B

Industria y Comercio

SUPERINTENDENCIA

Delegatura de Propiedad Industrial

División de Nuevas Creaciones

PCT

SOLICITUD

PATENTE DE INVENCION

21. **EXPEDIENTE No.** 04-568841

54. **TITULO** *ACIDO NUCLEICO DE ADENOVIRUS DE SIMIO Y SECUENCIAS DE AMINOACIDOS, VECTORES QUE CONTIENEN LOS MISMOS Y METODOS DE USO.*

51. **CLASIFICACION INTERNACIONAL** A61K 47/42

71. **SOLICITANTE** *THE TRUSTEES OF THE UNIVERSITY OF PENNSYLVANIA*

DOMICILIO *PHILADELPHIA, PENNSYLVANIA, ESTADOS UNIDOS DE AMERICA*

74. **APODERADO** *JIMENA ESCOBAR URIBE*

22. **BOGOTA, D.C.,** *JUNIO 17 , 2004*

(FORMA P-10)

Industria y Comercio
SUPERINTENDENCIA

SUPERINDUSTRIA Y COMERCIO
Radicacion : 04056841 00000000 Folios: 418
Fecha (AMD): 2004-05-17 16:14:24
Tramite : 011 PCT-PATENT 379 FASE NACI 411 PRESENTAC
Dependencia: 2000 DIVISION DE NUEVAS CREACIONES

FORMULARIO ÚNICO DE SOLICITUD DE PATENTE
PCT
ENTRADA EN FASE NACIONAL

1 SOLICITUD DE:

[X] Patente de Invención [ ] Patente de Modelo de Utilidad

[ ] Capítulo I. [X] Capítulo II.

2 SOLICITANTE (71)	Nombre: **The Trustees of the University of Pennsylvania** Dirección: 3160 Chestnut Street, Suite 200, Philadelphia, Pennsylvania 19104 Nacionalidad o Domicilio: Philadelphia, Pennsylvania , Estados Unidos de América. Lugar de Constitución : Pennsilvania, Estados Unidos de América. Teléfono: Fax: E-mail	**IDENTIFICACIÓN** C.C. [ ] NIT [ ] C.E. [ ] Otro [ ] **Número**
3 REPRESENTANTE O APODERADO (74)	Nombre: **JIMENA ESCOBAR URIBE** Dirección: Cra. 11A No. 94-76 (Of. 305) Teléfono: 6 162051/61 Fax: 6 161889 E-mail: tm-patent@escobaruribe.com	**IDENTIFICACIÓN** C.C. [X] NIT [ ] C.E. [ ] Otro [ ] **Número 39.779.452 T.P.68.228**
4 INVENT R (ES) (72)	Nombre: **VER ANEXOS** Dirección: Nacionalidad o domicilio:	

5 PCT (86)

Solicitud Internacional No. **PCT/US02/33645** Fecha **Noviembre 20, 2002**

Publicación Internacional No. **WO 03/046124 A2** Fecha **Junio 5, 2003**

WO 03/046124 A3 Fecha **Junio 5, 2003**

6 TITULO (54) ÁCIDO NUCLÉICO DE ADENOVIRUS DE SIMIO Y SECUENCIAS DE AMINOÁCIDOS, VECTORES QUE CONTIENEN LOS MISMOS Y MÉTODOS DE USO.

7 Clasificación Internacional (51) C12N 7/00, A61K 47/42

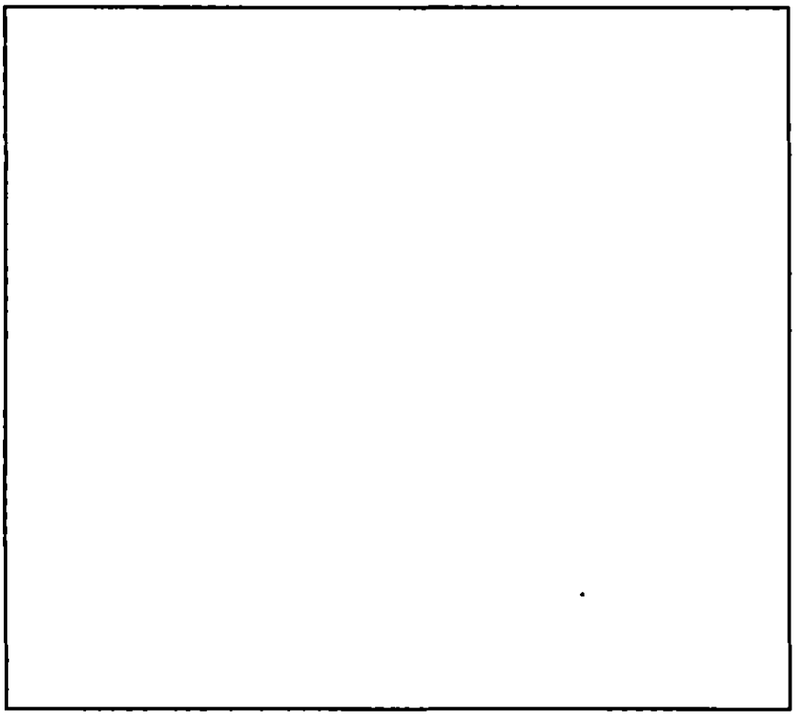
8 Pri ridad SI [X] NO [ ]	(33) País de Origen	(32) Fecha	(31) Número de Solicitud
	Estados Unidos de América	Noviembre 21, 2001	60/331,951
	Estados Unidos de América	Marzo 22, 2002	60/366,798

9 Comprobante de pago No. 04 - 41.543 **Fecha** 25 de mayo de 2004

10 **ANEXOS**

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

11 **FIGURA CARACTERÍSTICA**



12 **Solicito la concesión de la patente**

NOMBRE: JIMENA ESCOBAR URIBE

FIRMA:

C.C. 39.779.452 de Usaquén T.P. 68.228

2670/PCT

3670

2

SUPERINDUSTRIA Y COMERCIO
Radicacion : 04056841 00000000 Folios: 418
Fecha (AMD): 2004-06-17 16:14:34
Tramite : 011 PCT-PATENT 379 FASE NACI 411 PRESENTAC
Dependencia: 2020 DIVISION DE NUEVAS CREACIONES

NIT : 800176089-2

RECIBO OFICIAL DE CAJA : 04 - 41,543
FECHA : MAYO 25 DE 2004

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

DEPOSITANTE	TIPO PAGO	BANCO	CUENTA	No. PAGO	FECHA PGO	Vr. PAGO
ESCOBAR URIBE ASOCIADOS	CONSIGNACION	BANCO POPULAR	050-00110-6	4992974	25/05/2004	5,178,000.00

***** C O N C E P T O *****

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

SON: CUATROCIENTOS VEINTIDOS MIL PESOS

RESPONSABLE : _____

RECIBO DE CAJA APLICADO AL EXPEDIENTE No._____

ANEXOS

INVENTORES

1.) **WILSON, James, M.**
(1350N. Avignon Drive, Gladwyne, Pennsylvania 19035, Estados Unidos de América);

2.) **GAO, Guangping**
(408 Yorkshire Road, Rosemont,Pennsylvania 19010, Estados Unidos de América);

3.) **ROY, Soumitra**
(240 Pugh Road, Wayne, Pennsylvania 19087, Estados Unidos de América).

Solicito expresamente a nombre de la sociedad **THE TRUSTEES OF THE UNIVERSITY OF PENNSYLVANIA,** sociedad constituida de conformidad con las leyes del Estado de Pennsylvania, Estados Unidos de América, domiciliada en 3160 Chestnut Street, Suite 200, Philadelphia, Pennsylvania 19104, Estados Unidos de América, el otorgamiento de Patente para la invención titulada "ÁCIDO NUCLÉICO DE ADENOVIRUS DE SIMIO Y SECUENCIAS DE AMINOÁCIDOS, VECTORES QUE CONTIENEN LOS MISMOS Y MÉTODOS DE USO".

Clase: C12N

JIMENA ESCOBAR URIBE
C.C. 39.779.452 de Usaquen.
T.P. 68.228

5

ÁCIDO NUCLÉIC DE ADENOVIRUS DE SIMIO Y SECUENCIAS DE AMINOÁCIDOS, VECTORES QUE CONTIENEN LOS MISMOS Y MÉTODOS DE USO.

ANTECEDENTES DE LA INVENCIÒN.

El adenovirus es un virus a ADN de doble banda, con un tamaño de genoma de cerca de 36 kilo-bases (kb), el cual ha sido usado ampliamente para aplicaciones de transferencia génica, debido a su habilidad para alcanzar transferencias génicas altamente eficientes en una variedad de tejidos blanco y a su gran capacidad transgénica. Convencionalmente, los genes E1 del adenovirus se borran y reemplazan con una cinta transgénica que consiste de un promotor de selección, la secuencia cADN del gen de interés, y una señal poli A, dando por resultado una replicación defectuosa del virus recombinante.

Los adenovirus tienen una característica morfológica con una cápside icosaédrica que consiste de tres proteínas principales, exón (II), pentón base (III) y una fibra nudosa (knobbed fibre) (IV) a lo largo con un número de otras proteínas menores, VI, VIII, IX, IIIa y IVa2 [W.C.Russel, *J. Gen Virol.*, 81:2573-2604 (Nov 2000)]. El genoma del virus es un ADN de doble banda lineal entrelazada con una proteína terminal unida covalentemente al terminal 5', en el cual se han invertido las repeticiones terminales (ITRs). El ADN del virus está íntimamente asociado con la

proteína VII altamente básica y con un pequeño péptido llamado mu. Otra proteína, V, se empaca con este complejo ADN-proteína y proporciona una estructura unida a la cápside vía proteína VI. El virus también contiene una proteasa virus-codificada, la cual es necesaria para el procesamiento de algunas de las proteínas estructurales, para producir el virus infeccioso maduro.

Los adenovirus recombinantes se han descrito para la entrega de moléculas a las células huésped. Ver, la Patente US 6, 083, 716, la cual describe el genoma de dos los adenovirus de dos chimpancés.

Lo que se necesita en el arte son más vectores efectivos que eviten el efecto de inmunidad pre-existente a los serotipos de adenovirus seleccionados en la población y/o los cuales sean útiles para administración repetida y para elevar el título por segunda vacunación, si se requiere.

Resumen de la Invención.

La presente invención proporciona las secuencias de los ácidos nucleicos aislados y las secuencias de aminoácidos de seis adenovirus de simio, los vectores que contienen estas secuencias y las líneas celulares que expresan los genes de adenovirus de simio. También proporciona un número de métodos para usar los vectores y las células de la invención.

Los métodos de la invención incluyen la entrega de uno o más gene(s) heterólogos seleccionados para un paciente mamífero mediante la administración de un vector de la invención.

A causa de que los constructos de varios vectores, se derivan de los adenovirus de simios más que de los adenovirus humanos, el sistema inmune de pacientes no simios, humanos y veterinarios no se prepara inmediatamente para responder al vector como un antígeno extraño. El uso de las composiciones de esta invención, permite así una expresión más estable del transgen seleccionado cuando se administra a un paciente no-simio. El uso de las composiciones de esta invención para vacunación permite la presentación de un antígeno seleccionado para la producción de respuestas inmuno protectoras. Sin desear estar limitados por la teoría, la capacidad de los adenovirus de la invención para transducir las células dendríticas humanas, es al menos parcialmente responsable de la capacidad de los constructos recombinantes de la invención para inducir una respuesta inmune. Los adenovirus recombinantes de simio de esta invención también se pueden usar para producir productos de genes heterólogos *in Vitro*. Tales productos génicos producidos son en si mismos útiles en una variedad y para una variedad de propósitos tales como se describen aquí.

Estas y otras incorporaciones y ventajas de la invención se describen con más detalle más abajo.

Breve Descripción de los Dibujos.

La Fig. 1 proporciona una alineación de las secuencias de los aminoácidos de L1 y una porción de los bucles (loops) L2 del exón de la proteína cápside del adenovirus C1 de chimpancé [SEQ ID NO: 13], el adenovirus de chimpancé C68 (Pan-9) [SEQ ID NO: 14] y los nuevas secuencias de adenovirus de chimpancé Pan5 [SEQ ID NO: 15], Pan6 [SEQ ID NO: 16], Pan 7 [SEQ ID NO: 17], de la invención. La región conservada que interviene es parte del dominio fundamental conservado entre serotipos deladenovirus.

La Fig. 2 proporciona una alineación de de las secuencias de aminoácidos de los dominios nudo de la fibra (fiber Knob), de los adenovirus de chimpancé C68 (Pan-9) [SEQ ID NO: 18], Pan-6 [SEQ ID NO: 19], Pan-7 [SEQ ID NO: 20], y Pan-5 [SEQ ID NO: 21] y los serotipos 2 de adenovirus humano [SEQ ID NO: 22] y 5 [SEQ ID NO: 23].

DESCRIPCION DETALLADA DE LA INVENCION

La invención proporciona un ácido nucleico nuevo y las secuencias de aminoácidos de Ad Pan5 [SEQ ID NO: 1-4, 15 y 21], Ad Pan6 [SEQ ID NO: 5-8, 16, 19] y serotipo Ad Pan7 [SEQ ID NO: 9-12, 17, 20], los cuales fueron originalmente aislados de los nódulos linfoides de chimpancé. En varios casos a través de la especificación estos adenovirus, alternativamente son llamados aquí

como C5, C6 y C7, respectivamente. También se proporcionan secuencias de adenovirus: SV1 [SEQ ID NO: 24-28], las cuales originalmente fueron aisladas de las células de riñón del mono cynomolgus. La invención también proporciona las secuencias de los adenovirus SV-25 [SEQ ID NO: 29- 33] y SV-39 [SEQ ID NO: 34-37], las cuales originalmente fueron aisladas de las células de riñón del mono rhesus.

La presente invención proporciona nuevos vectores de adenovirus y las líneas de células que los empacan, para producir aquellos vectores para uso en la producción *in vitro* de proteínas recombinantes o fragmentos u otros reactivos. La invención además proporciona composiciones para uso en entregas de una molécula heteróloga para propósitos terapéuticos o vacunas. Tales composiciones para uso terapéutico o vacunas contienen los vectores adenovirales llevando insertada una molécula heteróloga. En adición, las secuencias nuevas de la invención son útiles al proporcionar las funciones esenciales auxiliares requeridas para la producción de vectores virales adenoasociados (AAV). Así, la invención proporciona constructos auxiliares, métodos y líneas celulares las cuales usan estas secuencias en tales métodos de producción.

El término "homología sustancial" o "similaridad sustancial" cuando se refiere a un ácido nucleico a un fragmento del mismo, indica que, cuando se alinean óptimamente con las inserciones o las supresiones del nucleótido apropiado con

otro ácido nucleico (o su banda complementaria), hay identidad en la secuencia de nucleótido en al menos cerca del 95 al 99% de las secuencias alineadas.

El término "homología sustancial" o "similaridad sustancial" cuando se refiere a aminoácidos o fragmentos de los mismos, indican que cuando se alinean óptimamente con las insersiones o supresions aproiadas de aminoácidos con otro aminoácido (o su banda complementaria), hay una identidad en la secuencia de los aminoácidos en por lo menos cerca del 95 al 99% de las secuencias alineadas. Preferiblemente, la homología esta sobre la longitud total de la secuencia, una proteína de la misma, o un fragmento de la misma el cual tiene por lo menos 8 aminoácidos, o más deseablemente, al menos 15 aminoácidos de longitud. Ejemplos de los fragmentos adecuados se describen aquí.

El término "porcentaje de identidad de secuencia" o "idéntico" en el contexto de las secuencias de ácidos nucleicos se refiere a los residuos de las dos secuencias que son el mismo cuando están alineados para la correspondencia máxima. Es deseable que la comparación de la longitud de la identidad de secuencia pueda ser sobre la longitud total del genoma (e.g., cerca 36 kbp), la longitud completa de una estructura abierta de lectura de un gen, proteína, subunidad o enzima [Ver, e.g., las tablas que proporcionan las regiones de codificación adenoviral] o un fragmento de al menos cera de 500 a 5000 nucleótidos. Sin embargo es deseable que la identidad entre pequeños fragmentos sea de, e.g., por lo menos 9 nucleótidos, usualmente al menos cerca de 20 a 24 nucleótidos, al menos de

cerca de 28 a 32 nucleótidos, por lo menos cerca de 36 o más nucleótidos. Similarmente el "porcentaje de identidad de secuencia" se puede determinar fácilmente para secuencias de aminoácidos, sobre la longitud total de una proteína o un fragmento de la misma. Adecuadamente un fragmento es de por lo menos una longitud de 8 aminoácidos, y puede ser hasta de cerca de 700 aminoácidos. Ejemplos de fragmentos adecuados se describen aquí.

La identidad se determina fácilmente usando tales algoritmos y programas de computador como se definen aquí por defecto. Preferiblemente tal identidad es sobre la longitud total de la proteína, enzima, subunidad, o sobre un fragmento de una longitud de por lo menos 8 aminoácidos. Sin embargo, la identidad puede estar basada en regiones cortas, donde se acomode al uso para el cual ha sido puesto el producto del gen idéntico.

Como aquí se describe, los alineamientos se llevan a cabo utilizando una gran variedad de programas de alineamiento de secuencia múltiple disponibles pública o comercialmente tales como "Clustal W", accesible a través de los servidores de web en internet. Alternativamente, también se usan las utilidades Vector NTI. Hay también un número de algoritmos conocidos en el arte que se pueden utilizar para medir la identidad de secuencia de nucleótidos, incluyendo aquellos contenidos en los programas descritos anteriormente. Como otro ejemplo, las secuencias de nucleótidos se pueden comparar utilizando Fasta, un programa en GCG versión 6.1. Fasta proporciona alineamientos y porcentajes de identidad de secuencias de

las regiones del mejor sobrelaparniento entre las secuencias preguntadas y las investigadas. Por ejemplo el porcentaje de identidad de secuencia entre secuencias de ácidos nucleicos se puede determinar usando Fasta con sus parámetros por defecto (un tamaño de palabra de seis y el factor NOPAM para la matriz de anotación), como la proporcionada en GCG versión 6.1, se incorpora aquí por referencia. Programas similares están disponibles para realizar alineamientos de aminoácidos. Generalmente, estos programas se utilizan en condiciones por defecto, sin embargo una persona versada en el arte puede cambiar estas condiciones según las necesite. Alternativamente una persona versada en el arte puede utilizar otro algoritmo o un programa de computador que proporcione por lo menos el nivel de identidad o alineamiento tanto como los proporcionados por los algoritmos y programas referenciados.

Como se usa a lo largo de esta especificación y en las reivindicaciones, el término "comprender" y sus variantes que incluyen "comprende" y "que comprende" entre otras variantes, es inclusivo de otros componentes, elementos, enteros, etapas y similares. El término "consiste de" o "que consiste de" son exclusivos de otros componentes, elementos, enteros, etapas y similares.

I. Secuencias de Adenovirus de Simio.

La invención proporciona secuencias de ácidos nucleicos y secuencias de Pan5, Pan6, Pan7, SV1, SV25 y SV39 los cuales son aislados de otros materiales virales con los que están asociados en la naturaleza.

A. Secuencias de Ácidos Nucleicos.

Las secuencias de ácido nucleico Pan5 de la invención incluyen nucleótidos 1 a 36462 de SEQ ID NO: 1. Las secuencias de ácido nucleico Pan6 de la invención incluyen nucleótidos 1 a 36604 de SEQ ID NO: 5. Las secuencias de ácido nucleico Pan7 de la invención incluyen nucleótidos 1 a 36535 de SEQ ID NO: 9. Las secuencias de ácido nucleico SV1 de la invención incluyen nucleótidos 1 a 34264 de SEQ ID NO: 24. Las secuencias de ácido nucleico SV25 de la invención incluyen nucleótidos 1 a 31044 de SEQ ID NO: 29. Las secuencias de ácido nucleico SV39 de la invención incluyen nucleótidos 1 a 34115 de SEQ ID NO: 34. Ver, lista de secuencia, la cual se incorpora aquí por referencia.

Las secuencias de ácido nucleico de la invención además encierran la banda que es complementaria a las secuencias de SEQ ID NO: 5, 9, 24, 29, y 34, como también las secuencias de ARN y cADN correspondientes a las secuencias de esos guarismos de secuencias y sus bandas complementarias. Además se incluyen en esta invención secuencias de ácido nucleico que son mayores de 95 a

98%, y más preferiblemente homólogos o idénticos en cerca del 99 al 99.9 % a la secuencia que se lista. También están incluidas en las secuencias de ácido nucleico de la invención las variantes naturales y las modificadas por ingeniería de las secuencias proporcionadas en SEQ ID NO: 5, 9, 24, 29, y 34 y sus bandas complementarias. Tales modificaciones incluyen, por ejemplo etiquetas que son conocidas en el arte, metilación, y sustitución de uno o más de los nucleótidos de ocurrencia natural con un nucleótido degenerado.

La invención además encierra fragmentos de las secuencias de Pan5, Pan6, Pan7, SV1, SV25 y SV39, sus bandas complementarias, cADN y ARN de los mismos. Los fragmentos adecuados son de por lo menos 15 nucleótidos de longitud y encierran fragmentos funcionales, i.e., fragmentos que son de interés biológico. Por ejemplo, un fragmento funcional puede expresar un producto adenoviral deseado o puede ser útil en la producción de vectores virales recombinantes. Tales fragmentos incluyen las secuencias de genes y fragmentos listados en las tablas más adelante.

