three FGF mutations decreases the thermal stability of the protein, probably because the introduced serines are substantially desolvated and are not positioned to form intramolecular hydrogen bonds. Rational approaches can identify the amino acids that are more precisely compatible with the local structural environment.

Dramatic improvements in the global stability of a protein can be obtained by optimizing intramolecular interactions. Early examples used rational computational design methods to optimize packing interactions and hydrophobic burial in the protein core [7-9]. Optimizing secondary structure propensity, hydrogen bonds and electrostatic interactions can also improve protein stability substantially [10-12]. More recently, these principles have been applied to the clinically relevant proteins G-CSF and human growth hormone (hGH) by using Protein Design Automation[®] (PDA[™]) technology (Box 1). The designed hGH variants are active in cell proliferation studies and are up to 16°C more thermostable than the wild type protein [13]. Optimized G-CSF variants with 10-14 mutations display enhanced thermal stability and five to tenfold increases in shelf life while maintaining the desired biological activity [14].

An additional stabilization strategy is reduction of proteolytic susceptibility. If a specific site in the protein is known to be especially prone to proteolysis, it can be modified so that it no longer matches the substrate specificity of the putative protease. The protease cleavage sites are often



Box 1. Xencor's PDA™ technology: a state of the art rational engineering platform

PDA™ technology, originally developed at Caltech [10,66,67] and further optimized at Xencor [13,14], couples computational design algorithms that generate quality sequence diversity with experimental high-throughput screening to discover proteins with improved properties. The computational component uses atomic level scoring functions, side chain rotamer sampling and advanced optimization methods to capture the relationships between protein sequence, structure and function accurately. Calculations begin with the 3D structure of the protein and a strategy to optimize one or more properties of the protein. PDA™ technology then explores the sequence space comprising all pertinent amino acids (including unnatural amino acids, if desired) at the positions targeted for design.

This is accomplished by sampling conformational states of allowed amino acids and scoring them using a parameterized and experimentally validated function that describes the physical and chemical forces governing protein structure. Powerful combinatorial search algorithms are then used to search through the initial sequence space, which can constitute 10^{50} sequences or more, and quickly return a tractable number of sequences that are predicted to satisfy the design criteria. Useful modes of the technology span from combinatorial sequence design to prioritized selection of optimal single site substitutions. PDA™ technology has been applied to numerous systems including important pharmaceutical and industrial proteins and has a demonstrated record of success in protein optimization.

located in flexible loops; therefore, another approach is to introduce mutations that decrease flexibility. Thrombolytics are a class of protein therapeutics for which proteolytic susceptibility is especially important because many clotting factors are both activated and inactivated by specific proteases. For example, an engineered variant of coagulation factor VIIIa with increased resistance to proteolytic inactivation was generated by mutating two arginines required for cleavage by thrombin, factor Xa and activated protein C [15].

Solubility

Protein therapeutics are typically expressed, formulated and administered at high concentrations. Under such conditions many proteins form inclusion bodies during expression or aggregates after formulation. Improving the solubility of a protein can facilitate discovery efforts, whereas enabling soluble prokaryotic expression can reduce production costs dramatically and increase yields. It is far more critical to ensure the solubility of a protein therapeutic once it is administered. Aggregation can cause decreased activity, decreased bioavailability and increased immunogenicity. Several strategies have been applied successfully to reduce protein aggregation and enable soluble expression. Replacement of unpaired cysteine residues can prevent the formation of unwanted intermolecular disulfide bonds, as described above. Post-translational and chemical modifications, which are discussed in a later section, can also help to prevent aggregation. Substituting exposed nonpolar residues with polar residues can enable soluble expression and improve the solubility of the purified protein. This strategy was applied successfully to the A1 domain of cholera toxin, a powerful adjuvant. Of the six variants produced, one retained full biological activity

and stability and also displayed a significant improvement in solubility [16]. Altering the net charge and isoelectric point (pI) of a protein can also affect its solubility. For example, a single chain antibody targeting renal cell carcinoma was altered to increase solubility by adding five glutamic acid residues to the C-terminus, thus lowering the pI from 7.5 to 6.1 [17]. Although there are a few examples of rational solubility engineering, the majority of the published successes in solubility optimization have been anecdotal. Until now solubility obstacles have been more or less considered to be formulation problems that can be surmounted with an exhaustive protein chemistry effort. We anticipate that systematic structure-guided optimization efforts will lead to the emergence of well-defined strategies that consistently yield proteins with improved solubility and minimally perturbed structure and function.

Pharmacokinetics

The first generation of protein therapeutics frequently suffered from poor PK. As our understanding of protein clearance processes improves, it becomes possible to rationally modify proteins to tailor their PK profiles. Properly controlling the serum concentration of a therapeutic protein over time can lead to improved efficacy and decreased side effects. In fact, improvements in PK properties can be so vital to the efficacy of a protein drug that they are often made at the expense of specific activity. In addition to eliminating proteolytic susceptibility (see above), several strategies have been developed to alter PK, including polyethylene glycol (PEG) attachment (PEGylation), glycosylation, fusion to proteins with long serum half-lives, alteration of oligomerization state and modulation of receptor-mediated uptake and