Las siguientes tablas proporcionan las regiones de transcripción y las estructuras de lectura abierta en las secuencias de adenovirus de simio de la invención. Para ciertos genes la transcripción y las estructuras de lectura abierta (ORFs) están localizadas en la banda complementaria a la presentada en SEQ ID NO: 5, 9, 24, 29, y 34. Ver, e.g., E2b, E4 y E2a. Se muestran también los pesos moleculares *(M.W.)* calculados de las proteínas codificadas. Note que la estructura de lectura

abierta E1a Pan5 [nt 576-1436 de SEQ ID NO: 1], Pan6 [nt 576 a 1437 de SEQ ID NO: 5] y Pan7 [nt 576 a 1437 de SEQ ID NO: 9] contienen sitios de unión interno. Estos sitios de unión se observan en las siguientes tablas.

Ad Pan-5 [SEQ ID NO:1]				
Regiones		Inicio (nt)	Final (nt)	*M. W. (Daltons)*
ITR		1	120	-
E1a	Transcrito	478		-
	13S	576-664,1233-1436		*28120*
	12S	576-1046,1233-1436		*24389*
	9S	576-644,1233-1436		*9962*
	Transcrito		1516	-
E1b	Transcrito	1552		-
	Pequeña T	1599	2171	*22317*
	Grande T	1904	3412	*55595*
	LX	3492	3920	*14427*
	Transcrito		3959	-
E2b	Transcrito	10349		-
	PTP	10349	8451	*72930*
	Polimerasa	8448	5083	*127237*
	IVa2	5604	3980	*50466*
	Transcrito		3960	
28.1 kD		515	5979	*28141*
Agnoproteína		7864	8580	*25755*

Ad Pan-5 (continuación) [SEQ ID NO:1]				
Regiones		Inicio (nt)	Final (nt)	*M. W. (Daltons)*
L1	Transcrito	10849		-
	52/55D	10851	12025	-
	IIIa	12050	13819	*65669*
	Transcrito		13832	-
	Transcrito	13894		-
L2	Penton	13898	15490	*59292*
	VII	15494	16078	*21478*
	V	16123	17166	*39568*
	Mu	17189	17422	*8524*
	Transcrito		17442	-
	Transcrito	17488		-
L3	VI	17491	18222	*26192*
	Exón	18315	21116	*104874*
	Endoproteasa	20989	21783	*28304*
	Transcrito		21811	-

Ad Pan-5 (continuación) [SEQ ID NO:1]				
Regiones		Inicio (nt)	Final (nt)	*M. W. (Daltons)*
E2a	Transcrito	10849		-
	DBP	10851	12025	-
	Transcrito	12050	13819	*65669*
L4	Transcrito	13898	15490	*59292*
	100kD	15494	16078	*21478*
	Homólogo 33kD	16123	17166	*39568*
	VIII	17189	17422	*8524*
	Transcrito		17442	-
E3	Transcrito	26788	-	-
	Orf # 1	27112	27432	*12098*
	Orf # 2	27386	28012	*23040*
	Orf # 3	27994	28527	*19525*
	Orf # 4	28577	29156	*22567*
	Orf # 5	29169	29783	*22267*
	Orf # 6	29798	30673	*31458*
	Orf # 7	30681	30956	*10477*
	Orf # 8	30962	31396	*16523*
	Orf # 9	31389	31796	*15236*
	Transcrito		31837	-

Ad Pan-5 (continuación) [SEQ ID NO:1]				
Regiones		Inicio (nt)	Final (nt)	*M. W. (Daltons)*
L5	Transcrito	32032		-
	Fibra	32035	33372	*47670*
	Transcrito		33443	-
E4	Transcrito	36135		-
	Orf7	33710	33462	*9191*
	Orf6	34615	33710	*35005*
	Orf4	34886	34521	*13878*
	Orf3	35249	34896	*13641*
	Orf2	35635	35246	*14584*
	Orf1	36050	35676	*13772*
	Transcrito		33437	-
ITR		36343	36462	-

Ad Pan-6 [SEQ ID NO:5]				
Regiones		Inicio (nt)	Final (nt)	*M. W. (Daltons)*
ITR		1	123	-
E1a	Transcrito	478		-
	13S	576-1143, 1229-1437		*28291*
	12S	576-1050,1229-1437		*24634*
	9S	576-645,1229-1437		*10102*
	Transcrito		1516	-
E1b	Transcrito	1553		-
	T pequeña	1600	2172	*22315*
	T grande	1905	3413	*55594*
	IX	3498	3926	*14427*
	Transcrito		3965	-
E2b	Transcrito	10341		-
	PTP	10340	8451	*72570*
	Polimerasa	8445	5089	*126907*
	IVa2	5610	3986	*50452*
	Transcrito		3966	-
L1	Transcrito	10838		-
	52/55 kD	10840	12012	*44205*
	IIIa	12036	137979	*65460*
	Transcrito		13812	-

Ad Pan-6 (continuación) [SEQ ID NO:5]				
Regiones		Inicio (nt)	Final (nt)	*M. W. (Daltons)*
28.1 kd		5161	5985	*28012*
Agnoproteína		7870	8580	*25382*
L2	Transcrito	13874		-
	Penton	13878	15467	*59314*
	VII	15471	16055	*21508*
	V	16100	17137	*39888*
	Mu	17160	17393	*8506*
	Transcrito		17415	-
L3	Transcrito	17466		-
	VI	17469	18188	*25860*
	Exón	18284	21112	*106132*
	Endoproteasa	21134	21754	*23445*
	Transcrito		21803	-
E2a	Transcrito	26780	-	-
	DBP	23375	21837	*57299*
	Transcrito		21780	-

Ad Pan-6 (continuación) [SEQ ID NO:5]				
Regiones		Inicio (nt)	Final (nt)	*M. W. (Daltons)*
L4	Transcrito	23398		-
	100kD	23404	25806	*88577*
	33 kD homólogo	25523	26357	*24609*
	VIII	26426	27109	*24749*
	Transcrito		27419	-
E3	Transcrito	26786		-
	Orf1	27110	27430	*12098*
	Orf2	27384	28007	*22880*
	Orf3	27989	28519	*19460*
	Orf4	28553	29238	*25403*
	Orf5	29249	29860	*22350*
	Orf6	29875	30741	*31028*
	Orf7	30749	31024	*10469*
	Orf8	31030	31464	*16540*
	Orf9	31457	31864	*15264*
	Transcrito		31907	-
L5	Transcrito	32159		-
	Fibra	32162	33493	*47364*
	Transcrito		33574	-

Ad Pan-6 (continuación) [SEQ ID NO:5]				
Regiones		Inicio (nt)	Final (nt)	*M. W. (Daltons)*
E4	Transcrito	36276		-
	Orf7	33841	33593	*9177*
	Orf6	34746	33841	*35094*
	Orf4	35017	34652	*13937*
	Orf3	35380	35027	*13627*
	Orf2	35766	35377	*14727*
	Orf1	36181	35807	*13739*
	Transcrito		33558	-
ITR		36482	36604	-

Ad Pan-7 [SEQ ID NO:9]				
Regiones		Inicio (nt)	Final (nt)	*M. W. (Daltons)*
ITR		1	132	-
E1a	Transcrito	478		-
	13S	576-1143, 1229-1437		*28218*
	12S	576-1050,1229-1437		*24561*
	9S	576-645,1229-1437		*10102*
	Transcrito		1516	-
E1b	Transcrito	1553		-
	T pequeña	1600	2178	*22559*
	T grande	1905	3419	*55698*
	IVa2	3492	5616	*50210*
	Transcrito		3971	-
E2b	Transcrito	10341		-
	PTP	10340	8457	*72297*
	Polimerasa	8451	5095	*126994*
	IX	3504	3932	*14441*
	Transcrito		3972	-
28.1 kd		5167	5991	*28028*
Agnoproteína		7876	8586	*25424*
L1	Transcrito	10834		-
	52/55 kD	10836	12011	*44302*
	IIIa	12035	13795	*65339*
	Transcrito		13808	-

Ad Pan-7 (continuación) [SEQ ID NO:9]				
Regiones		Inicio (nt)	Final (nt)	*M. W. (Daltons)*
L2	Transcrito	13870		-
	Penton	13874	15469	*59494*
	VII	15473	16057	*21339*
	V	16102	17139	*39414*
	Mu	17167	17400	*8506*
	Transcrito		17420	-
L3	Transcrito	17467		-
	VI	17470	18198	*26105*
	Exón	18288	21086	*104763*
	Endoproteasa	21106	21732	*23620*
	Transcrito		21781	-
E2a	Transcrito	26764		-
	DBP	23353	21815	*57199*
	Transcrito		21755	-
L4	Transcrito	23370		-
	100 kD	23376	25781	*88520*
	33 kD homólogo	25489	26338	*25155*
	VIII	26410	27093	*24749*
	Transcrito		27403	-

Ad Pan-7 (continuación) [SEQ ID NO:9]				
Regiones		Inicio (nt)	Final (nt)	*M. W. (Daltons)*
E3	Transcrito	26770		-
	Orf1	27094	27414	*12056*
	Orf2	27368	27988	*22667*
	Orf3	27970	28500	*19462*
	Orf4	28530	29150	*22999*
	Orf5	29163	29777	*22224*
	Orf6	29792	30679	*32153*
	Orf7	30687	30962	*10511*
	Orf8	30968	31399	*16388*
	Orf9	31392	31799	*15205*
	Transcrito	-	31842	-
L5	Transcrito	32091		-
	Fibra	32094	33425	*47344*
	Transcrito		33517	-

Ad Pan-7 (continuación) [SEQ ID NO:9]				
Regiones		Inicio (nt)	Final (nt)	*M. W. (Daltons)*
E4	Transcrito	36208		-
	Orf7	33784	33536	*9191*
	Orf6	34689	33784	*35063*
	Orf4	34960	34595	*13879*
	Orf3	35323	34970	*13641*
	Orf2	35709	35320	*14644*
	Orf1	36123	35749	*13746*
	Transcrito		33501	-
ITR		36404	36535	-

	Ad SV-1 [SEQ ID NO:24]		Ad SV-25 [SEQ ID NO:29]		Ad SV-39 [SEQ ID NO:34]	
Región	Inicio	Final	Inicio	Final	Inicio	Final
ITR	1	106	1	133	1	150
E1a	352	1120	-	-	404	1409
E1b	1301	2891	359	2273	1518	3877
E2b	9257	2882	9087	2754	10143	3868
E2a	24415	20281	24034	20086	25381	21228
E3	24974	27886	24791	25792	25790	29335
E4	33498	30881	30696	28163	33896	31157
ITR	34145	34264	30912	31044	33966	34115

	Ad SV-1 [SEQ ID NO:24]		Ad SV-25 [SEQ ID NO:29]		Ad SV-39 [SEQ ID NO:34]	
Región	Inicio	Final	Inicio	Final	Inicio	Final
ITR	1	106	1	133	1	150
L1	9513	12376	9343	12206	10416	13383
L2	13453	15858	12283	15696	13444	16877
L3	15910	20270	15748	20080	17783	21192
L4	21715	25603	21526	25420	22659	26427
L5	28059	30899	25320	28172	29513	31170
ITR	34145	34264	30912	31044	33966	34115

	Proteína	Ad SV-1, SEQ ID NO: 24		
		Inicio	Fin	*M. W.*
ITR		1	106	-
E1a	13S	459	953	*18039*
	12S			
E1b	T pequeña			
	T grande	1301	2413	*42293*
	IX	2391	2885	*16882*
E2b	IVa2	4354	2924	54087
	Polimerasa	6750	4027	*102883*
	PTP	9257	7371	*72413*
	Agnoproteína	6850	7455	*20984*
L1	52/55 kD	9515	10642	*42675*
	IIIa	10663	12372	*636568*

	Proteína	Ad SV-1, SEQ ID NO: 24 Cont.		
		Inicio	Fin	*M. W.*
L2	Pentón	12454	13965	*56725*
	VII	13968	14531	*20397*
	V	14588	15825	*39374*
	Mu	15645	15857	*7568*
L3	VI	15911	16753	30418
	Exón	16841	19636	*104494*
	Endoproteasa	19645	20262	*23407*
2a	DBP	21700	20312	52107
L4	100kD	21721	24009	*85508*
	VIII	24591	25292	*25390*
E3	Orf#1	25292	25609	*11950*
	Orf#2	25563	26081	*18940*
	Orf#3	26084	26893	*30452*
	Orf#4	26908	27180	*10232*
	Orf#5	27177	17512	*12640*
	Orf#6	27505	27873	*13639*
L5	Fibra #2	28059	29150	*39472*
	Fibra #1	29183	30867	*61128*
E4	Orf7	31098	30892	*7837*
	Orf6	31982	31122	*33921*
	Orf4	32277	31915	*14338*
	Orf3	32629	32279	*13386*
	Orf2	33018	32626	*14753*
	Orf1	33423	33043	*14301*
ITR		34145	34264	

	Proteína	Ad SV-25 [SEQ ID NO:29]			Ad SV-39 [SEQ ID NO:34]		
		Inicio	Fin	*M. W.*	Inicio	Fin	*M. W.*
ITR		1	133	-	1	150	-
E1a	13S				492	1355	*28585*
	12S				492	1355	*25003*
E1b	T pequeño	478	1030	*20274*	1518	2075	*21652*
	T grande	829	2244	*52310*	1823	3349	*55534*
	IX	2306	2716	*13854*	3434	3844	*14075*
E2b	IVa2	4208	2755	*54675*	3912	5141	*46164*
	Polimerasa	6581	3858	*102839*	7753	5033	*103988*
	PTP	9087	7207	*71326*	10143	8335	*69274*
	Agnoproteína	6681	7139	*16025*	-	-	-
L1	52/55 kD	9345	10472	*42703*	10418	11608	*44232*
	IIIa	10493	12202	*63598*	11574	13384	*66078*
L2	Pentón	12284	13801	*56949*	13448	14959	*56292*
	VII	13806	14369	*20369*	14960	15517	*20374*
	V	14426	15463	*39289*	15567	16628	*39676*
	Mu	15483	15695	*7598*	16650	16871	*7497*
L3	VI	15749	16591	*30347*	16925	17695	*28043*
	Exón	16681	19446	*104035*	17785	20538	*102579*
	Endoproteasa	19455	20072	*23338*	20573	21181	*22716*
2a	DBP	21511	20123	*52189*	22631	21231	*53160*
L4	100kD	21532	23829	*85970*	22659	25355	*100362*
	VIII	24408	25109	*25347*	25410	26108	*25229*

	Proteína	Ad SV-25 [SEQ ID NO:29] Continuación			Ad SV-39 [SEQ ID NO:34] Continuación		
		Inicio	Fin	*M. W.*	Inicio	Fin	*M. W.*
E3	Orf#1	25109	25426	*11890*	26375	27484	*42257*
	Orf#2				27580	28357	*29785*
	Orf#3				28370	28645	*10514*
	Orf#4				28863	29333	*18835*
	Orf#5						
	Orf#6						
L5	Fibra #2	25380	26423	*37529*			
	Fibra #1	26457	28136	*60707*	29515	31116	*56382*
E4	Orf7				31441	31118	*11856*
	Orf6	29255	28395	*33905*	32292	31438	*33437*
	Orf4	29550	29188	*14399*	32587	32222	*13997*
	Orf3	29902	29552	*13284*	32954	32607	*13353*
	Orf2	30291	29899	*14853*	33348	32959	*14281*
	Orf1	30316	30696	*14301*	33764	33378	*14235*
ITR		30912	31044		33966	34115	

Las secuencias de ácido nucleico adenoviral Pan5, Pan6, Pan7, SV1, SV25, y SV39 son útiles como agentes terapéuticos y en la construcción de una variedad de sistemas de vector y de células huésped. Como se usa aquí un vector incluye cualquier molécula adecuada de ácido nucleico que incluye ADN desnudo, un

plásmido, un virus un cósmido o un episoma. Estas secuencias y los productos se pueden usar solas o en combinación con otras secuencias adenovirales o fragmentos, o en combinación con elementos de otras secuencias adenovirales o no adenovirales. Las secuencias adenovirales de la invención, también son útiles como vectores de liberación en contrasentido, vectores para terapia génica o vectores de vacuna. Así la invención proporciona, además de moléculas de ácido nucleico, vectores de entrega de genes y células huésped las cuales contienen las secuencias Ad de la invención.

Por ejemplo, la invención abarca una molécula de ácido nucleico que contiene secuencias del AdITR de simio de la invención. En otro ejemplo, la invención proporciona una molécula de ácido nucleico que contiene secuencias Ad de simio de la invención que codifican un producto genético Ad deseado. Aun otra molécula de ácido nucleico construida usando las secuencias de la invención será clara para una persona versada en el arte, con base en la información aquí proporcionada.

En una incorporación, las regiones del gen Ad de simio identificadas aquí pueden utilizarse en una variedad de vectores para la entrega de una molécula heteróloga a una célula. Por ejemplo, los vectores son generados por expresión de una proteína cápside adenoviral (o fragmento de la misma) con el propósito de generar un vector viral en una célula huésped empacada. Tales vectores se pueden diseñar para expresión en trans. Alternativamente tales vectores se diseñan para

proporcionar células las cuales establemente contienen secuencias que expresan funciones adenovirales deseadas, e.g., una o más de E1a, E1b, las secuencias terminales de repetición, E2a,. E2b, E4, y la región E4ORF6.

En adición, las secuencias del gen adenoviral y fragmentos de la misma se utilizan para proporcionar las funciones auxiliares necesarias para la producción de virus dependientes del auxiliar (e.g., vectores adenovirales suprimidos de las funciones esenciales o virus adenoasociados (AAV)). Para tales métodos de producción, las secuencias adenovirales de simio de la invención son utilizadas en dichos métodos de una manera similar a los descritos para el Ad humanos sin embargo, debido a las diferencias en las secuencias entre las secuencias adenovirales de simio de la invención y de aquellas de Ad humano, el uso de las secuencias de la invención esencialmente elimina la posibilidad de recombinación homóloga con funciones auxiliares en una célula huésped llevando funciones Ad E1 humanas, e.g., células 293 las cuales pueden producir contaminantes adenovirales infecciosos durante la producción de rAAV.

Los métodos de producción de rAAV que usan funciones auxiliares adenovirales se han descrito ampliamente en la literatura con serotipos adenovirales humanos. Ver, e.g., Patente US 6, 258, 595 y las referencias aquí citadas. Ver, también, las Patentes US 5, 871, 982; WO 99/14354; WO 99/15685; WO 99/47691. Estos métodos también se pueden usar en la producción del serotipo no humano AAV que incluyen serotipos AAV de primates no humanos. Las secuencias del gen

adenoviral de simio de la invención la cuales proporcionan las funciones auxiliares necesarias (e.g., E1a, E1b, E2a y / o E4 ORF6) pueden ser particularmente útiles en proporcionar la función adenoviral necesaria mientras minimiza o elimina la posibilidad de recombinación con cualquiera de otros adenovirus presentes en la célula empacada rAAV, las cuales son, típicamente, de origen humano. Entonces los genes seleccionados o las estructuras de lectura abierta de las secuencias adenovirales de la invención se pueden utilizar en estos métodos de producción de rAAV.

Alternativamente, los vectores recombinantes adenovirales de simio de la invención pueden ser utilizados en estos métodos. Tales vectores adenovirales de simio pueden incluir, e.g., un híbrido Ad/AAV de chimpancé en el la cual las secuencias Ad de chimpancé flanquean una cinta de expresión compuesta de, e.g., AAV3' y/o 5' ITRs y un transgen bajo el control de las secuencias reguladoras las cuales controlan su expresión. Una persona versada en el arte reconocerá que aún otros vectores adenovirales de simio y /o secuencias de genes de la invención serán útiles para la producción de rAAV y otros virus dependientes del auxiliar adenoviral.

Aún en otra incorporación, las moléculas de ácido nucleico se diseñan para entrega y expresión de productos del gene adenoviral seleccionado en una célula huésped para lograr el efecto fisiológico deseado. Por ejemplo, una molécula de ácido nucleico que contiene secuencias que codifican una proteína del adenovirus

E1a de la invención puede ser entegada a un sujeto para utilizarla como terapéutico del cáncer. Opcionalmente, tal molécula se formula en un vehículo basado en un lípido y preferencialmente actúa en las células cancerosas blanco. Tal formulación se puede combinar con otros terapéuticos del cáncer (e.g., cisplatino, taxol, o similares). Aún otros usos para las secuencias adenovirales proporcionadas aquí serán claros para una persona versada en el arte.

Además, una persona versada en el arte fácilmente entenderá que las secuencias Ad de la invención se pueden adaptar fácilmente para usarlas en una variedad de sistemas de vectores virales y no virales para la entrega *in vitro*, *ex vivo* o *in vivo* de moléculas terapéuticas e inmunogénicas. Por ejemplo, el Pan5, Pan6, Pan7, SV1, SV25, y/o genomas Ad de simio SV39 de la invención se pueden utilizar en una variedad de sistemas de vectores rAd y no-rAd. Tales sistemas de vectores pueden incluir, e.g., plásmidos, lentivirus, retrovirus, poxvirus, virus de vacuna y sistemas virales adenoasociados, entre otros. La selección de estos sistemas de vectores no es una limitación de la presente invención.

La invención además proporciona moléculas utilizadas para la producción de proteínas de simio y simio derivadas de la invención. Tales moléculas pueden transportar polinucleótidos incluyendo secuencias Ad ADN de simio de la invención en la forma de ADN desnudo, un plásmido, un virus o cualquier otro elemento genético.

B. Proteínas adenovirales de simio de la invención

La invención además proporciona productos génicos de los adenovirus anteriores tales como proteínas, enzimas, y fragmentos de las mismas los cuales están codificados por los ácidos nucleicos adenovirales de la invención. La invención además abarca proteínas de Pan5, Pan6, y Pan7, SV1, SV25, y SV39, enzimas, y fragmentos de las mismas, que tienen las secuencias de aminoácidos codificadas por éstas secuencias de ácidos nucleicos las cuales se generan por otros métodos. Tales proteínas incluyen aquellas codificadas por las estructuras de lectura abierta identificadas en las tablas anteriores, en las Figs. 1 y 2 y fragmentos de las mismas.

Así en un aspecto, la invención proporciona proteínas adenovirales de simio únicas las cuales son sustancialmente puras, i.e., están libres de otras proteínas virales y proteináceas. Preferiblemente, estas proteínas son al menos en un 10% homogéneas, más preferiblemente homogéneas en un 60% y mucho más preferiblemente homogéneas, en un 95%.

En otra incorporación, la invención proporciona proteínas cápside simio derivadas únicas. Como se usa aquí una proteína cápside simio derivada incluye cualquier proteína cápside adenoviral que contiene un proteína cápside de Pan5, Pan6 Pan7, SV1, SV25 o SV39 o un fragmento de las mismas, como se definió anteriormente, que incluye sin limitación proteínas cápsides quiméricas, proteínas

de fusión, proteínas cápsides artificiales, proteínas cápsides sintéticas, y proteínas cápsides recombinantemente, sin limitación de los medios de generación de estas proteínas.

Adecuadamente, estas proteínas cápside simio derivadas contienen uno o más regiones Pan5, Pan6, y Pan7, SV1, SV25, y SV39, o fragmentos de las mismas (e.g., un exón, pentón, fibra o fragmentos de los mismos) en combinación con regiones cápside o fragmentos de las mismas de diferentes serotipos adenovirales o proteínas cápside de simio modificadas o fragmentos como se describen aquí. Como se usa aquí, una "modificación de una proteína cápside asociada con tropismo alterado" incluye una proteína cápside alterada, i.e., un pentón, exón o región de la proteína fibra, o un fragmento de los mismos, tal como el dominio nudoso de la región fibra, o de un polinucleótido que codifica la misma, tal que esa especificidad se altera. La cápside simio derivada se puede construir con uno o más de los Ad de simio de la invención u otros serotipos Ad los cuales pueden ser de origen humano o no humano. Dicho Ad se puede obtener de una variedad de fuentes que incluyen ATCC, fuentes comerciales y académicas o las secuencias de los Ad se pueden obtener del GenBank o de otras fuentes adecuadas.

Las secuencias de aminoácidos de las proteínas pentón de los adenovirus de simio de la invención se proporcionan aquí. La proteína pentón del AdPan5 se proporciona en SEQ ID NO: 2. La pentón AdPan7 se proporciona en SEQ ID NO: 6. La pentón AdPan6 se proporciona en SEQ ID NO: 10. La pentón SV1 es

proporcionada en SEQ ID NO: 25. La proteína pentón SV25 se proporciona en SEQ ID NO: 30. La pentón SV39 es proporcionada en SEQ ID NO: 35. Adecuadamente cualquiera de estas proteínas pentón, o fragmentos únicos de las mismas, se pueden utilizar para una variedad de propósitos. Ejemplos de fragmentos adecuados incluyen la pentón que tiene trocamientos N/terminal y/o C/terminal de cerca de 50, 100, 150, 0 200 aminoácidos, basados en la numeración de aminoácidos proporcionada anteriormente y en SEQ ID NO:2; SEQ ID NO:6, SEQ ID NO:25, SEQ ID NO:30, o SEQ ID NO:35. Otros fragmentos adecuados incluyen fragmentos internos más cortos, C-terminal ó N-terminal. Además, la proteína pentón se puede modificar para una variedad de propósitos conocidos por aquellos ilustrados en el arte.

La invención además proporciona las secuencias de aminoácidos de la proteína exón de Pan5 [SEQ ID NO: 3], Pan6 [SEQ ID NO: 7], Pan7 [SEQ ID NO: 11], SV1 [SEQ ID NO: 26], SV25 [SEQ ID NO: 31] y/o SV39 [SEQ ID NO: 36]. Adecuadamente estas proteínas exón, o fragmentos únicos de las mismas, se pueden utilizar para una variedad de propósitos. Ejemplos de fragmentos adecuados incluyen la exón que tiene trocamientos N/terminal y/o C/terminal de cerca de 50, 100, 150, 200, 300, 400, o 500 aminoácidos, basados en la numeración de aminoácidos proporcionada anteriormente y en SEQ ID NO: 3, 7, 11, 26, 31 y 36. Otros fragmentos adecuados incluyen fragmentos internos más cortos, C-terminal ó N-terminal. Por ejemplo, un fragmento adecuado, de la región (dominio) bucle de la proteína exón, designado como DE1 Y FG1, o una región

hipervariable del mismo. Tales fragmentos incluyen las regiones que abarcan cerca de 125 a 443 residuos de aminoácidos; cerca de 138 a 441, o fragmentos más pequeños, tal como aquellos que abarcan cerca de 138 residuos a 163 residuos; cerca de 170 a cerca de 176; cerca de 195 a cerca de 203; cerca de 233 a cerca de 246; cerca de 253 a cerca de 264; cerca de 287 a cerca de 297; y cerca de 404 a cerca de a 430 de las proteínas exón de simio, referidas a SEQ ID NO: 3, 7, 11, 26, 31, o 36. Otros fragmentos adecuados pueden ser fácilmente identificados por una persona versada en el arte. Además la proteína exón se puede modificar para una variedad de propósitos, conocidos por aquellos versados en el arte. Por que la proteína exón es la determinate para el serotipo de un adenovirus, tales proteínas exón artificiales resultarían en adenovirus que tienen serotipos artificiales. También se pueden construir otras proteínas cápside artificiales utilizando secuencias pentón Ad de chimpancé y/o secuencias fibra de la invención y/o fragmentos de las mismas.

En un ejemplo, podría ser deseable generar un adenovirus que tenga una proteína exón alterada utilizando las secuencias de una proteína exón de la invención. Un método adecuado para alterar las proteínas exón, se describe en la patente US 5, 922, 315, la cual se incorpora aquí por referencia. En este método, al menos una región bucle de la exón de adenovirus se cambia con al menos una región bucle de otro serotipo de adenovirus. Así, al menos una región bucle de la proteína exón de adenovirus alterada es una región bucle de exón Ad de simio de la invención (e.g., Pan 7). En una incorporación, una región bucle de la proteína exón

Pan 7 se reemplaza por una región bucle de otro serotipo de adenovirus. En otra incorporación, la región bucle de la exón Pan7 se usa para reemplazar una región bucle de otro serotipo de adenovirus. Serotipos de adenovirus adecuados se pueden seleccionar fácilmente de entre serotipos humanos y no humanos, como aquí se describe. La Pan7 se seleccionó para propósitos de ilustración solamente; las otras proteínas exón Ad de simio de la invención pueden ser alteradas similarmente, o usadas para alterar otras exón Ad. La selección de un serotipo adecuado no es una limitación de la presente invención. Aún otros usos para las secuencias de las proteínas exón de la invención serán claros para aquellos versados en el arte.

La invención además abarca las proteínas fibra de los adenovirus de simio de la invención. La proteína fibra de AdPan 5 tiene la secuencia de aminoácidos de SEQ ID NO: 4. La proteína fibra de AdPan6 tiene la secuencia de aminoácidos de SEQ ID NO: 8. La proteína fibra de AdPan7 tiene la secuencia de aminoácidos de SEQ ID NO: 12. La SV-1 tiene dos proteínas fibra; la fibra 2 tiene la secuencia de aminoácidos de SEQ ID NO: 27 y la fibra 1 tiene la secuencia de aminoácidos de SEQ ID NO: 28. El SV-25 también tiene dos proteínas fibras; la fibra 2 tiene la secuencia de aminoácidos de SEQ ID NO: 32 y la fibra 1 tiene la secuencia de aminoácidos de SEQ ID NO: 33. La proteína fibra de SV-39 tiene la secuencia de aminoácidos de SEQ ID NO: 37.

40
41

Adecuadamente las proteínas fibra, o fragmentos únicos de las mismas, se pueden utilizar para una variedad de propósitos. Un fragmento adecuado es el fragmento nudoso, el cual abarca cerca de 247 a 425 aminoácidos de SEQ ID NO: 4, 8, 12, 28, 32, 33 y 37. Ver Fig. 2. Ejemplos de otros fragmentos adecuados incluyen la fibra que tiene trocamientos N-terminal y/o C-terminal de cerca de 50, 100, 150, o 200 aminoácidos, basados en la numeración de aminoácidos proporcionada antes y en SEQ ID NO: 4, 8, 12, 28, 32, 33 y 37. Aún otros fragmentos adecuados incluyen fragmentos internos. Además, la proteína fibra se puede modificar utilizando una variedad de técnicas conocidas por aquellos versados en el arte.

La invención además abarca fragmentos únicos de las proteínas de la invención los cuales son al menos de 8 aminoácidos de longitud. Sin embargo, fácilmente se pueden utilizar fragmentos de otras longitudes deseadas. En adición, la invención abarca tantas modificaciones como se puedan introducir para mejorar el producto y/o la expresión producto de los genes Pan5, Pan6, Pan7, SV1, SV25 o SV39, e.g., la construcción de una molécula de fusión en la cual todo o un fragmento del producto de los genes Pan5, Pan6, Pan7, SV1, SV25 o SV39 se funde (bien directamente o vía un ligante) con una pareja de fusión para mejorarla. Otras modificaciones adecuadas incluyen, sin limitación, truncación de una región que codifica (e.g., una proteína o enzima) para eliminar una pre o pro-proteína ordinariamente degradada y proporcionar la proteína o enzima madura y/o la mutación de una región que codifica para proporcionar un producto segregable por

el gen. Aún otras modificaciones serán claras para alguien versado en el arte. La invención además abarca proteínas que tienen al menos cerca del 95% a cerca del 99% de identidad con las proteínas Pan5, Pan6, Pan7, SV1, SV25 o SV39, aquí proporcionadas.

Como aquí se describe, los vectores de la invención que contienen proteínas cápside adenovirales de la invención son particularmente bien adecuados para uso en aplicaciones en las cuales los anticuerpos de neutralización disminuyen la efectividad de otros vectores basados en serotipos Ad así como también otros vectores virales. Los vectores rAd de la invención son particularmente ventajosos en readministración para repetir la terapia génica o para promover la respuesta inmune (títulos de vacunas).

Bajo ciertas circunstancias, puede ser deseable usar uno o más de los productos génicos Pan5, Pan6, Pan7, SV1, SV25 y/o SV39 (e.g., una proteína cápside o un fragmento de la misma) para generar un anticuerpo. El término "un anticuerpo", como se usa aquí, se refiere a una molécula de inmunoglobulina la cual es capaz de enlazarse específicamente a un epitope. Los anticuerpos de la invención unidos, preferiblemente específicamente y sin reactividad cruzada a un epitope Pan5, Pan6, Pan7, SV1, SV25 y/o SV39. Los anticuerpos en la presente invención existen en una variedad de formas que incluyen, por ejemplo, anticuerpos policlonales de alta afinidad, anticuerpos monoclonales, anticuerpos sintéticos, anticuerpos quiméricos, anticuerpos recombinantes, anticuerpos humanizados.

Tales anticuerpos se originan de las clases de inmunoglobulinas IgG, IgM, IgA, IgD e IgE.

Tales anticuerpos se pueden generan usando cualquiera de un número de métodos conocidos en el arte. Anticuerpos adecuados se pueden generar mediante técnicas convencionales bien conocidas, e.g., Kohler y Milstein y las muchas modificaciones conocidas de la misma. Similarmente títulos altos deseables de anticuerpos se generan por aplicación de técnicas recombinantes conocidas a los anticuerpos monoclonales o policlonales desarrollados para estos antígenos [ver, e.g., Patente PCT Aplicación No. PCT/GB85/00392; Patente Británica Aplicación Publicación No GB2188638A; Amit *et al.*, 1986 *Science*, 233: 747-753; Queen *et al.*, 1989 *Proc, Nat'l. Acad. Sci. USA*, 86:10029–10033; Patente PCT Aplicación No. PCT/WO9007861; y Riechmann *et al.*, *Nature*, 332: 323–327 (1988); Huse *et al.*, 1988a *Science*, 246:1275-1281]. Alternativamente, los anticuerpos se pueden producir manipulando las regiones que determinan la complementariedad de anticuerpos animales o humanos para el antígeno de esta invención. Ver, e.g., E. Mark y Padlin, "Humanization of Monoclonal Antibodies", capítulo 4, The Handbook of Experimental Pharmacology, Vol. 113, The Pharmacology of Monoclonal antibodies, Springer-Verlag (June, 1994); Harlow *et al.*, 1999, Using Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press, NY; Harlow *et al.*, 1989, Anticuerpos: A Laboratory Manual, Cold Spring Harbor Laboratory Press, NY; Houston *et al.*, 1988, *Proc. Natl.Acad Sci USA:* 85: 5879-5883; I Bird *et al.*, 1988, *Science 242:423–426.* Además la presente

invención proporciona anticuerpos anti-idiotipo (Ab2) y anticuerpos anti-anti-idiotipo (Ab3). Ver.e.g. M. Wettendorff et al., "Modulation of anti-tumor immunity by anti-idiotypic antibodies." En Idiotipic Network and Diseasses, ed, by J. Cerny y J. Hiernaux, 1990 *J. Am. Soc. Microbiol.*, Washington DC: pp 203-229]. Estos anti-idiotipos y anticuerpos anti-anti-idiotipo se producen utilizando técnicas bien conocidas por aquellos versados en el arte. Estos anticuerpos se pueden utilizar para una variedad de propósitos, que incluyen métodos de diagnóstico, clínicos y Kits.

Bajo ciertas circunstancias, puede ser deseable introducir una marca detectable o un marbete sobre un producto génico de Pan5, Pan6, Pan7, SV1, SV25 y/o SV39, anticuerpo u otro constructo de la invención. Como aquí se usa, una marca detectable, es una molécula la cual es capaz, sola o sobre interacción con otra molécula, de proporcionar una señal detectable. Más deseablemente, la marca se detecta visualmente, e.g., por fluorescencia, para fácil uso en análisis inmunohistoquímico o microscopía de inmunofluorescencia. Por ejemplo, marcas adecuadas incluyen isotiocianato de fluoresceína (FITC), ficoeritrina (PE), aloficocianina (APC), corifosfina-O (CPO) o colorantes en tándem, PE-cianina-5 (PC5), y rojo Texas-PE (ECD). Todos estos colorantes fluorescentes están comercialmente disponibles, y sus usos son conocidos en el arte. Otras marcas útiles incluyen una marca de oro coloidal. Aún otras marcas útiles incluyen compuestos o elementos radiaactivos. Adicionalmente, las marcas incluyen una variedad de sistemas de enzimas que operan para revelar una señal colorimétrica

en un ensayo, e.g., la glucosa oxidasa (la cual usa glucosa como sustrato) libera peróxido como un producto, el cual en presencia de peroxidasa y un donor de hidrógenos tal como trimetil bencidina (TMB) produce una TMB oxidada que se ve de color azul. Otros ejemplos incluyen peroxidasa de rábano picante (HRP) o fosfatasa alcalina (AP), y hexocinasa en conjunción con glucosa-6-fosfato deshidrogenasa la cual reacciona con ATP, glucosa y NAD^+ para producir, entre otros productos, NADH que se detecta por incremento en la absorbancia a una longitud de onda de 340 nm.

Otros sistemas de marca que se utilizan en los métodos de esta invención son detectables por otros medios, e.g., micro-partículas de látex coloreadas [Bangs Laboratories, Indiana] en las cuales se embebe un colorante, se usan en lugar de enzimas para formar conjugados con las secuencias blanco proporcionando una señal visual indicativa de la presencia de los complejos que resultan en ensayos aplicables.

Los métodos para acoplar o asociar las marcas con la molécula deseada son similarmente convencionales y conocidos por aquellos ilustrados en el arte. Métodos conocidos de unión de marcas se describen en [Ver, por ejemplo, Handbook of Fluorescent Probes ande Research Chemicals, 6th Ed., R.P.M. Haugland, Molecular Probes, Inc., Eugene, OR, 1996; Pierce Catalog and Handbook, Life Science and Analytical Research Products, Pierce Chemical

Company, Rockford, IL, 1994/1995]. Por lo tanto la selección de la marca y los métodos de acoplamiento no limitan esta invención.

Las secuencias, proteínas, y fragmentos de la invención se pueden producir por cualquier medio adecuado, que incluyen producción recombinante, síntesis química, u otros medios de síntesis. Las técnicas de producción adecuadas son bien conocidas por aquellos versados en el arte. Ver, e.g., Sambrook et al, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Press (Cold Spring Harbor, NY). Alternativamente, también se pueden sintetizar péptidos por los bien conocidos métodos de síntesis de péptidos en fase sólida, (Merrifield, *J. Am. Chem. Soc.*, **85**: 2149 (1962); Stewart and Young, Solid Phase Peptide Synthesis (Freeman, San Francisco, 1969) pp. 27-62. Estos y otros métodos de producción adecuados están en el conocimiento de aquellos versados en el arte y no son una limitación de la presente invención.

En adición, alguien versado en el arte entenderá fácilmente que las secuencias Ad de la invención se pueden adaptar fácilmente para uso en una variedad de sistemas de vectores virales y no virales para entrega *in vitro, ex vivo o in vivo* de moléculas terapéuticas e inmunogénicas. Por ejemplo, en una incorporación, las proteínas cápside Ad de simio y otras proteínas de adenovirus de simio descritas aquí se usan para la entrega basada en proteínas no-virales de genes, proteínas, y otras moléculas deseables de diagnóstico, terapéuticas e inmunogénicas. En una de tales incorporaciones, una proteína de la invención se une, directamente o

indirectamente, a una molécula para apuntar a células con un receptor para adenovirus. Preferiblemente, una proteína cápside tal como un exón, un pentón, fibra o fragmento de la misma que tiene un ligando para un receptor de la superficie celular se selecciona para tal blanco. Moléculas adecuadas para la entrega se seleccionan de entre las moléculas terapéuticas descritas aquí y sus productos génicos. Una variedad de ligandos que incluyen lípidos, poliLys y similares, se pueden utilizar como ligandos. Por ejemplo, la proteína pentón de simio se puede utilizar fácilmente para tal propósito por producción de una proteína de fusión usando las secuencias pentón de simio en una manera análoga a la descrita en Medina-Kauwe LK, et al, *Gene Ther.* **2001** May; 8 (10): 795-803 y Medina-Kauwe LK, et al, *Gene Ther.* **2001** Dec:8(23):1753-1761. Alternativamente, las secuencias de aminoácidos de la Ad proteína IX de simio se pueden utilizar para vectores que apuntan a un receptor de la superficie celular, como se describe en la Patente US Aplicación 20010047081. Ligandos adecuados incluyen un antígeno CD40, uno que contiene RGD o la secuencia que contiene polilisina, y similares. Aún otras Ad proteínas, que incluyen, e.g., la proteína exón y/o la proteína fibra, se pueden utilizar para estos y propósitos similares.

Aún otras proteínas adenovirales de la invención se pueden usar solas, o en combinación con otras proteínas adenovirales, para una variedad de propósitos los cuales serán claros para alguien versado en el arte. En adición, aún otros usos para las proteínas adenovirales de la invención serán claros para alguien versado en el arte.

47
48

II Vectores Adenovirales Recombinantes.

Las composiciones de esta invención incluyen vectores que entregan una molécula heteróloga a las células, ya sea con propósitos terapéuticos o de vacuna. Como aquí se usa, un vector puede incluir cualquier elemento genético, que incluye, sin limitación, ADN desnudo, un fago, un transposon, un cósmido, un episoma, un plásmido o un virus. Tales vectores contienen un ADN de adenovirus de simio de Pan5, Pan6, Pan7, SV1, SV25 y/o SV39 y un minigen. Por"minigen" se entiende la combinación de un gen heterólogo seleccionado y de otros elementos reguladores necesarios para conducir la traducción, trascripción y/o expresión del producto del gen en una célula huésped.

Típicamente un vector adenoviral de la invención se designa tal que el minigen está localizado en una molécula de ácido nucleico el cual contiene otras secuencias adenovirales en la región nativa de un gen adenoviral seleccionado. El minigen se puede insertar en una región del gen existente para desorganizar la función de esa región, si se desea. Alternativamente, el minigen se puede insertar en el sitio de un gen adenoviral parcial o completamente suprimido. Por ejemplo, el minigen puede estar localizado en el sitio tal como el sitio de una supresión E1 funcional o una supresión E3 funcional, entre otras que se pueden seleccionar. El término"funcionalmente suprimido" o "supresión funcional" significa que una cantidad suficiente de la región génica se remueve o es dañada, e.g., por mutación o modificación, tal que la región génica no es más capaz de producir

productos funcionales de la expresión del gen. Si se desea la región entera del gen se puede remover. Otros sitios adecuados para la desorganización o supresión del gen se discuten en otra parte en la aplicación.

Por ejemplo, para una producción útil de un vector para la generación de un virus recombinante el vector puede contener el minigen o bien el terminal 5' del genoma adenoviral o el terminal 3' del genoma adenoviral, o ambos terminales 5' y 3' del genoma adenoviral. El terminal 5' del genoma adenoviral contiene los elementos cis 5' necesarios para el empacamiento y la replicación; i.e., las secuencias (ITR) de repetición terminal invertidas 5' (las cuales funcionan como origen de la replicación) y los dominios mejorados de empaquetamiento nativo 5' (que contienen las secuencias necesarias para el empaquetamiento lineal de genomas Ad lineales y elementos mejoradores para el promotor E1). La terminal 3' del genoma adenoviral incluye los elementos cis 3' (que incluyen los ITRs) necesarios para empaquetamientos y encapsidación. Adecuadamente un adenovirus recombinante contiene ambos elementos cis adenovirales el 5' y el 3' y el minigen está localizado entre las secuencias adenovirales 5' y 3'. Cualquier vector adenoviral de la invención puede también contener secuencias adenovirales adicionales.

Adecuadamente, estos vectores adenovirales de la invención contienen uno o más elementos adenovirales de un genoma adenoviral de la invención. En una incorporación, los vectores contienen ITRs adenovirales de Pan5, Pan6, Pan7,

SV1, SV25 o SV39 y secuencias adenovirales adicionales de los mismos serotipos adenovirales. En otra incorporación, los vectores contienen secuencias adenovirales que se derivan de un serotipo adenoviral diferente mas que los que proporcionan los ITRs. Como se define aquí, un adenovirus seudotipo se refiere a un adenovirus en el cual la proteína cápside del adenovirus es de un serotipo diferente del serotipo que proporciona los ITRs. La selección de los serotipos de los ITRs y los serotipos de cualquiera otra secuencia adenoviral presente en el vector no es una limitación de la presente invención. Una variedad de tipos de adenovirus están disponibles en American Type Culture Collection Manassas, Virginia, o disponibles por solicitud a una variedad de fuentes comerciales e institucionales. Además, las secuencias de algunas de tales tipos están disponibles de una variedad de bases de datos que incluyen e.g., PubMed y GenBank. Vectores de adenovirus homólogos preparados de otros adenovirus de simio o de humanos están descritos en la literatura publicada [ver, por ejemplo, Patente US No. 5, 240,846]. Las secuencias ADN de un número de tipos de adenovirus están disponibles en GenBank, incluyendo el tipo Ad5 [GenBank No de acceso M73260]. Las secuencias de adenovirus pueden ser obtenidas de cualquier serotipo de adenovirus conocido, tal como serotipos 2,3,4,7,12,y 40, y además incluyendo cualquiera de los tipos humanos identificados actualmente. Similarmente adenovirus conocidos para infectar animales no humanos (e.g., simios) también se pueden emplear en la construcción del vector de esta invención. Ver, e.g., Patente US No 6, 083,716.

Las secuencias virales, virus auxiliares si se necesitan, y partículas virales recombinantes, y otros componentes del vector y secuencias empleadas en la construcción de los vectores descritos aquí se obtienen como se describió anteriormente. Las secuencias de ADN de Pan5, Pan6, Pan7, SV1, SV25 y/o las secuencias de adenovirus de simio SV39 de esta invención se emplean para construir vectores y líneas celulares útiles en la preparación de tales vectores.

Modificaciones de las secuencias de ácido nucleico que forman los vectores de esta invención, incluyen supresión de secuencias, inserciónes y otras mutaciones se pueden generar usando las técnicas de biología molecular estándar y están en el alcance de esta invención.

A. *El "Minigen".*

Los métodos empleados para la selección del transgen, la clonación y la construcción del "minigen" y su inserción en el vector viral están dentro de la ilustración en el arte dando las enseñanzas aquí proporcionadas.

1. El Transgen

Es una secuencia de ácido nucleico, heteróloga a las secuencias del vector que flanquean el transgen, el cual codifica un polipéptido, proteína u otro producto de interés. La secuencia de ácido nucleico que codifica está operativamente unida a

S1
S2

componentes reguladores de una manera que permite la transcripción del transgen, la traducción, y/o la expresión en una célula huésped.

La composición de la secuencia transgénica dependerá del uso para el cual el se coloque el vector resultante. Por ejemplo, un tipo de secuencia transgénica incluye una secuencia reportera la cual produce sobre la expresión una señal detectable. Tales secuencias reporteras incluyen, sin limitación, secuencias de ADN que codifican -lactamasa, –galactosidasa (LacZ), fosfatasa alcalina, cinasa-timidina, proteína fluorescente verde (GFP), acetil-transferasa cloranfenicol (CAT), luciferasa, proteínas unidas a membranas incluyendo, por ejemplo, CD2, CD4, CD8, proteína hemaglutinina de la influenza, y otras bien conocidas en el arte, las cuales tienen anticuerpos de alta afinidad dirigidos a existir o poder ser producidos por medios convencionales, y proteínas de fusión que comprenden una proteína unida a una membrana apropiadamente fundida a un dominio marca de un antígeno de, entre otros, hemaglutinina o Myc. Estas secuencias codifican, en asociación con elementos reguladores los cuales dirigen sus expresiones, proporcionan señales detectables por medios convencionales incluyendo ensayos enzimáticos, radiográficos, colorimétricos, de fluorescencia, u otros ensayos espectrográficos, y ensayos de clasificación celular por activación de fluorescencia, y ensayos inmunológicos, incluyendo una enzima unida a un ensayo inmunosorbente (ELISA), un radioinmunoensayo (RIA) e inmunohistoquímico. Por ejemplo, donde la secuencia marcadora es el gen LacZ, la presencia del vector que lleva la señal se detecta por ensayos de la actividad de

la beta-galactosidasa. Donde el transgen es GFP o luciferasa, el vector que lleva la señal se puede medir visualmente por color o por la producción de luz en un luminómetro.

Sin embargo, deseablemente, el transgen es una secuencia no marcadora que codifica un producto que es útil en biología y medicina, tal como proteínas, péptidos, ARN, enzimas, o ARNs catalíticos. Moléculas de ARN deseables que incluyen tARN, dsARN, ARN ribosomal, ARNs catalíticos, y ARNs contrasentido. Un ejemplo de una secuencia ARN útil es una secuencia, la cual extingue la expresión de una secuencia de ácido nucleico puesta como blanco en el animal tratado.

El transgen puede ser útil para tratamiento e.g., de deficiencias genéticas, como un terapéutico del cáncer o vacuna, por inducción de una respuesta inmune, y/o, para propósitos de vacuna profiláctica. Como aquí se usa, la inducción de una respuesta inmune se refiere a la capacidad de una molécula (e.g., un producto génico) para inducir una célula T y/o una respuesta inmune humoral a la molécula. La invención, además incluye utilizando múltiples transgenes, e.g., para corregir o mejorar una condición de una proteína de subunidad múltiple. En ciertas situaciones un transgen diferente se puede usar para codificar cada subunidad de una proteína o para codificar diferentes peptidos o proteínas, esto es deseable cuando el tamaño del ADN que codifica la subunidad de proteína es grande, e.g., para una inmunoglobulina, el factor de crecimiento derivado de la plaqueta o una

proteína distrofina. En orden a que la célula produzca la proteína de subunidad múltiple, una célula se infecta con el virus recombinante que contiene cada una de las diferentes subunidades. Alternativamente, diferentes subunidades de una proteína pueden ser codificadas por el mismo transgen. En este caso, un gen particular incluye el ADN que codifica cada una de las unidades, con el ADN para cada subunidad separada por un sitio de entrada al ribosoma interno (IRES). Esto es deseable cuando el tamaño del ADN que codifica cada una de las subunidades es pequeño, e.g., el tamaño total del ADN que codifica las subunidades y el IRES es menor de 5 Kb. Como una alternativa a un IRES, el ADN se puede separar por secuencias que codifican un péptido 2 A, el cual se autodegrada en un evento post-traduccional. Ver, e.g., M. L. Donnelly, *et al, J. Gen. Virol.*, **78** (Pt 1): 13-21 (Jan 1997); Furler, S., *etal, Gene Ther.*, **8** (11): 864-873 (Jun 2001); Klump H., *et al., Gene Ther;* **8** (10): 811-817 (May 2001). Este péptido 2A es significativamente más pequeño que un IRES, haciéndolo bien adecuado para utilizarlo cuando el espacio es un factor limitante. Sin embargo el transgen seleccionado puede codificar cualquier producto biológicamente activo u otro producto, e.g., un producto deseable para estudio.

Genes adecuados se pueden seleccionar fácilmente por alguien versado en el arte. La selección del transgen no se considera limitante de esta invención.

54
55

2. Elementos Reguladores

En adición a los elementos principales identificados antes para el minigen, el vector también incluye elementos de control convencional necesarios los cuales están unidos operablemente al transgen de una manera que permite su transcripción, traducción y/o expresión en una célula transfectada con el vector plásmido o infectada con el virus producido por la invención. Como aquí se utiliza las secuencias "unidas operablemente" incluyen ambas las secuencias de control de expresión que están contiguas al gen de interés y las secuencias de control de expresión que actúan en *trans* o a una distancia para controlar el gen de interés.

Las secuencias de control de expresión incluyen apropiadas secuencias promotoras y mejoradoras de transcripción, iniciación, terminación; las señales de procesamiento eficientes del ARN tales como señales de unión y señales de poliadenilación (poliA), secuencias que estabilizan el mARN citoplasmático; secuencias que mejoran la eficiencia de traducción, (i.e., la secuencia de consenso de Kozak); secuencias que mejoran la estabilidad de la proteína; y cuando se desean secuencias que mejoran la secreción del producto codificado. Un gran número de secuencias del control de expresión que incluyen promotores los cuales son nativos, constitutivos, inducibles, y/o específicos para tejidos, se conocen en el arte y se pueden utilizar.

Ejemplos de promotores constitutivos incluyen, sin limitación, el promotor LTR del virus (RSV) del sarcoma Rous (opcionalmente con el mejorador RSV) el promotor del citomegalovirus (CMV) (opcionalmente con el mejorador CMV)[ver, e.g., Boshart *et al, Cell,* **41**: 521-530 (1985)], el promotor SV40, el promotor de la dihidrofolato reductasa, el promotor de la -actina, el promotor de la fosfoglicerocinasa (PGK) y el promotor de EF1 [Invitrogeno].

Promotores inducibles que permiten la regulación de la expresión génica y pueden ser regulados por compuestos suministrados exógenamente, factores ambientales, tales como la temperatura, o la presencia de un estado fisiológico específico, e.g., fase aguda, un estado de diferenciación particular de la célula o solamente en células en replicación. Los promotores inducibles y sistemas inducibles estan disponibles en una variedad de fuentes comerciales, que incluyen, sin limitación, Invitrogen, Clontech y Ariad. Algunos otros sistemas se han descrito y pueden ser fácilmente seleccionados por alguien versado en el arte. Por ejemplo, promotores inducibles incluyen metalotionina de oveja inducible por cinc (MT) y el promotor del virus del tumor mamario de ratón (MMTV) inducible por dexametasona (Dex). Otros sistemas inducibles incluyen rl sistema promotor de polimerasa T7 [WO 98/10088] el promotor del ecdysone de los insectos [No *et al, Proc. Natl. Acad. Sci. USA,* **93**: 3346-3351 (1996)], sistema tetraciclina-represible [Gossen *et al, Proc. Natl. Acad. Sci. USA,* **89**: 5547-5551 (1992)], sistema tetraciclina-inducible [Gossen *et al, Science,* **268:1766**-1769(1995), ver también, Harvey *et al,*Curr. Opin. Chem. Biol., **2**:512-518 (1998)]. Otros sistemas

incluyen el dímero FK506, VP16 o p65 usando castradiol, murislerona difenol, el sistema inducible-RU486 [Wang *et al Nat. Biotech.*, 15: 239-243 (1997) y Wang *et al, Gene Ther.*, 4:432-441 (1997)] y el sistema rapamicina-inducible [Magar i *et al, J.Clin. Invest.*, 100:2865-2872(1997). La efectividad de algunos promotores inducibles se aumenta en un cierto tiempo. En tales casos se puede mejorar la efectividad de tales sistemas insertando múltiples represores en tándem, e.g., TetR unido a TetR por un IRES. Alternativamente, uno puede esperar al menos 3 días antes de tamizar para la selección para la función deseada. Uno puede mejorar la expresión de las proteínas deseadas al conocer los medios para mejorar la efectividad de este sistema. Por ejemplo, utilizando el elemento regulador Postranscripcional del Virus de la Hepatitis Woodchuk (WPRE).

En otra incorporación, será utilizado el promotor nativo para el transgen. El promotor nativo se puede preferir cuando se desea que la expresión del transgen deba imitar la expresión del nativo. El promotor nativo se puede usar cuando la expresión del transgen debe ser regulada temporalmente o en el desarrollo, o de una manera específica para el tejido, o en respuesta a un estímulo transcripcional específico. En otra incorporación, otros elementos de control de la expresión nativos, tales como elementos mejoradores, secuencias de consenso Kosak o sitios de poliadenilación se pueden utilizar para imitar la expresión nativa.

Otra incorporación del transgen incluye un transgen unido operablemente a un promotor de tejido específico. Por ejemplo, si se desea la expresión en el músculo

esquelético, se debe usar un promotor activo en el músculo. Estos incluyen los promotores de los genes que codifican -actina esquelética, la cadena 2 A liviana de miosina, distrofina, creatin-cinasa del músculo, así como también promotores musculares sintéticos con actividades mas altas que los promotores de ocurrencia natural (ver Li *et al., Nat biotech.*,17:241-245 (1999)). Ejemplos de promotores que son específicos para tejido son conocidos para el hígado (albúmina, Miyatake *et al., J.Virol.* 71:5124-32 (1997); promotor base del virus de la hepatitis B, Sandig *et al., Gene Ther.*, 3: 1002-9 (1996); Alfa-fetoproteína (AFP), Arbuthnot *et al., Hum. Gene Ther.*, 7:1503-14 (1996), osteocalcina de hueso (Stein *et al., Mol. Biol. Rep.*, 24:185-96 (1997); sialoproteína de hueso (Chen *et al., J. Bone Miner. Res.*, 11:654-664 (1996), linfocitos (CD2) Hansal *et al., J. Immunol.* 161:1063-8 (1998); cadenas pesadas de inmunoglobulinas; células T cadena receptora), neuronal tal como el promotor de la enolasa específica para la neuronas (NSE) (Andersen *et al., Cell. Mol. Neurobiol.*, 13:503-15(1993)), gen de cadena liviana del neurofilamento (Piccioli *et al., Proc. Natl. Acad. Sci.USA*, 88:5611-5 (1991)), y el gen vgf neurona-específico (Piccioli *et al., Neuron*, 15:373-84 (1995), entre otros.

Opcionalmente, vectores que llevan transgenes que codifican productos útiles terapéuticamente o inmunogénicos pueden también incluir marcadores seleccionables o genes reporteros que pueden incluir secuencias que codifican geneticina, higromicina o de resistencia a la puromicina, entre otros. Tales reporteros seleccionables o genes marcadores (preferiblemente localizados fuera del genoma viral para ser empacados en una partícula viral) se pueden usar para

señalar la presencia de los plásmidos en las células bacterianas, tales como los de resistencia a la ampicilina. Otros componentes del vector pueden incluir un origen de replicación. La selección de estos y otros promotores y elementos de vectores son convencionales y están disponibles muchas de tales secuencias [ver, e.g., Sambrook et al, y las referencias citadas aquí].

Estos vectores se generan utilizando las técnicas y secuencias propuestas aquí, en conjunción con las técnicas conocidas por aquellos versados en el arte. Tales técnicas incluyen técnicas de clonación convencionales de cADN tales como aquellas que se describen en los textos [Sambrook et al, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Press, Cold Spring Harbor, NY], uso del sobrelapamiento de secuencia de oligonucleotidos de los genomas de adenovirus, reacción en cadena de polimerasa, y cualquier método adecuado que proporciona la secuencia de nucleótidos deseada.

III. Producción de la Partícula Viral Recombinante

En una incorporación, los plásmidos adenovirales de simio (u otros vectores) se utilizan para producir partículas adenovirales recombinantes. En una incorporación, los adenovirus recombinantes son funcionalmente suprimidos en los genes E1a o E1b, y opcionalmente soportan otras mutaciones, e.g., mutaciones por sensibilidad a la temperatura o supresiones en otros genes. En otras incorporaciones, es deseable retener una región intacta E1a y/o E1b en los

adenovirus recombinantes. Tal región intacta E1 se puede localizar en su locación nativa en el genoma adenoviral o colocada en el sitio de una supresión en el genoma adenoviral nativo (e.g., en la región E3).

En la construcción de vectores de adenovirus de simio útiles para la entrega de un gen a una célula humana (u otro mamífero), se pueden emplear en los vectores un intervalo de secuencias de adenovirus de ácido nucleico. Por ejemplo, toda o una porción del adenovirus retardado tempranamente del gen E3 se puede eliminar de la secuencia del adenovirus de simio la cual forma parte de un virus recombinante. La función E3 de simio se cree sea irrelevante para la función y producción de la partícula de virus recombinante. Los vectores de adenovirus de simio también se pueden construir teniendo una supresión de por lo menos la región ORF6 del gen E4, y mas deseablemente a causa de la redundancia en la función de esta región, de la región completa E4. Aún otro vector de esta invención contiene una supresión el gen E2a tempranamente retardado. Las supresiones también se pueden hacer en cualquiera de los genes retardados L1 hasta L5 del genoma adenovirus de simio. Similarmente, las supresiones en los genes intermedios IX y IVa_2 pueden ser útiles para algunos propósitos. Otras supresiones se pueden hacer en los otros genes de adenovirus estructurales o no estructurales. Las supresiones discutidas anteriormente se pueden utilizar individualmente, i.e., una secuencia de adenovirus para uso en la presente invención, puede contener supresiones en solamente una región particular. Alternativamente, supresiones de genes completos o porciones de los mismos

efectivas para destruir sus actividades biológicas se pueden usar en cualquier combinación. Por ejemplo, en un vector ejemplar, la secuencia del adenovirus puede tener supresiones de los genes E1y el gen E4, o de los genes E1, E2a y E3, o de los genes E1 y E3, o de los genes E1, E2a y E4, con o sin supresión de E3, y etc. Como se discutió antes, tales supresiones pueden ser utilizadas en combinación con otras mutaciones, tales como mutaciones por sensibilidad a la temperatura, para alcanzar los resultados deseados.

Un vector adenoviral que carece de cualquier secuencia adenoviral esencial (e.g., E1a, E1b, E2a, E2b, E5 ORF6, L1, L2, L3, L4 y L5) se puede cultivar en presencia de los productos génicos adenovirales que se perdieron los cuales se requieren para la infectidad viral y la propagación de una partícula adenoviral. Estas funciones auxiliares pueden ser proporcionadas cultivando el vector adenoviral en presencia de uno o más constructos auxiliares (e.g., un plásmido, o virus) o una célula huésped empacada. Ver, por ejemplo, las técnicas descritas para la preparación de un vector Ad humano "mínimo" en Internacional Patent Application WO96/13597, publicada en mayo 9, 1996, e incorporada aquí por referencia.

1. Virus Auxiliares.

Así, dependiendo del contenido génico del adenovirus de simio de los vectores virales empleados para llevar el minigen, puede ser necesario un adenovirus auxiliar o fragmento de virus que no replica para proporcionar suficientes secuencias génicas de adenovirus de simio necesarias para producir una partícula viral recombinante infectiva que contenga el minigen. Virus auxiliares útiles contienen secuencias génicas de adenovirus seleccionadas no presentes en el constructo del vector de adenovirus y/o no expresadas por la línea celular empacada en la cual el vector es transfectado. En una incorporación, el virus auxiliar es defectuoso en la replicación y contiene una variedad de genes de adenovirus en adición a las secuencias descritas anteriormente. Tale virus auxiliar se usa deseablemente en combinación con una línea celular que expresa a E1.

Los virus auxiliares también se pueden formar en conjugados policatiónicos como se describe en Wu *et al, J. Biol. Chem.*, **264**: 16985-16987 (1989); K. J. Fisher and J. M. Wilson, *Biochem. J.*,**299: 49** (April 1, 1994). Los virus auxiliaries opcionalmente pueden contener un Segundo minigen reportero. Un número de tales genes reporteros son conocidos en el arte. La presencia de un gen reportero sobre el virus auxiliar el cual es diferente del transgen sobre el vector del adenovirus permite que ambos el vector Ad y el virus auxiliar sean monitoreados independientemente. El segundo reportero se usa para permitir la separación entre el virus recombinante resultante y el virus auxiliar en la purificación.

62
63

2. Complementación de Líneas Celulares

Para generar adenovirus de simio recombinante (Ad) suprimido en cualquier de los genes descritos anteriormente, la función de la región del gen borrada, si es necesaria para la replicación la infectividad del virus debe ser suministrada para el virus recombinante por un virus auxiliar o línea celular, i.e., una línea celular empacada o de complementación. En muchas circunstancias, una línea celular que expresa el E1 humano se puede usar para transcomplementar el vector Ad de chimpancé. Esto es particularmente ventajoso porque, debido a la diversidad entre las secuencias Ad de chimpancé de la invención y las secuencia AdE1 humanas encontradas en células empacadas disponibles corrientemente, el uso de las células que contienen E1 humano corrientes previene, la generación de adenovirus competentes para la replicación durante los procesos de replicación y producción. Sin embargo en ciertas circunstancias, será deseable utilizar una línea celular la cual exprese los productos génicos E1 que se pueda utilizar para la producción de un adenovirus de simio E1-suprimidos. Tales líneas celulares han sido descritas. Ver, e.g., Patente US 6,083, 716.

Si se desea, uno puede utilizar las secuencias proporcionadas aquí para generar células o líneas celulares que expresen, en un mínimo, el gen E1 de adenovirus de Pan5, Pan6, Pan7, SV1, SV25 o SV39 bajo el control transcripcional de un

promotor de expresión en una línea celular padre seleccionada. Promotores inducibles o constituvos se pueden emplear para este propósito. Ejemplos de tales promotores están descritos en detalle en su lugar en esta especificación. Una célula padre se selecciona para la generación de una nueva línea celular que exprese cualquier gen deseado Pan5, Pan6, Pan7, SV1, SV25 o SV39. Sin limitación, tal línea celular padre puede ser células Hela [ATCC No. de acceso CCL2], A549 [ATCCC No. De acceso CCL185], HEK 293, KB [CCL 17], Detroit [e.g., Detroit 510, CCL 72] y WI-38 [CCL 75]. Estas líneas celulares están todas disponibles de American Type Culture Collection, 10801 University Boulevard, Manassas, Virginia 20110-2209. Otras líneas celulares padre adecuadas se pueden obtener de otras fuentes.

Tales líneas celulares que expresan E1 son útiles en la generación de vectores suprimidos E1 de adenovirus de simio recombinante. Adicionalmente, o alternativamente, la invención proporciona líneas celulares que expresan uno o más productos génicos adenovirales de simio, e.g., E1a, E1b, E2a, E2b y/o E4ORF6, se pueden construir utilizando esencialmente los mismos procedimientos para uso en la generación de vectores virales de simio recombinantes. Tales líneas de células se pueden utilizar para transcomplementar, vectores virales suprimidos en los genes esenciales que codificaban aquellos productos, o para proporcionar funciones necesarias para empacar los virus dependientes del auxiliar (e.g., virus adeno-asociados). La preparación de las células huésped de acuerdo con la invención implica técnicas tales como ensamble de secuencias de

ADN seleccionadas. Este ensamble se puede acompañar de la utilización de técnicas convencionales. Tales técnicas incluyen clonación de cADN y genómica, las cuales son bien conocidas y se describen en Sambrook et al., citado antes, el uso de secuencias de oligonucleotidos de translapamiento de los genomas de adenovirus, combinadas con la reacción en cadena de polimerasa, métodos sintéticos, y otros métodos adecuados que proporcionan la secuencia de nucleótidos necesaria.

En aún otra alternativa, los productos de gen adenovirales esenciales se proporcionan en *trans* por el vector adenoviral y/o por el virus auxiliar. En tal caso, una célula huésped apropiada se puede seleccionar de cualquier organismo biológico, incluyendo células procariotas (e.g., bacteriales), células eucariotas, incluyendo, células de insectos, células de levaduras y células de mamíferos. Particularmente las células huésped deseables son las seleccionadas de entre cualquier especie de mamífero. Incluyendo sin limitación, células tales como A549, WEHI, 3T3, 10T1/2, células HEK 293, o PERC6 (ambas expresan E1 adenoviral funcional) [Fallaux, FJ *et al.*, (1998), *Hum Gene Ther*, 9:1909-1917], Saos, C2C12, células L, HT1080, HepG2 y fibroblastos primarios, células de hepatocitos y mioblastos derivadas de mamíferos incluyendo humanos, monos, ratón, rata, conejo y hámster. La selección de las es'pecies de mamíferos que proporcionan las células no es una limitación de esta invención; ni tampoco lo es el tipo de célula del mamífero, i.e., fibroblasto, hepatocito, célula tumoral, etc.

3. Ensamble de la Partícula viral y Transfección de una línea celular

Generalmente, cuando se entrega el vector que comprende el minigen por transfección, el vector se entrega en una cantidad desde cerca de 5 g hasta cerca de 100 g de ADN y preferiblemente cerca de 10 g hasta cerca de 50 g de ADN a a cerca de 1 x10^4 células, hasta cerca 1 x 10^{13}células y preferiblemente cerca de 10^5células. Sin embargo, las cantidades relativas de vector ADN a las células huésped se pueden ajustar, tomando en consideración factores tales como el vector seleccionado, el método de entrega y las células huésped seleccionadas.

El vector puede ser cualquier vector conocido en el arte o divulgado anteriormente., que incluye ADN desnudo, un plásmido, fago, transposon, cósmidos, episomas, virus, etc.

La introducción en la célula huésped del vector puede ser alcanzada por cualquier medio conocido en el arte o como se divulgó antes. Que incluyen transfección, e infección. Uno o más de los genes adenovirales puede ser integrado establemente en el genoma de la célula huésped, establemente expresados como episomas o expresados transitoriamente. Todos los productos génicos pueden ser expresados transitoriamente, sobre un episoma o integrados establemente, o algunos de los productos génicos pueden ser expresados establemente mientras otros son expresados transitoriamente. Además los promotores para cada uno de los genes

adenovirales se puede seleccionar independientemente desde un promotor constitutivo, un promotor inducible, o un promotor adenoviral nativo. Los promotores pueden ser regulados por estados fisiológicos específicos del organismo o de la célula (i.e., por el estado de diferenciación, o en replicación o quietud de las células) o por factores adicionados exógenamente, por ejemplo.

La introducción de las moléculas (como plásmidos o virus) en la célula huésped también se puede realizar utilizando las técnicas conocidas por los versados en el arte y como se ha discutido a lo largo de la especificación. En una incorporación preferida, se usan las técnicas de transfección estándar, e.g., transfección por $CaPO_4$ o electroporación.

La unión de las secuencias de ADN seleccionadas del adenovirus, así como también el transgen y otros elementos del vector en varios plásmidos intermedios y el uso de los plásmidos y vectores para producir una partícula viral recombínate se logran todas usando técnicas convencionales. Tales técnicas incluyen técnicas de clonación convencional de cADN tales como aquellas descritas en los textos [Sambrook et al, citado anteriormente] el uso de sobrelapamiento de las secuencias de oligonucleotidos de los genomas de los adenovirus, reacción en cadena de la polimerasa, y cualquier método adecuado que proporcione la secuencia de nucleótidos deseada. Se emplean las técnicas de transfección estándar y co-transfección, e.g., técnicas de precipitación con $CaPO_4$. Otros métodos convencionales que se emplean incluyen recombinación homóloga de los

genomas virales, plaqueo de virus en recubrimientos de agar, métodos de generación de señales que se miden, y similares.

Por ejemplo, siguiendo la construcción y unión de los vectores virales que contienen los minigenes deseados, el vector se transfecta *in vitro* en presencia de un virus auxiliar en la línea celular empacada. La recombinación homóloga ocurre entre el auxiliar y las secuencias del vector, lo cual permite que las secuencias transgénicas del adenovirus en el vector se repliquen y empaquen en cápsides virión, produciendo las partículas del vector viral recombinante. El método corriente para la producción de tales partículas de virus se basa en la transfección. Sin embargo, la invención no se limita a tales métodos.

Loa adenovirus de simio recombinantes que resultan son útiles transfiriendo un transgen seleccionado a una célula seleccionada. En experimentos *in vivo* con el virus recombinante creciendo en las líneas celulares empacadas, los vectores adenovirales de simio recombinantes E1-suprimidos de la invención demuestran utilidad en la transferencia de un transgen a un no simio, preferiblemente a una célula humana.

IV. Us de Vectores de Adenovirus Recombinantes.

Los vectores de adenovirus de simio recombinantes de la invención son útiles para transferir un gen a un paciente humano o veterinario no simio *in vitro, ex vivo,* e *in vivo.*

Los vectores de adenovirus recombinantes descritos aquí se pueden usar como vectores de expresión para la producción de productos codificados por los genes heterólogos *in vitro.* Por ejemplo, los adenovirus recombinantes que contienen un gen insertado en la locación de una supresión E1 pueden ser transfectados en una línea celular que exprese E1 como se describió anteriormente. Alternativamente, los adenovirus de replicación competente se pueden usar en otras líneas celulares seleccionadas. Luego las células transfectadas se cultivan de la manera convencional, permitiendo que el adenovirus recombinante exprese el producto génico del promotor. Entonces el producto génico se puede recuperar del medio de cultivo por los métodos convencionales conocidos de extracción de proteínas y recuperación desde el cultivo.

Los vectores adenovirales de simio recombinantes derivados de Pan5, Pan6, Pan7, SV1, SV25 o SV39 de la invención proporcionan un vehículo de transferencia génica que puede entregar un transgen seleccionado en una célula huésped seleccionada *in vivo o ex vivo* donde el organismo tiene anticuerpos que neutralizan a uno o más serotipos AAV. En una incorporación, el rAAV y las

células se mezclan *ex vivo;* las células infectadas se cultivan utilizando metodologías convencionales; y las células transducidas son re-infundidas en el paciente. Estas composiciones son particularmente bien situadas para la entrega de un gen para propósitos terapéuticos y para inmunización, que incluye inmunidad protectora inducida.

Más comúnmente los vectores adenovirales de Pan5, Pan6, Pan7, SV1, SV25 o SV39 de la invención se utilizarán para la entrega de moléculas terapéuticas o inmunogénicas, como se describe mas adelante. Se entenderá fácilmente para ambas aplicaciones, que los vectores adenovirales recombinantes de la invención son particularmente bien apropiados para uso en regimenes que implican entregas repetidas de los vectores adenovirales recombinantes. Típicamente, tales regimenes implican la entrega de una serie de vectores virales en los cuales las cápsides virales están alteradas. Las cápsides virales se pueden cambiar para cada administración subsiguiente, o después de un número preseleccionado de administraciones de una cápside de un serotipo particular (e.g., una, dos, tres, cuatro o más). Así, un régimen puede implicar la entrega de un rAd con una primera cápside de simio, la entrega de un rAd con una segunda cápside de simio y la entrega de un rAd con una tercera cápside de simio. Una variedad de otros regimenes los cuales utilizan las cápsides Ad, de la invención, solas, en combinación con otra, o en combinación con otros serotipos Ad serán claras para aquellos versados en el arte. Opcionalmente, tal régimen puede implicar la administración de Ad con cápsides de otros adenovirus de primates no humanos,

adenovirus humanos, o serotipos artificiales tales como los aquí descritos. Cada fase del régimen puede implicar la administración de una serie de inyecciones (u otras rutas de entrega) con una cápside de serotipo Ad particular seguida de una serie con otras cápsides de serotipo Ad. Alternativamente, los vectores Ad recombinantes de la invención se pueden utilizar en regimenes que implican sistemas de entrega mediados no adenovirales, que incluyen otros sistemas virales, sistema de entrega no viral, proteína, peptidos, y otras moléculas activas biológicamente.

Las siguientes secciones se enfocarán en las moléculas ejemplares las cuales pueden ser entregadas vía vectores adenovirales de la invención.

A. Moléculas Terapéuticas de Entrega Ad.-Mediadas.

En una incorporación, los vectores recombinantes descritos anteriormente se administran a humanos de acuerdo con los métodos publicados para terapia génica. Un vector viral de simio que porta el transgen seleccionado se puede administrar a un paciente, preferiblemente suspendido en una solución compatible biológicamente o en un vehículo de entrega farmacéuticamente aceptable. Un vehículo adecuado incluye solución salina estéril. Otras soluciones estériles isotónicas no acuosas para inyección y suspensiones estériles no acuosas conocidas que sean vehículos farmacéuticamente aceptables y bien conocidos por aquellos versados en el arte se pueden emplear para éste propósito.

Los vectores adenovirales de simio se administran en cantidades suficientes para transducir las células que actúan como blanco y para proporcionar niveles suficientes de transferencia y expresión del gen para proporcionar un beneficio terapéutico sin adversos indebidos o con efectos fisiológicos médicamente aceptables, los cuales pueden ser determinados por aquellos versados en las artes médicas. Las rutas de administraciones convencionales y aceptables farmacéuticamente incluyen, pero no están limitadas a, rutas de entrega directa a la retina, y otros métodos de entrega intraocular, entrega directa al hígado, inhalación, intranasal, endovenosa, intramuscular, intratraqueal, subcutánea, intradérmica, oral y o parenteral. Las rutas de administración se pueden combinar, si se desea, o ajustar dependiendo del transgen o la condición. La ruta de administración dependerá primariamente de la naturaleza de la condición que este siendo tratada.

Las dosificaciones del vector viral dependerán primariamente de factores tales como la condición que este siendo tratada, la edad, el peso y la salud del paciente, y puede así variar entre los pacientes. Por ejemplo, una dosificación terapéuticamente efectiva para un humano adulto o una dosis veterinaria del vector viral está generalmente en el intervalo de cerca de 100 L a cerca de 100 mL de un vehículo que contiene concentraciones de cerca de 1×10^{6} a cerca de 1×10^{15} partículas, de cerca de 1×10^{11} a 1×10^{13} partículas, o de cerca de 1×10^{9} a 1×10^{12} partículas de virus. El intervalo de la dosificación dependerá del tamaño del animal y de la tuta de administración. Por ejemplo, una dosificación humana

o veterinaria adecuada (para un animal de cerca de 80 Kg) para inyección intramuscular está en el intervalo de cerca de 1×10^9 a cerca de 5×10^{12} partículas por mL, para una situación particular. Opcionalmente, múltiples sitios de administración pueden ser entregados. En otro ejemplo una dosificación humana o veterinaria adecuada puede estar en el intervalo de de cera de 1×10^{11} a 1×10^{15} partículas para una formulación oral. Alguien versado en el arte puede ajustar estas dosificaciones, dependiendo de la ruta de administración, y de la aplicación terapéutica o vacuna para la cual se emplea el vector recombinante. Los niveles de expresión del transgen, o para un inmunógeno, el nivel de un anticuerpo circulando, se puede monitorear para determinar la frecuencia de administración de la dosificación. Aún otros métodos para determinar la regulación del tiempo de frecuencia serán claros para alguien versado en el arte.

Un método adicional por etapas implica la co-administración al paciente, ya sea concurrentemente con, o antes o después de la administración del vector viral, de una cantidad adecuada de un modulador inmune de acción corta. El modulador inmune seleccionado se define aquí como un agente capaz de inhibir la formación de anticuerpos neutralizantes dirigidos contra el vector recombinante de esta invención o capaz de inhibir la eliminación linfocitos T citolíticos (CTL) del vector. El modulador inmune puede interferir con las interacciones entre los subconjuntos auxiliares T (T_{H1} o T_{H2}) y las células B para inhibir la formación de anticuerpo neutralizante. Alternativamente, el modulador inmune puede inhibir la interacción entre las células T_{H1} y las CTLs para reducir la ocurrencia de eliminación CTL del

vector. Una variedad de moduladores inmunes útiles y dosificaciones para uso de los mismos se divulgan, por ejemplo, en Yang *et al, J. Virol.*, **70** (9) (Sept., 1996); Internacional Patent Application No. WO 96/12406, publicada en mayo 2 de 1996; Patente Internacional Aplicación No. PCT/US96/03035, todas incorporadas aquí por referencia.

1. Transgenes Terapéuticos

Productos terapéuticos útiles codificados por el transgen incluyen hormonas y factores de crecimiento y diferenciación que incluyen, sin limitación, insulina, glucagón, hormona de crecimiento (GH) , hormona paratiróidea (PTH), factor de liberación de hormona de crecimiento (GRF), hormona estimulante del folículo (SH), hormona luteinizante (LH), gonadotropina coriónica humana (HCG), factor de crecimiento del endotelio vascular (VEGF), angiopoyetinas , angiostatina, factor estimulante de colonia granulocítica (GCSF), eritropoyetinas (EPO), factor de crecimiento de tejido conectivo (CTGF), factor de crecimiento de fibroblasto básico (bFGF), factor de crecimiento de fibroblasto ácido (aFGF), factor de crecimiento epidérmico (EGF), factor de crecimiento transformante (TGF), factor de crecimiento derivado de plaquetas (PDGF), factores de crecimiento de insulina I y II (IGF-I e IGF-II), cualquiera de la superfamilia de factores de crecimiento transformante, incluyendo TGF, activinas, inhibinas, o cualquiera de las proteínas morfogénicas del hueso (BMP) BMPs 1-15, cualquiera de los factores de crecimiento de la familia (NDF) de los factores de diferenciación

heregluina/neurogulina/ARIA/neu, factores de crecimiento del nervios (NGF), factor neurotrófico derivado del cerebro (BDNF), neurotrofinas NT-3 y NT-4/5, factor neurotrófico ciliar (CNTF), factor neurotrófico derivado de la línea celular glial (GDNF), neurturina, agrina, cualquiera de la familia de las semaforinas/colapsinas, netrina-1 y netrina-2, factor de crecimiento de hepatocitos (HGF), efrinas, noggi, hedgehoh sónico y tirosina hidroxilasa.

Otros productos transgénicos útiles incluyen proteínas que regulan el sistema inmune las cuales incluyen, sin limitación, citoquinas y linfoquinas tales como trombopoietina (TPO), interleuquinas (IL) IL-1 hasta IL-25 (incluyendo e.g., IL-2, IL-4, IL-12, e IL-18), proteína quimoatractora monocítica, factor inhibidor de la leucemia, factor estimulante de la colonia macrofago granulocítica, ligando Fas, factores de necrosis tumoral e, interferones, y, factor de céluas de tallo, ligandos flk-2/flt3. Los productos génicos producidos por el sistema inmune también son útiles en la invención. Estos incluyen, sin limitación, inmunoglobulinas IgG, IgM, IgA, IgD e IgE, inmunoglobulinas quiméricas, anticuerpos humanizados, anticuerpos de cadena liviana, receptores de células T, receptores de células T quiméricas, receptores de células T de cadena sencilla, moléculas MHC de clase I y clase II, así como también inmunoglobulinas por ingeniería y moléculas MHC. Los productos génicos útiles también incluyen proteínas reguladoras de complemento tales como proteínas reguladoras de complemento, proteínas de cofactor de membrana (MCP), factor acelerante de decaimiento (DAF), CR1, CF2 y CD59.

Aún otros productos génicos incluyen cualquiera de los receptores para hormonas, factores de crecimiento, citoquinas, linfoquinas, proteínas reguladoras y proteínas del sistema inmune. La invención abarca receptores para regulación de colesterol, que incluyen receptor de lipoproteínas de baja densidad (LDL), receptor de lipoproteínas de alta densidad (HDL), receptor de lipoproteínas de muy baja densidad (VLDL), y el receptor limpiador. La invención también abarca productos génicos tales como miembros de la superfamilia de receptores de hormonas esteroidales que incluyen receptores de glucocorticoides y receptores de estrógeno, vitamina D y otros receptores nucleares. En adición, los productos génicos útiles incluyen factores de transcripción tales como *jun, fos,* max, mad, factor de respuesta del suero (SRF), AP-1, AP2, *myb,* MyoD, y miogenina. Proteínas que contienen caja ETS, TFE3, E2F, ATF1, ATF2, ATF3, ATF4, ZF5, NFAT, CREB, HNF-4, C/EBP, SP1, proteínas enlazantes de caja CCAAT, factores de regulación del interferón (IRF-1), proteínas del tumor de Wilms, proteínas enlazantes de ETS, STAT, proteínas enlazantes de caja de GATA, e,g., GATA-3, y la familia de cabeza horquillada de proteínas de hélice alada.

Otros productos génicos incluyen, carbamoil sintetasa I, ornitina transcarbamilasa, arginosuccinato sintetasa, arginosuccinato liasa, arginasa, fumarilacetoacetato hidrolasa, fenilalanina hidroxilasa, alfa-1-antitripsina, glucosa-6-fosfatas, porfobilinógeno desaminasa, factor VIII, factor IX, cistationa beta-sintasa, descarboxilasa de cetoácidos de cadena ramificada, albúmina, isovaleril-CoA deshidrogenas, propionilCoA carbixilasa, metil malonilCoA mutasa, glutarilCoA

deshidrogenasa, insulina, beta-glucosidasa, piruvato carboxilato, fosforilasa hepática, cinasa fosforilasa, glicina descarboxilasa, H-proteína, T-proteína, una secuencia del regulador transmembranar de fibrosis quística (CFTR) y una secuencia de cADN distrofina.

Otros productos génicos incluyen polipéptidos de ocurrencia no natural, tales como polipéptidos quiméricos o polipéptidos híbridos que tienen una secuencia de aminoácidos de ocurrencia no natural que contienen inserciones, supresiones o substitución de aminoácidos. Por ejemplo, las inmunoglobulinas de cadena única hechas por ingeniería podrían ser útiles en ciertos pacientes comprometidos inmunológicamente. Otros tipos de secuencias génicas de ocurrencia no natural incluyen moléculas contrasentido y ácidos nucleicos catalíticos, tales como ribosimas, las cuales pueden ser usadas para reducir sobre-expresión de un blanco.

La reducción o modulación de la expresión de un gen son particularmente deseables para el tratamiento de condiciones hiperproliferativas caracterizadas por células hiperproliferantes, tales como en el cáncer y la soriasis. Los polipéptidos que actúan como blanco incluyen aquellos que son producidos exclusivamente o a los más altos niveles en células hiperproliferativas al compararlas con las células normales. Los antígenos blanco incluyen polipéptidos codificados por oncogenes tales como myb, myc, fyn y el gen de translocación bcr/abl, ras, src, P53, neu, trk y EGRF. En adición a los productos oncogénicos como antígenos

blanco, los polipéptidos blanco para tratamientos contra el cáncer y regimenes protectores incluyen regiones variables de anticuerpos hechos por las células B de linfoma y regiones variables de receptores de células T de células de linfoma las cuales, en alguna incorporaciones, también se usan como antígenos blanco para enfermedad autoinmune. Otros polipéptidos asociados a tumores se pueden utilizar como polipéptidos blanco tales como polipéptidos que se encuentran a niveles muy altos en células tumorales que incluyen el polipéptido reconocido por el anticuerpo 17-1A monoclonal y polipéptidos enlazantes de folato.

Otros polipéptidos terapéuticos y proteínas incluyen aquellos que pueden ser útiles para tratar individuos que sufren enfermedades autoinmunes y desórdenes conferidos por una repuesta inmune protectora amplia contra blancos que están asociados con autoinmunidad que incluyen receptores celulares y células que producen en si mismas anticuerpos. Las enfermedades autoinmunes mediadas por células T incluyen artritis reumatoide (RA), esclerosis múltiple (MS), síndrome de Sjögren's, sarcoidosis, diabetes isulinodependiente (IDDM), tiroiditis autoinmune, artritis reactiva, espondilitis anquilosante, escleroderma, polimiositis, dermatomiositis, soriasis, vasculitis, granulomatosis de Wegener, enfermedad de Crohn y colitis ulcerativa. Cada una de estas enfermedades se caracteriza por receptores de células T (TCRs) que unen los antígenos endógenos e inician la cascada inflamatoria asociada con enfermedades autoinmunes.

Los vectores adenovirales de simio de la invención son particularmente bien apropiados para regimenes terapéuticos en los cuales son deseable entregas adenovirales-mediadas múltiples de transgenes, e.g., en regimenes que implican re-entregas del mismo transgen o en regimenes de combinación que implican la entrega de otros transgenes. Tales regimenes pueden implicar la administración de un vector adenoviral de simio de Pan5, Pan6, Pan7, SV1, SV25 o SV39, seguida de re-administración con un vector del mismo serotipo de adenovirus. Regimenes particularmente deseables implican la administración de un vector adenoviral de simio de Pan5, Pan6, Pan7, SV1, SV25 o SV39 de la invención en el cual el serotipo del vector viral entregado en la primera administración difiere del serotipo del vector viral utilizado en una o más de las administraciones subsiguientes. Por ejemplo, un régimen terapéutico implica la administración de un vector de Pan5, Pan6, Pan7, SV1, SV25 o SV39 y repetir la administración con uno o más vectores adenovirales del mismo o diferente serotipo. En otro ejemplo, un régimen terapéutico implica la administración de un vector adenoviral seguida de administración repetida con un vector de Pan5, Pan6, Pan7, SV1, SV25 o SV39 de la invención el cual difiere del serotipo del vector adenoviral entregado primero, y además opcionalmente la administración con otro vector que es el mismo o, preferiblemente, difiere del serotipo del vector en las etapas de administración anteriores. Estos regimenes no están limitados para la entrega de vectores adenovirales construidos usando los serotipos de simio Pan5, Pan6, Pan7, SV1, SV25 o SV39 de la invención. Es decir, estos regimenes pueden fácilmente utilizar serotipos adenovirales de otros vectores, que incluyen, sin limitación, otros

serotipos adenovirales de simio (e.g., Pan9 o C68, Cl, etc), otros serotipos adenovirales de primate no humano, o serotipos adenovirales humanos, en combinación con uno o más de los vectores Pan5, Pan6, Pan7, SV1, SV25 o SV39 de la invención. Ejemplos de tales vectores de simio, otros serotipos adenovirales de primate no humano y humano se discuten en su lugar en este documento. Además, estos regimenes terapéuticos pueden implicar bien sea entregas simultáneas o secuenciales de vectores adenovirales Pan5, Pan6, Pan7, SV1, SV25 y/o SV39 de la invención en combinación con vectores no adenovirales, vectores no virales, y/o una variedad de otros compuestos o moléculas útiles terapéuticamente. La presente invención no esta limitada a aquellos regimenes terapéuticos, una variedad de los cuales será claro para alguien versado en el arte.

B. Transgenes Inmunogénicos de Entrega Ad-mediada.

Los adenovirus de simio recombinantes también se pueden emplear como composiciones inmunogénicas. Como se usa aquí, una composición inmonugénica es una composición en la cual una respuesta humoral (e.g., anticuerpo), o celular (e.g., una célula T citotóxica) se muta a un producto transgénico entregado por la composición inmunogénica siguiendo la entrega a un mamífero, y preferiblemente un primate. La presente invención proporciona un Ad de simio recombinante que puede contener en cualquier de sus supresiones de secuencia un gen que codifica un inmunógeno deseado. Estos adenovirus de

símio probablemente son mucho más apropiados para uso como una vacuna del virus recombinante vivo en diferentes especies animales comparado con un adenovirus de origen humano, pero no están limitados a tal uso. Los adenovirus recombinantes se pueden usar como vacunas profilácticas o terapéuticas contra cualquier patógeno para el cual el antígeno(s) es(son) crucial(es) para la inducción de una respuesta inmune y capaz de limitar el despliegue del patógeno que ha sido identificado y para el cual el cADN está disponible.

Tales composiciones de vacuna (u otro inmunogénico), se formulan en un vehículo de entrega adecuado, como se describió anteriormente. Generalmente, las dosis para las composiciones inmunogénicas están en los intervalos definidos anteriormente para composiciones terapéuticas. Los niveles de inmunidad del gen seleccionado se pueden monitorear para determinar la necesidad, si la hay, de aumentarlos. Siguiendo un ensayo de título de anticuerpos en el suero, pueden ser deseadas, inmunizaciones aumentadas opcionalmente.

Opcionalmente una composición para vacuna de la invención se puede formular para que contenga otros componentes, que incluyen, e.g., adyuvantes, estabilizantes, ajustadores de pH, preservativos y similares. Tales componentes son bien conocidos por aquellos versados en el arte de las vacunas. Ejemplos de adyuvantes adecuados incluyen, sin limitación, liposomas, alumbre, lípidos A monofosforilados, y cualquier factor activo biológicamente, tal como citoquina, una interleuquina, una quimioquina, un ligando, y combinaciones óptimas de los

mismos. Algunos de aquellos factores activos biológicamente se pueden expresar *in vivo*, e.g., vía un plásmido o un vector viral. Por ejemplo, tal adyuvante se puede administrar con una vacuna de ADN cebadora que codifica un antígeno para mejorar la respuesta inmune antígeno-específica comparada con la respuesta inmune generada por el cebador con una vacuna de ADN que codifica el antígeno únicamente.

Los adenovirus recombinantes se administran en una "cantidad inmunogénica", esto es, una cantidad de adenovirus recombinante que es efectiva en una ruta de administración para transfectar las células deseadas y proporcionar niveles suficientes de expresión del gen seleccionado para inducir una respuesta inmune. Cuando se proporciona la inmunidad protectora, se considera a los adenovirus recombinantes para estar en composiciones de vacuna útiles para prevenir la infección y/o la enfermedad recurrente.

Alternativamente, o en adición, los vectores de la invención pueden contener un transgen que codifica un péptido, polipéptido o proteína la cual induce una respuesta inmune para un inmunógeno seleccionado. Se espera que los adenovirus recombinantes de esta invención sean altamente eficaces en inducir las células T citolíticas y los anticuerpos para la proteína antigénica heteróloga insertada expresada por el vector.

Por ejemplo, los inmunógenos se pueden seleccionar de una variedad de familias virales. Ejemplos de familias virales deseables contra las cuales una respuesta inmune debería ser deseable incluyen , la familia picornavirus, la cual incluye los géneros rhinovirus, que son los responsables de cerca del 50% de los casos del resfriado común; los géneros enterovirus, el cual incluye polivirus, coxsackievirus, echovirus, y los enterovirus humanos tal como el virus de la hepatitis A; y los géneros apthovirus, los cuales son responsable de enfermedades en boca y pie, primariamente en animales no humanos. En la familia de virus picornavirus, los antígenos blanco incluyen los VP1, VP2, VP3, VP4 y VPG. Otra familia viral incluye la familia de los calcivirus, que abarca el grupo de virus Norwalk, los cuales son importantes porque son agentes que causan gastroenteritis epidémica. Aún otra familia viral deseable para ser usada como antígenos blanco para inducir respuestas inmunes en humanos y animales no humanos es la familia togavirus, que incluye los géneros alfavirus, la cual incluye los virus Sindbis, el virus RossRiver, y el Venezolano, el de la encefalitis equina del Oeste & dell Este, y rubivirus, incluyendo el virus de la rubeola. La familia flaviviridae incluye dengue, fiebre amarilla, encefalitis japonesa, Encefalitis St Louis y virus de la encefalitis llevadores de señal. Otro antígenos blanco se pueden generar de la familia de la Hepatitis C o de coronavirus, la cual incluye un número de virus no humanos tales como el de la bronquitis infecciosa (de aves de corral), virus de la gastroenteritis porcina transmisible (cerdo), virus de la encefalomielitis hemaglutinante porcina (cerdo), virus de la peritonitis infecciosa felina (gatos), corona virus entérica felina (gato), corona virus canina (perro) y corona virus respiratorios en humanos, los

cuales pueden causar el resfriado común y/o hepatitis no-A,B o C. En la familia de los corona virus, los antígenos blanco pueden incluir el E1 (también llamado M o proteína matriz), E2 (también llamado S proteína Spike), E3 (también llamado HE o hemaglutinina-elterosa) glicoproteína (no presente en todos los corona virus), o N (núcleo-cápside). Aún otros antígenos se pueden utilizar como blancos contra la familia rhabdovirus, la cual incluye los géneros vesiculovirus (e.g., Virus de la Estomatitis Vesicular),y el género general Lyssavirus (e.g., rabia) En la familia rhabdovirus antígenos adecuados se pueden derivar de la proteína G o de la proteína N. La familia filoviridae, la cual incluye el virus de la fiebre hemorrágica tales como los virus Marburg y Ebola, pueden ser fuentes adecuadas de antígenos. La familia paramyxovirus incluye los virus Tipo1 de la parainfluenza, virus Tipo3 de la parainfluenza, virus Tipo3 de la parainfluenza bovina, rubulavirus (virus de las paperas), virus Tipo2 de la parainfluenza, virus Tipo4 de la parainfluenza, virus de la enfermedad Newcastle (pollos), peste rinder, morbillivirus, el cual incluye la enfermedad de los cerdos producida por las triquinas y moquillo canino, y neumovirus, el cual incluye virus sincitial respiratorio, El virus de la influenza está clasificado en la familia ortomivirus y es una adecuada fuente de antígeno (e.g., La proteína HA, La proteína N1). La familia bunyavirus incluye los géneros bunyavirus (encefalitis de California, La Crosse), flebovirus (fiebre del Valle Rift), hantavirus (puremala es un virus de la fiebre hemahagin), nairovirus (enfermedad de las Ovejas de Nairobi) y varios bungavirus no asignados. La familia arenavirus proporciona una fuente de antígenos contra LCM y virus de la fiebre Lassa y LCM. La familia reovirus incluye

los géneros reovirus, rotavirus (el cual causa gastroenteritis aguda en niños), orbivirus, y cultivirus (Fiebre Tick de Colorado, Lebombo (humanos), encefalosis equina, lengua azul).

La familia de los retrovirus incluye la subfamilia oncorivirinal la cual abarca enfermedades tanto humanas como veterinarias como el virus de la leucemia felina, HTLVI y HTLVII, lentivirinal (el cual incluye el virus de inmunodeficiencia humana (HIV), virus de inmunodeficiencia de los simios (SIV), virus de inmunodeficiencia felina (FIV), virus de la anemia infecciosa equina, y espuma virinal). Entre los lentivirus, algunos antígenos adecuados se han descrito y se pueden seleccionar fácilmente. Ejemplos de antígenos HIV y SIV incluyen, sin limitación, las proteínas gag, pol, Vif, Vpx, VPR, Env, Tat, Nef y Rev, así como también varios fragmentos de los mismos. Por ejemplo, fragmentos adecuados de la proteína Env pueden incluir cualquiera de sus subunidades tales como la gp 120, gp160, gp41, o fragmentos más pequeños de los mismos. E.g., de al menos cerca de 8 aminoácidos de longitud. Similarmente, se pueden seleccionar fragmentos de la proteína tat. [Ver, Patente US 5, 891,994 y Patente US 6,193, 981.]. Ver, también, las proteínas HIV y SIV descritas en D.H. Barouch et al, J. Virol., 75(5):2462-2467 (Marzo 2001), y R.R. Amara, et al, *Science,* **292**:69-74 (6 de Abril 2001). En otro ejemplo, las proteínas inmunogénicas HIV y/o SIV o péptidos se pueden usar para formar proteínas de fusión u otras moléculas inmunogénicas. Ver, e.g., las proteínas de fusión tat HIV-1 y/o Nef y regímenes de inmunización descritos en WO 01/54719, publicada en Agosto 2, 2001, y WO

99/16884, publicada Abril 8, 1999. La invención no está limitada a las proteínas inmunogénicas HIV y/o SIV o a los péptidos aquí descritos. En adición, una variedad de modificaciones de estas proteínas han sido descritas o podrían ser hechas fácilmente por alguien versado en el arte. Ver, e.g., la proteína gag modificada que esta descrita en la Patente US 5, 972,596. Además, cualquiera de los inmunogenos deseados HIV y/o SIV puede entregarse solos o en combinación. Tales combinaciones pueden incluir la expresión de un vector único o de vectores múltiples. Opcionalmente, otra combinación puede involucrar la liberación de uno o mas inmunogenos expresados con la liberación de una o mas de los inmunote nos en forma de proteína. Tales combinaciones se discuten con más detalle, mas adelante.

La familia papovavirus incluye la sub–familia poliomavirus (virus BKU y JCV) y la familia papillomavirus (asociada con cánceres o progresión maligna de papiloma). La familia adenovirus incluye virus (EX, AD7, ADR, O.B.) los cuales causan enfermedad respiratoria y/o enteritis. La familia parvovirusfelino (enteritis felina), palenco–pemiavirus felino, parvovirus canino, y parvovirus porcino. La familia herpesvirus incluye la sub-familia alfaherpesvirinae la cual abarca los géneros simplex virus (HIV, HSVII), varicellovirus (seudorabies, varicela zoster) y la subfamilia betaherpesvirinae, la cual incluye los géneros citomegalovirus (HCMV, muromegalovirus) y la subfamilia gammaherpesvirinae, la cual incluye los géneros linfocriptovirus, EBV (Linforma de BurKitts), rinotraqueitis infecciosa, virus de la enfermedad de Marek y rhadinovirus. La familia poxvirus incluye la subfamila

chordopoxvirinae, la cual abarca los géneros orthopoxvirus (varicela) y vaccina (coupox), parapoxvirus, avipoxvirus, capripoxvirus, leporipoxvirus, suipoxvirus, y la subfamilia entomopoxvirinae. La familia hepadnavirus incluye el virus de la Hepatitis B. Un virus no clasificado que puede ser una fuerte adecuada de antigenos es el virus de la hepatitis delta. Aún otras fuentes virales pueden incluir virus de la enfermedad bursal infecciosa en aves y respiratoria porcina y virus del síndrome reproductivo. La familia alfavirus incluye virus de la artritis equina y varis virus de encefalitis.

La presente invención puede abarcar inmunógenos que son útiles para inmunizar humanos y animales no humanos, contra otros patógenos incluyendo bacterias, hongos, microorganismos parasiticos o parásitos multicelulares los cuales infectan vertebrados humanos y no humanos, o de célula cancerosa o célula tumoral. Ejemplos de patógenos bacteriales incluyen cocos gram–positivos patógenos, incluyen neumococos, estafilococos; y estreptococos. Cocos patógenos gram–negativos incluyen meningococos, gonococos, bacilos, gram–negativos entéricos patógenos incluyen enterobacteriaceae: pseudomonas, acinetobacterias y elkenella; melioidiosis, salmonela, shíguela, ahemophilus, moraxelle; *H. ducreyi* (la cual causa chancroides); brucela, *Franicella tularensis* (la cual causa tularemia, yersinia (pasteurlla); estreptobacillus moniliformes y spirillium; bacilos gram–positivos que incluyen listeria monocitogenes; erysipelothrix rhusiopathiae; *Coryneabacterium diphteria* (difteria); cólera; *B.anthracis* (ántrax); donovanosis (granuloma inguinal); y bartonellosis. Enfermedades causadas por bacterias

patógenas anaeróbicas incluyen tétanos; botulismo, otros clostridios, tuberculosis, lepra; y otras micobacterias.

Las enfermedades de la espiroqueta patógena incluyen sífilis, treponematosis.; pinta y sífilis endémica; y leptospirosis. Otras infecciones causadas por bacterias patógenas más altas y hongos patógenos incluyen actinomicosus, norcadiosis, criptococosis, blastomicosis, histoplasmosis y coccidioidomicosis; candidiasis, aspergillosis y mucormicosis; esporotricosis; paracoccidiodomicosis, petrielidiosis, torulopsosis, micetoma y cromomicosis y dermatofitosis. Infecciones ricketsiales incluyendo fiebre tifoidea, fiebre de las Montañas Rocosas, fiebre Q y ricketsialpox. Ejemplos de infecciones de micoplasma y clamidiales incluyen: neumonía micoplasma; linfogranuloma venéreo, sitacosis, infecciones perinatales clamidiales. Los eucariotes patogénicos, abarcan protozoos patogénicos y helmintos y las infecciones producidos por estos incluyen: amebiasis; malaria; leishmaniasis; tripanosomiasis; toxoplasmosis; *pneumocystis carinii; trichans; toxoplasma gondii*; babesiosis, giardisis, triquinosis, filariasis, esquistosomiasis, nematodos, trematodos; e infecciones cestodo (gusano).

Muchos de estos organismos y/o toxinas producidas han sido identificadas por los Centros de Control de Enfermedades, (CDC), Departament Of Health And Human Services, USA), como agentes que tienen un uso potencial en ataques biológicos. Por ejemplo, algunos de estos agentes biológicos, incluyen, *Bacillus anthracis* (ántrax), Clostridium *botulinum* y sus toxinas (botulismo), *Yersinia pestis* (plaga)

viruela mayor(smallpox), *Francisella tularensis* (Tularemia) y fiebres virales hemorrágicas [filovirus (e.g., Ebola, Marburg], y arenavirus [e.g., Lassa, Machupo], todas están clasificadas corrientemente como agentes de categoría A; *Coxiella burnetti* (Fiebre Q); especies de Brucella (brucelosis), *Burkholderia mallei* (muermo), *Burkholderia pseudomalles*; (meloidosis), Ricinus communis y sus toxinas (toxina de ricino), *Clostridium perfringenes* y su toxina (toxina epsilón), Especies de *Staphylococcus* y sus toxinas (enterotoxina B), *Clamydia psittaci* (sitacosis), seguridad del agua (e.g., *Vibrio cholerae*, *Crytosporidium parvum*), fiebre tyfoidea (*Rickettsia powasekii*), y encefalitis viral (alfavirus, e.g., encefalitis equina venezolana, encefalitis equina del este, encefalitis equina del oeste); todas las cuales están clasificadas como agentes de categoría B; y los virus Nipan, hantavirus, los cuales están clasificados en la categoría C. En adición, otros organismos, los cuales están clasificados así, o clasificados de manera diferente, se pueden identificar y/o usan para tal propósito en el futuro. Se deberá entender fácilmente que los vectores virales y otros constructos descritos aquí son útiles para entregar antígenos desde estos organismos, virus, sus toxinas u otros subproductos, los cuales prevendrán y/o trataran la infección u otras reacciones adversas con estos agentes biológicos.

La administración de los vectores de la invención para entregarar inmunógenos contra la región variable de las células T para producir una respuesta inmune incluyendo CTLs para eliminar aquellas células T. En RA, algunas regiones variables características de los TCRs los cuales están involucrados en la

enfermedad han sido caracterizadas. Estos TRCs incluyen V-3, V-14, V-17 y Vα-17. Entonces la, entrega de un ácido nucleico que codifique por lo menos uno de estos polipéptidos producirá una respuesta inmune que dará en el blanco en las células T involucradas en RA. En MS, algunas regiones variables específicas de TCRs las cuales están involucradas en la enfermedad, se han caracterizado. Estos TCRs incluyen V-7 Y Vα10. Entonces, la entrega de una secuencia de ácido nucleico que codifica al menos uno de estos polipéptidos producirá una respuesta inmune que dará en el blanco de las células T involucradas en MS. En escleroderma, algunas regiones variables específicas de TRCs las cuales están involucradas en la enfermedad han sido caracterizadas. Estos TCRs incluyen V-6, V-8, V-14 y V -16, V -3C, V -7, V -14, V -15, V -16, V -28 Y V -12. Entonces la entrega de adenovirus recombinante de simio que codifica por lo menos uno de estos polipéptidos que producirán una respuesta inmune que dará en el blanco de las células T involucradas en la escleroderma.

C. Métodos de liberación Ad-mediados.

Los niveles terapéuticos, o niveles de inmunidad, de los genes seleccionados se pueden monitorear para determinar la necesidad, si hay alguna, para los suplementos, siguiendo el impacto de la respuesta de las células T CD8+ de opcionalmente, títulos de anticuerpos, en el suero, se pueden desear inmunizaciones suplementarias.

Opcionalmente, los vectores adenovirales recombinantes de simio de la invención, se pueden entregar en una administración simple o en varios regímenes combinados e. g., en combinación con un régimen o curso de tratamiento que involucra otros ingredientes activos o en un régimen de inicio-suplemento. Una variedad de tales regímenes, se han descrito en el arte y se pueden seleccionar fácilmente. Por ejemplo, los regímenes de inicio y suplemento, pueden involucrar la administración de un vector basado en un ADN (e. g., plasmido) basado en un vector para iniciar el sistema inmune al segundo, suplementar, la administración con un antígeno tradicional, tal como una proteína o un virus recombinante que lleve las secuencias que codifican tal antígeno. Ver e. g. No.WO 00/11140, publicado Marzo 2, 2000, incorporado por referencia. Alternativamente, un régimen de inmunización puede envolver la administración de un vector adenoviral recombinante de simio de la invención suplementar la respuesta inmune a un vector (ya sea basado en ADN o viral) llevando un antígeno, o una proteína. En aún, otra alternativa, un régimen de inmunización involucra la administración de una proteína seguida por un suplemento con un vector que codifica el antígeno.

En otra incorporación, la invención proporciona un método de iniciación y suplementación de una respuesta inmune a un antígeno seleccionado por entrega de un vector plásmido de ADN que lleva dicho antígeno, seguido por la suplementacióno con un vector adenoviral recombinante de simio de la invención. En otra incorporación, el régimen inicio-suplemento, involucra la expresión de multiproteínas del inicio, y/o el vehículo suplementador. Ver, e. g., R. R. Amara,

Science **292**: 69-74 (6 de Abril de 2001) el cual describe un régimen multiproteínico para la expresión de unidades de proteína para generar una respuesta inmune contra HIV y SIV. Por ejemplo, un ADN iniciador puede ser entregado a Gag., Pol., Vif., VPX y Vpr y Env. Tat., y Rev de una trascripción simple. Alternativamente, el Gag SIV, Pol y HIV-1 Env se entrega en un constructo de adenovirus recombinante de la invención. Aún otros regímenes se describen en WO 99/16884 y WO 01/54719.

Sin embargo, los regímenes iniciación-suplemento, no se limitan a la inmunización para HIV o para entrega de estos antígenos. Por ejemplo, una iniciación puede involucrar la entrega con un primer vector de chimpancé de la invención seguida por la suplementación con un segundo vector de chimpancé, o con una composición que contiene el antígeno por sí mismo en forma de proteína. En un ejemplo, el régimen iniciación–suplemento proporciona una respuesta inmune protectora contra el virus, bacteria u otro organismo del cual se deriva el antígeno. En otra incorporación deseada, el régimen inicio–suplemento proporciona un efecto terapéutico que se puede medir usando ensayos convencionales para la detención de la presencia de la condición para la cual se administra la terapia.

La composición iniciadora se puede administrar en varios sitios en el cuerpo en dosis que dependen de la manera, la cual depende del antígeno que se coloca como blanco para la respuesta inmune deseada. La invención no está limitada a la cantidad o sitios de inyección(es) o al transportador farmacéutico. Además, el

92
93

régimen puede involucrar pasos de iniciación y/o paso de suplementación, cada uno de los cuales pueden incluir una dosis sencilla o una dosis que se administra horariamente, diariamente, semanalmente, o mensualmente, o anualmente. Como un ejemplo, los mamíferos pueden recibir una dosis conteniendo entre cerca de 10 g a cerca de 50 g de plásmido en el transportador. Una cantidad deseable de rangos de composición de ADN está entre cerca de 1 g hasta cerca de 10,000 g del vector ADN. Las dosis pueden varias desde cerca de 1 g hasta 1000 g por kilo de peso del cuerpo del sujeto. La cantidad o sitio de la entrega se selecciona deseablemente basado en la identidad y la condición del mamífero.

La unidad dosis del vector apropiado para entrega del antígeno al mamífero, se describe aquí. El vector se prepara para administración suspendiendo o disolviendo en un transportador aceptable farmacéuticamente o fisiológicamente tal como una solución salina isotónica; las soluciones salinas isotónicas y otras formulaciones que aparentemente son conocidas en tal administración. El transportador apropiado será evidente para aquellos versados en el arte y dependerá en gran parte de la ruta de administración. Las composiciones de la invención se pueden administrar a los mamíferos de acuerdo con las rutas descritas antes, en una formulación de entrega sostenida utilizando un polímero biocompatible, biodegradable, o por entrega en el sitio utilizando micelas, geles y liposoma. Opcionalmente el paso de iniciación de esta invención, también influye

la administración con la composición iniciante, una cantidad deseable de un adyuvante, tal como se definió aquí.

Preferiblemente una composición de suplementación, se administra cerca de 2 hasta cerca de 7 semanas después de la administración de la composición iniciante para los sujetos mamíferos. La administración de la composición de suplementación, es acompañada usando una cantidad efectiva de una composición de suplementación, conteniendo o que sea capaz de entregar el mismo antígeno como se administró por la vacuna del ADN inicial. La composición de suplementación se puede componer de un vector viral recombinante derivado de la misma fuente viral (e. g., secuencias adenovirales de la invención) o de otra fuente. Alternativamente, la "composición de la suplementación" puede ser una composición conteniendo el mismo antígeno como se codificó en la vacuna de ADN inicial, pero en la forma de una proteína o de un péptido, cuya composición induce una respuesta inmune en el huésped. En otra incorporación, la composición de suplementación contiene una secuencia de ADN que codifica el antígeno bajo el control de una secuencia regulatoria dirigiendo la expresión en la célula del mamífero, e. g., vectores tales como los bien conocidos vectores virales o bacteriales. Los requerimientos primarios de la composición de suplementación, son esos en los cuales la composición del antígeno es el mismo antígeno, o una sección reactiva del antígeno como la codificada por la composición inicial.

En otra incorporación, los vectores adenovirales de simio de la invención, son también apropiados para usarlos en una variedad de regímenes de otra inmunización y terapéuticos. Tales regímenes pueden involucrar lla entrega de vectores adenovirales de simio de la invención simultánea o secuencialmente con vectores Ad de diferentes cápsides de serotipo, regímenes en los cuales los vectores adenovirales de la invención son entregados simultáneamente o secuencialmente con vectores no-Ad, regímenes en los cuales los vectores adenovirales de la invención son entregados simultáneamente o secuencialmente con proteínas, péptidos, y/o otros compuestos terapéutica o inmunogénicos útiles biológicamente. Tales usos son fácilmente claros para una persona versada en el arte.

Los siguientes ejemplos ilustran la clonación de los adenovirus de simio y la construcción de vectores de adenovirus recombinantes ejemplarmente de la presente invención. Estos ejemplos son solamente ilustrativos, y no limitan el alcance la presente invención.

Ejemplo 1 – Propagación Viral.

Los virus, Pan5 [ATTC Acceso No. VR-591], Pan6 [ATTC Acceso No. VR-592] y Pan7 [ATTC Acceso No. VR-593] originalmente aislados de nodos linfáticos de chimpancés, fueron propagados en células 293 [ATTC CRL 1573]. Típicamente estas células son cultivadas en un medio Dulbecco's Modified Eagles Mediurm

(DMEM; Sigma, St. Louis, MO). Suplementado con suero fetal calf al 10% (FCS) [Sigma o Hyclone, Logan, UT] y Penicilina–Estreptomicina al 1% (Sigma). La infección de 293 células, se lleva a cabo fuera en un de DMEM suplementado con 2% de FCS por las primeras 24 horas, después de lo cual se adiciona FCS para llevar la concentración final al 10%. Las células infectadas, son cosechadas cuando el 100 por ciento de las células presentan efectos citopáticos inducidos por virus (CPE), entonces se colectan, y se concentran por centrifugación. Los pelets de células son resuspendidos en 10 nM de Tris (pH 8.0, y se les hace lisis por tres ciclos de congelación y descongelación. Las preparaciones de virus se obtienen siguiendo dos pasos de ultracentrifugación sobre gradientes de densidad de cloruro de cesio y los patrones de virus se diluyen de 1 a 5×10^{12} partículas/ml en 10 mM de Tris / 100 mM NaCl/50% de glicerol y se almacenan a -70°C. La habilidad de las 293 células para propagar estos adenovirus, excedió las expectativas las cuales estaban basadas en el conocimiento de otros serotipos de adenovirus no humanos.

Virus	Producción (partículas de virus producidas en $8X10^8$ Células)
Pan5	8.8×10^{13}
Pan6	1.6×10^{14}
Pan7	8.8×10^{13}

Ejemplo 2 – Caracterización del ADN Viral Genómico.

El ADN genómico fue aislado de preparaciones de virus purificadas el Ejemplo 1 y fueron digeridas con enzimas de restricción HindIII o BamHI siguiendo las recomendaciones de los fabricantes. Los resultados (no se muestran) revelaron que los genomas del Pan5, Pan6, Pan7 de la invención y el publicado genoma Pan9 (C68) mostró diferentes patrones de restricción y entonces, son distintivos de cada uno.

Las secuencias de nucleótidos de Pan5, Pan6 y Pan7 fueron determinadas. La secuencia de nucleótido de la parte alta de la banda del ADN de Pan5 se reporta en SEQ ID NO: I. La secuencia de nucleótido de la parte alta de la banda del ADN de Pan6 se reporta en SEQ ID NO: 5. La secuencia de nucleótido de la parte alta de la banda del ADN de Pan7 se reporta en SEQ ID NO: 9.

Las regiones reguladas y codificadas en las secuencias de ADN viral, fueron identificadas por homología con secuencias adenovirales conocidas utilizando el programa "Clustan W" descrito antes en las condiciones convencionales. Ver las tablas anteriores que proporcionan las secuencias adenovirales. Las estructuras de lectura abierta, fueron traducidas y las secuencias de los aminoácidos predichas se examinaron para la homología para las secuencias de proteínas adenovirales previamente descritas, Ad4, Ad5, Ad7, Ad12 y Ad40.

97
98

Los análisis de la secuencia revelaron una organización del genoma que es similar a la presentada en los adenovirus humanos, con la mayor similitud a los humanos Ad4. Sin embargo, las diferencias sustanciales en las regiones hipervariables del exón fueron notadas entre los adenovirus del chimpancé y otros adenovirus conocidos, incluyendo AdHu4. Estas diferencias bien ajustadas con los datos de reactividad cruzada serológica que se han obtenido (ver abajo). Una alineación de una porción de las secuencias del exón se muestra en la Fig. 1. la porción mostrada es desde la región del exón que corresponde a la dispuesta exteriormente extendida a los bucles DEI y FG1 donde se observó la mayor variabilidad entre los serotipos.

Una porción que interviene y que contribuye a la base de exón (correspondientes a los residuos 308-368 de la secuencia AdC68 de la publicación; patente US 6.083716), y es altamente conservada entre los serotipos, está también presente. La siguiente tabla resume las comparaciones por pares adecuados de los aminoácidos en las proteínas de exón.

Comparación		Similaridad de los aminoácidos de laexón (%)
# 1	#2	
AdC5	AdC7	99.0
AdC5	AdC68	98.3
AdC5	AdC6	88.0
AdC5	AdC1	84.9
AdC6	AdC7	87.7
AdC6	AdC68	87.3
AdC6	AdC1	84.9
AdC7	AdC68	97.5
AdC7	AdC1	84.8
AdC68	AdC1	84.9

Los análisis del dominio de la fibra de nudo (la cual es responsable de la unión al receptor) de los adenovirus del chimpancé muestran total similaridad en la estructura (Fig. 2).

En grados de similaridad de la secuencia entre las proteínas E1 de huAd5 y C68. (Ver las tablas más adelante), es similar a las que hay entre huAd5 y Pan-5, Pan-6 y Pan-7.

Comparación		Porcentaje de identidad de los aminoácidos en E1a (13S)
# 1	#2	
AdHu5	AdC5	36.6
AdHu5	AdC6	28.5
AdHu5	AdC7	34.9
AdHu5	AdC68	35.6
AdHu5	AdC1	35.6
AdC5	AdC6	68.3
AdC5	AdC7	96.9
AdC5	AdC68	80.4
AdC5	AdC1	51.3
AdC6	AdC7	69.3
AdC6	AdC68	59.4
AdC6	AdC1	37.7
AdC7	AdC68	81.5
AdC7	AdC1	51.0
AdC68	AdC1	54.9

	Identidad de la secuencia con Ad5 humano	
	Proteína T pequeña E1b	Proteína T grande E1b
C68	47.3%	55.8%
Pan-5	43.2%	54.5%
Pan-6	45.3%	54.5%
Pan-7	46.4%	53.8%

Las versiones de replicaciones defectuosas de AdC5, AdC6 y AdC7 fueron creados por métodos de clonación molecular descritos en los siguientes ejemplos en los cuales los casetes de minigenes fueron insertados en el lugar de los genes de E1a y E1b. Los clones moleculares de los virus recombinantes fueron rescatados y agrupados en 293 células para una purificación en gran escala, utilizando el método de sedimentación en CsCl publicado [K. Fisher et *al.*, J. Virol., **70**:520 (1996)]. La producción del vector se basó en 50 placas preparadas (150mm) en las cuales aproximadamente 1 x 10^9 y 293 células fueron infectados por los correspondientes virus. Los producidos se determinaron midiendo espectrofotométricamente la concentración de particular virales. Después teniendo los vectores E1- suprimidos construidos, se determinaron las 293 células HEK (las cuales expresaban funciones E1 del serotipo 5 de adenovirus humano)

transcomplementarios de las E1 suprimidos de los vectores virales nuevos y permitir la producción de patrones de alto título. Ejemplos de producción de virus para unos pocos de estos virus recombinantes se muestran en la tabla abajo.

Los transgenes de estos vectores, β - galactosidasa (LacZ), la proteína fluorescente (GFP), alfa – 1- anti-tripsina (A1AT), glicoproteína ébola (ebo), una variante de glicoproteína soluble carente de transmembrana y dominio citoplasmáticos (sEbo), los tres mutantes suprimidos de la glicoproteína ébola (Ebo 2, Ebo 3, y Ebo 4), fueron expresados por el promotor citomegalovirus (CMV). En la siguiente tabla ND indica que el estudio no se ha hecho aún.

Transgen	Columna viral / producción del vector (partículas virales X 10^{13})			
	AdHu5	AdC7	AdC68	AdC6
CMVLacZ	1.5	1.4	3.3	6.1
COMVGFP	2.5	3.6	8	10
CMVA1AT	3.7	6	10	ND
CMVEbo	1.1	4.3	ND	ND
CMVsEbo	4.9	5.4	ND	ND
CMVEbo 2	1	9.3	ND	ND
CMVEbo 3	0.8	9.5	ND	ND
CMVEbo 4	1.4	6.2	ND	ND

La habilidad del adenovirus humanos E1 para transcomplementar los virus E1-suprimidos de chimpancé de la invención es altamente ventajosa, como permite la producción de vectores adenovirales E1-suprimidos de chimpancé de la invención, mientras reduce o elimina el riesgo de recombinación homóloga debido a las diferencias en las secuencias entre adenovirus humano Ad y de chimpancé descrito aquí.

Ejemplo 3 – Estudios serológicos de virus Pan-5, 6 y 7.

A causa de las diferencias en la región hipervariable en el exón, se anticipó que los virus de C5, C6 y C7, deberían ser serológicamente diferentes de los adenovirus humanos incluyendo AdHu4.

1. Reactividad cruzada de virus tipo salvaje.

Para tamizar los virus tipo salvaje con el objeto de hacer una determinación de la reactividad cruzada del anticuerpo, los virus de replicación competente, se utilizaron y se midió la inhibición de los efectos citopáticos (CPE). Brevemente, las preparaciones de adenovirus (Adhu5, Pan-5, Pan-6, Pan-7 y AdC68) almacenados en 5 X 10^{12} partículas/ml se diluyeron 1/600 para los ensayos. Esta concentración de virus se seleccionó en los resultados en 100% de CPE en 48 horas en la ausencia de neutralización. Previa adición del virus a las 293 células (4 X 10^{4} células/pozo en una bandeja de 96 pozos), 1:20 diluciones de sueros se

adicionaron, el ensayo se leyó tanto en la presencia como en la ausencia de CPE; la completa neutralización debería leerse como que no tiene efectos citopáticos. Los resultados se resumen en la tabla más abajo. El hecho de que 9/36 sera humana neutralizada Adhu5 induce CPE es consisten con los estimativos previos de anticuerpos neutralizantes en la población humana. Los números indican los individuos totales que mostraron neutralización (numerador) vs.el total de números tamizados (denominador). ND = no determinado.

	Neutralización por 1/20 dilución de suero		
	Humano /N=36)	Rhesus (N=52)	Chimpancé (N=2.0)
Adhu5	9/36	ND	ND
AdC68	1/36	0/52	12/20
Pan 5	0/36	0/52	10/20
Pan 6	0/36	0/52	9/20
Pan 7	0/36	0/52	12/20

De todos los sueros humanos tamizados, los 35/36 fueron negativos para la neutralización de AdC68 mientras que los 36/36 fueron negativos para neutralización de Pan-5, Pan-6 y Pan-7. De los 52 monos Rhesus tamizados ninguno mostró neutralización de cualquier adenovirus de chimpancé; el mono

Rhesus es el modelo preclínico preferido para la evaluación de vacunas de HIV. Entre 9 a 12 de 20 chimpancés tuvieron sustancial neutralización de uno u otro de los adenovirus de chimpancé consistente con el hecho de que estos son verdaderamente patógenos endémicos y específicos para chimpancé. Interesantemente hay chimpancés con anticuerpos neutralizantes solamente a Pan-5, Pan-6 ó AdC68 que dan soporte a la hipótesis que algunos de estos vectores adenovirales de chimpancé, no tendrán neutralización cruzada mutuamente y son distintos serotipos distintos.

El mismo ensayo se llevó a cabo para 20 muestras de suero de chimpancé. El cincuenta por ciento (50%) de las muestras, fueron reactivas serológicamente, en diferentes grados a Pan-5; 40% a Pan-6; 55% a Pan-7 y 60% a C68. Entre las muestras de suero positivo, una de ellas tuvo una actividad neutralizantemente fuerte para 4 virus de chimpancé.

2. Neutralización cruzada con virus recombinantes.

Altos títulos de anticuerpos policlonales, se obtuvieron de cada uno de los adenovirus de simio con el objeto de medir precisamente el grado de neutralización cruzada entre los diferentes serotipos. Esto se hizo por inmunización intramuscular de conejos utilizando virus recombinante conteniendo GFP basado sobre los adenovirus C68 de chimpancé previamente descritos, como un adyuvante. El suero fue utilizado para ensayar la actividad neutralizante contra

cada uno de los 3 adenovirus de chimpancé de la invención, AdC5, AdC6 y AdC7. Un conejo fue inyectado intramuscularmente con 5 por 10^{12} partículas virales del vector C68CMVGFP por kilo y suplementados 5 semanas más tarde usando la misma dosis. Un sangrado colectado a las 9 semanas reveló un punto extremadamente potente de actividad neutralizante contra C68 así como también contra Pan-5 y Pan-7 pero no contra Pan-6 (ver. la tabla más adelante), que indica que la administración de la vacuna basada en C68 (Pan-5 – Pan-7) podría ser efectivamente seguida por un suplemento usando un vector basado en Pan-6. Sin embargo, se ha encontrado que este nivel de interrelación no necesariamente se previene con la readministración en una condición donde los títulos de anticuerpo antiviral no fueron tan altos como los alcanzados en este conejo. En la siguiente tabla + indica 33% de CPE; ++ indica 66% de CPE; +++ indica 100% de CPE.

106
107

Infección con virus sobre 293 células					
Pan 5	Pan 6	Pan 7	Pan 9 (C68)	C68 GFP	Dilución del suero
-	+++	-	-	-	1/20
-	+++	-	-	-	1/40
-	+++	-	-	-	1/80
-	+++	-	-	-	1/160
-	+++	-	-	-	1/320
-	+++	-	-	-	1/640
-	+++	-	-	-	1/1.280
-	+++	-	-	-	1/2.560
-	+++	-	-	-	1/5.120
+	+++	-	-	-	1/10.240
+	+++	++	-	-	1/20.480
++	+++	+++	-	-	1/40.960
++	+++	+++	+	+	1/81.920
+++	+++	+++	++	++	1/163.840
+++	+++	+++	+++	+++	1/327.680
+++	+++	+++	+++	+++	1/665.360
+++	+++	+++	+++	+++	1/1.310.720
+++	+++	+++	+++	+++	1/2.621.440

3. *Ensayo Cuantitativo para Detección del Anticuerpo Neutralizante*

El resultado fue validado por un ensayo basado más cuantitativamente para detectar anticuerpo neutralizante, el cual se basa en la transducción de un vector

GFP. Brevemente, grupos de ratones C57BL/6 se inmunizaron intramuscular o endovenosamente con 5.0 x 10^{10} partículas/mL Pan-5, Pan-6, Pan-7 o C68. Los sueros de 28 días de sangrado se ensayaron para la actividad neutralizante cruzada contra C68CMVEGFP en diluciones de 1/20 y 1/80. En resumen, cuando una preparación farmacéutica de inmunoglobulina humana se ensayo para reacciones serológicas a Pan-5, 6 y 7 y C68, se detectaron algunos de niveles bajos de actividades neutralizantes contra Pan-7 y C68. No se detectó actividad neutralizante contra Pan-5 o Pan-6. Muestras de suero de 36 sujetos humanos se corrieron para el mismo ensayo. Las muestras de suero fueron ensayadas en un dilución 1/20. Los resultados indicaron que solamente un individuo tiene clara actividad neutralizante a C68. No se detectó actividad neutralizante a Pan-5, Pan-6 o Pan-7.

4. *Neutralización cruzada in vitro*

Se ensayo la neutralización cruzada de los adenovirus de simio por los altos títulos de anticuerpos policlonales de conejo elevados contra cada uno de los adenovirus Pan-5, Pan-6, Pan-7 y C68.

Los conejos fueron inmunizados con inyecciones intramusculares de 10^{13} partículas de cada uno de los adenovirus de chimpancés y suplementados, 40 días más tarde con la misma dosis con un adyuvante de Freund incompleto. Los sueros fueron analizados o ensayados para la presencia de anticuerpos

neutralizantes por incubación de un serial de diluciones por duplicado con 10^9 copias de genomas de cada uno de los vectores de adenovirus de chimpancé apropiados que expresan GFP y ensayando la expresión de GFP para la atenuación cuando se aplicó a 293 células. La dilución del suero que produjo una reducción del 50% de la expresión GFP se midió como el título de anticuerpo neutralizante contra ese virus en particular.

Los resultados se muestran en la tabla. Los datos son consistentes con lo esperado del análisis de la secuencia de las secuencias de aminoácidos de la exón, lo cual indicó que Ad Pan-6 era probablemente el más diferente serológicamente al ser comparado con los otros adenovirus de chimpancé.

	Infección de células 293 con 10^9 copias de genoma			
Suero de Conejo Inmunizado con:	Ad Pan-5	Ad Pan-6	Ad Pan-7	Ad C68
Ad Pan-5	1/5120	<1/20	1/2560	1/2560
Ad Pan-6	No Neutralización	1/20.480	<1/20	<1/20
Ad Pan-7	1/2560	1/160	1/163.840	1/2560
Ad C68	No Neutralización	<1/20	<1/20	1/5120

En orden a determinar si la reacción cruzada de anticuerpos con los adenovirus de simio era posiblemente de baja prevalencia en humanos los adenovirus de simio SV1, SV39 y SV25, se ensayaron para su habilidad de mantener la neutralización cuando se incubaron con un pool de inmunoglobulinas (Ig) disponibles

comercialmente. El mismo ensayo también fue realizado con Adhu5 y los adenovirus de chimpancé Pan-5, Pan-6, Pan-7 y C68. En un estudio adicional, los sueros de ratón se inmunizaron con uno de los adenovirus de chimpancé C5, C6, C7 y C68 y han sido ensayadas sus habilidades para la neutralización cruzada de los adenovirus de simio SV-15, SV23, SA-17 y el adenovirus de mandril. En ninguno se observó reactividad cruzada.

Ejemplo 4- Generación Del Vector Pan5 E1 Suprimido Recombinante.

Un plásmido pX se preparó mediante la destrucción del sitio FspI en la región bla del gen de pX (Clontech) por mutagénesis sitio-dirigida. El plásmido modificado resultante, llamado pX`, es un plásmido circular de 3.000 bp el cual contiene un f1 ori y un gen resistente a la ampicilina (AmpR-cds).

A. *Producción del Plásmido de Adenovirus de Pan-5.*

Se creó un poli-ligante para clonación secuencial de fragmentos de ADN de Pan-5 en pX`. El poli-ligante se sustituye por el poli-ligante existente en pX` siguiendo la digestión *MluI y EcoRI.* El fragmento *FseI* romo de Pan-5 se inserta en los sitios *SmaI* y *FseI* del poli-ligante. Este fragmento contiene el extremo 5` del genoma adenoviral (bp1 a 3606, SEQ ID NO: 1). El fragmento *SnaBI-FspI* de Pan-5 (bp 455 a 3484, SEQ ID NO: 1) se reemplaza con una secuencia corta flanqueada por los sitios *I-CeuI y PI-SceI* de lanzadera p (pShuttle) (Clontech), para eliminar la

región del genoma adenoviral. El fragmento EcoRI-romo de Pan-5 (bp 28658 a 36462, SEQ ID NO: 1) se inserta en los sitios *EcoRI y EcoRV* del poli-ligante (para proporcionar el extremo 3` del genoma adenorival); el fragmento *FseI-MluI* (bp 3606 a 15135, SEQ ID NO: 1) se inserta en el poli-ligante; y el fragmento *MluI – EcoRI* se inserta en el poli-ligante (bp 15135 a 28658, SEQ ID NO: 1). Opcionalmente, un transgen deseado en los sitios *I-CeuI y PI-SceI* del vector pX`Pan-5 E1 recientemente creado.

B. *Método alternativo de generación pX`Pan-5 E1.*

El plásmido inicial pX se deriva del plásmido adenovirus pAdX disponible de Clontech como se describió anteriormente. Después, se suprimió una región *PacI-XhoI* de pX` y el poli-ligante Pan-5 de terminación roma se insertó en el sitio *FspI* para generar pX`PLNK (2994 bp). La región *FseI* 5`terminal de Pan 5 (bp 1-3607, SEQ ID NO: 1) se insertó en los sitios *SmaI* y *FseI* de pX`LNK para generar el plásmido pX`Pan5-5` (6591 bp). La región *SnaBI-NdeI* de pX`Pan5-5` se escindió y reemplazó con la cinta *CeuI/SceI* la cual fue amplificada por PCR de pRCS para crear pX`Pan5-5` E1 (4374 bp). Brevemente, una secuencia que contiene los sitios cortadores raros *I-CeuI* y *PI-SceI* fue amplificada por PCR de pRCS (3113 bp). El iniciador 3` PCR se introdujo en el sitio NdeI en el producto de PCR.

Para extender el ADN del Pan-5 en pX`Pan5-5` E1 (4374 bp), la región *FseI-MluI* del Pan 5 (bp 3607-15135, SEQ ID NO:1) se añadió, para crear pX`Pan5-

5`Mlu (15900 bp). La terminal residual Mlul-3` de la secuencia Pan5 (bp 15135 – 36462, SEQ ID NO: 1) se adiciona al vector entre los sitios Mlul y EcoRV del vector poli-ligando para formar pX`Pan5 E1 el cual contiene la secuencia completa Pan5 que contiene una supresión en la región E1.

C. Generación De Virus Recombinantes.

Para generar los adenovirus recombinantes de pX`Pan5 E1 el plásmido se cotransfecta con un auxiliar que expresa E1 o desde una línea celular empacada que expresa E1, como una línea celular 293 o una línea preparada como se describió aquí. La expresión de E1 en las células empacadas, permite la replicación y el empaquetamiento de Pan5 E1 en una cápside virión. En otra incorporación, las células empaquetadas transfectadas con pX`Pan5 E1 son transfectadas con un vector de adenovirus como se describió antes que suplementa el transgen de interés. La recomendación homóloga tiene lugar entre el auxiliador y el plásmido que permite que las secuencias de adenovirus – transgen en el vector, sean replicadas y empacadas en cápsides virión en el adenovirus recombinante.

La transfección se sigue sobre agar por 2 semanas, después de las cuales los virus se siembran, se expanden y se tamizan para la expresión del transgen. Algunos pases adicionales de purificación, se siguen para otra expansión de los cultivos. Finalmente las células son cosechadas, un extracto de virus preparado y el adenovirus recombinante de chimpancé que contiene el transgen deseado, se

purifica por ultracentrifugación de densidad flotante en un gradiente de CsCl o por un medio alternativo conocido por aquellos versados en el arte.

Ejemplo 5 – Generación Del Vector Recombinante E1-suprimido Pan-6.

A. Estrategia para la construcción del plásmido adenoviral Pan-6.

1. *Clonación de los fragmentos terminales.*

El virus Pan-6, se desproteiniza por tratamiento con pronasa y proteinasa K y extracción con fenol. Los ligandos sintéticos Pme de12 bp se ligan sobre el ADN viral como se describe por Berkner y Sharp, *Nucleic Acids Research*, 11. 6003 (1983). El ADN viral, entonces se digiere, con Xba I para aislar el fragmento terminal 5` (6043). El fragmento Ad6 XbaI 5` se entonces liga a la unión pX en los sitios Sma I y Xba I para formar pX-Ad Pan6-0-16.5. El ADN viral con ligandos Pme I también se digiere con Pac I para aislar el fragmento terminal 3` de 6475 bp y clonarlo en los sitios de la unión pX en Pac I y Sma I, resultando en pXAd Pan6-82-100.

2. *Supresión de E1 del clon 5`*

Para suprimir E1 (m.u. 1.2-9), el fragmento BsiWi-Xba I en pXAd Pan6-0-16.5 se reemplaza con un fragmento PCR que encaja con el fragmento m.u.9-16.7 tratado con BsiWi y Xba I, conduciendo al pX-Ad-Pan6 m.u. 0-1, 9-16.5.

3. *Fusión de clones 5`y 3` y crear un sitio de anclaje para aceptar el fragmento medio Posterior III (Hind III).*

Primero, el clon 5`, pX-Ad-Pan-6 m.u. 0-1, 9-16.5, además se expande mediante la inserción del segundo fragmento Xba (4350 bp, m.u. 16.5-28) del genoma Pan-6 en el sitio Xba en el pX-Ad-Pan-6 m.u. 0-1, 9-16.5 este constructo llamado pX-Ad-Pan-6 m.u. 0-1, 9-28.

Segundo, el clon 3` también se expande mediante la inserción del fragmento I Mlu I/pac de 15026 bp cubriendo m.u. 41-82 del genoma Pan-6 en los sitios Mlu I/Pac I de pXAdPan-6-82-100, generando pXAdPan-6-41-100.

Entonces, un fragmento Posterior (Hind III) /Eco 47III Pan-6 se aísla de pX-Ad-Pan-6 m.u. 0-1, 9-28 y se subclona en pXAdPan-6 m.u.41-6 a los sitios romos Posterior III y Xba1. Este clon de fusión 5` y 3` se llama pXAdPan-6m.u.0-1,9-19.5, 64-100.

4. *Caída del fragmento medio del genoma en la fusión del clon.*

Un fragmento Posterior III (Hind III) 16335 (m.u.19.5-64) del Pan-6 se inserta en el sitio Hind III de pXAdPan-6m.u.0-1, 9-19.5, 64-100 para formar pXAdPan-6m.u.0-1, 9-100.

5. *Introducción de un marcador selectivo PKGFP en el constructo final para clonar directamente el gen de interés y selección verde/blanco de los transformantes recombinante.*

Una cinta minigen que expresa GFP bajo un promotor lac y que está flanqueado con sitios de reconocimiento de intrón raro que codifica enzimas de restricción, PI-Sce-I y I-Ceu I, se aisló de la lanzadera-p (pShuttle)-pkGFP (desnuda) por digestiones con Sap I y Dra III seguida de una reacción de llenado. El plásmido de la lanzadera-p (pShuttle)-pkGFP (desnuda) tiene 4126 bp de longitud, y contienen un ColE1-Ori, un gen resistente a la Kanamicina, plac, un promotor LacZ-GFPmut3-1 cds (Clontech), y un GFPmut3-1 cds (Clontech). Esta cinta se subclona en SrfI se corta y se roma en pXAdPan-6 0-1, 9-100. Este constructo final se llama pX-Pan-6-pkGFPmu.0-1, 9-100 el cual es útil para generar clones moleculares E1 suprimidos de Pan-6 que llevan los genes de interés por unión directa y selección verde/blanco en combinación con los vectores genéricos lanzadera-p-pkGFP.

B. Estrategia Alternativa Para La Generación De Plásmido Pan-6

1. *Clonación de fragmento 5` terminal.*

El virus Pan-6 se desproteiniza mediante tratamiento con pronasa y proteinasa K y extracción con Fenol como se describió anteriormente y se unieron ligantes PmeI sobre el ADN viral como se describió. El fragmento AdPan-6 5` Xba I se aísla y se une en pX para formar pXAdPan-6-0-16.5 (9022 bp) como se describió en la parte A anterior.

2. *Supresión de E1 del clon 5`*

Para suprimir E1 (mu1.2-9), el pXAdPan-6-0-16.5 se digiere con SnaBI y NdeI para remover las regiones que codifican las proteínas E1a y E1b (3442-6310 bp) este vector es subsiguientemente digerido con BsiWI en preparación para que rome con la cinta minigen que lleva un marcador selectivo.

3. *Introducción de un marcador selectivo.*

Una cinta minigen que expresó GFP bajo un promotor lac y el cual está flanqueado con sitios de reconocimiento de intrón raro que codifica las enzimas de restricción, PI-XceI, se aisló de la lanzadera p-pkGFP como se describió anteriormente. Luego el fragmento DraIII-SapI se une con el digerido pXAdPan-6-0-16.5 para formar pXAdPan-6- MU 0-16.5 E1 (7749 bp).

4. *Extensión De Secuencias Adenovirales De Pan-6.*

pXAdPan-6-MU0-16.5 E1 fue sometida a digestión con XbaI para permitir la inserción de un ligante XbaI-RsrII. Un fragmento de digestión XbaI/RsrII del genoma Pan-6 se aisló (mu 28-100, 26240 bp) y se ligó en el digerido XbaI/RsrII pXAdPan-6- MU 0-16.5 E1 para proporcionar pXAdPan-6- MU 0-16.5, 28-100. Un segundo fragmento XbaI del genoma Pan-6 (mu 16.5-28, 4350 bp) se une entonces, en este plásmido para formar pXAdPan-6 MU 0-1, 9-100 (38551 bp).

C. Generación De Adenovirus Recombinantes.

Para generar los adenovirus recombinantes del plásmido E1-suprimido Pan-6 preparado como se describió en las partes A o b, el plásmido se cotransfecta con un auxiliar que expresa E1, o de una línea celular empacada que expresa E1, tal como una línea celular 293 o una línea celular preparada como aquí se describe. La expresión de E1 en la célula empacada permite la replicación y empaquetamiento del Pan-6-pkGFP mu.0-1, 9-100 en una cápside virión. Alternativamente, las células empacadas transfectadas con pX-Pan-6-pkGFP mu.0-1, 9-100 se transfectan con un vector de adenovirus como se describió antes en presencia de otro transgen de interés.

Ejemplo 6 – Generación del vector recombinante E1-suprimido de Pan-7.

A. *Generación De Plásmidos de Pan-7.*

Un ligante sintético que contiene los sitios de restricción PacI-SmaI-FseI-MluI-EcoRV-PacI se clonó en pBR322 que fue cortado con EcoRI y NdeI. El extremo izquierdo (bp a 3618) de Ad Pan-7 se clonó en el ligante entre los sitios SmaI y FseI. El adenovirus E1 se escindió del terminal izquierdo clonado cortando con SnaBI y NdeI e insertando una cinta I-CeuI-GFP-PI-SceI de lanzadera p (Clontech) en su lugar. El plásmido resultante se cortó con FseI y MluI y fragmento FseI (3618 bp) de Ad Pan-7 a MluI (bp 155114 se insertó para extender el extremo izquierdo. El constructo (pPan-7pGFP) se completó por inserción del fragmento terminal derecho Ad Pan-7 de 21421 bp del sitio MluI (15114 bp) en el plásmido anterior entre MluI y EcoRV para generar un clon molecular completo de adenovirus Pan-7 E1-suprimido, adecuado para la generación de los adenovirus recombinantes. Opcionalmente, un transgen deseado se insertó en los sitios I-CeuI y PI-SceI del plásmido vector pPan-7 recientemente creado.

B. *Construcción De Vectores Virales Pan-7 E1-Suprimidos.*

Para generar los adenovirus recombinantes de pPan-7 E1, el plásmido se cotransfecta con una auxiliar que exprese E1, o desde una línea de células empacadas que expresan E1, tal como la línea celular 293 o una línea celular preparada como aquí se describe. La expresión de E1 en la célula empacada permite la replicación y empaquetamiento del Pan7 E1 en una cápside virión. En otra incorporación, la célula empacada transfectada pX'Pan7 E1 se transfecta

con un vector de adenovirus como se describió antes en presencia del transgen de interés. La recombinación homóloga ocurre entre el auxiliar y el plásmido, lo cual permite que las secuencias adenovirus-transgen en el vector a ser replicado y empacado en las cápsides virión, resulten en adenovirus recombinantes. La transfección y la purificación son como se describió anteriormente.

Ejemplo 7- Generación De Vectores Plásmidos Que Expresan Los Genes E1.

Los vectores plásmidos construidos, los cuales codifican el gen de la región Pan-5 E1, y estos plásmido se usan para generar líneas celulares estables que expresen las proteínas E1 virales.

La región E1 de Pan5, está clonado en pX` esencialmente como se escribió en el ejemplo 4 anteriormente, previamente para reemplazar de esta región con el fragmento de lanzadera (Clontech). El plásmido de expresión contiene la secuencia del genoma adenoviral Pan5 que conduce a por lo menos de 1 a 3959 bp en las secuencias genómicas de Pan5. Entonces, el plásmido de expresión contiene la secuencia que codifica E1a y E1b de AdPan5 de chimpancé bajo el control de un promotor heterólogo. Similares plásmidos de expresión se pueden generar utilizando las regiones E1 de AdPan6 y AdPan7, identificadas en las tablas anteriormente.

Ejemplo 8 – Generación De Líneas De Célula Que Expresan Proteínas E1 De Adenovirus De Chimpancé.

Las líneas de células que expresan proteínas E1 virales, son generadas por transfectación HeLa (ATCCAcc.No.CCL2) con el plásmido del ejemplo 6. Estas líneas de células son útiles para la producción de adenovirus recombinantes E1-suprimidos de chimpancé por co-transfectación de ADN viral genómico y los plásmidos de expresión descritos anteriormente. La transfectación de estas líneas de células, como también el aislamiento y la purificación de los adenovirus recombinantes de chimpancé, son realizados por métodos convencionales para otros adenovirus, i.e., adenovirus humanos [ver, e.g., Horwits, citado antes y otros textos estándar].

A. *Líneas de células que expresan proteínas E1 de Pan5.*

Las células HeLa en cajas de 10 cm. son transfectadas con 10 microgramos de pX-Pan51-E1ADN utilizando un kit Cellphect™ (Pharmacia, Uppsala, Sweden) y siguiendo el protocolo del fabricante. 22 horas después de la transfección, las células son sujetas a un choque de 3 min., de glicerol (15% de glicerol en solución amortiguadora de Hepes, pH 7.5) lavadas una vez en medios de DMEM (HeLa) o F12K (A549; Life Technologies, Inc., Grand Island, NY) suplementados con 10% de FCS, 1% Pen-Strep, luego fueron incubados a 37° C en el medio descrito anteriormente. Las células transfectadas entonces fueron repartidas por duplicado

en placas de 15 cm., en las relaciones de 1:20, 1:40, 1:80, 1:320. Después de la incubación a 37°C por toda la noche, el medio se suplementó con G418 (Life Technologies, Inc.) a una concentración de 1 /mL. El medio se reemplazó cada 5 días y los clones se aislaron 20 días después de la transfección.

Los clones de las células HeLa E1 se aislaron y ensayaron para sus habilidades de aumentar la infección de virus adeno-sociados (AAV) y la expresión de la proteína LacZ recombinante como se describe más abajo.

B. Ensayo De Aumento AAV Por Tamizado De Líneas Celulares Que Expresan E1.

La AAV requiere proteínas codificadas por adenovirus en orden a completar su ciclo de vida. Las proteínas E1 adenovirales así como también las proteínas ORF6 codificadas por la región4 son necesarias para el aumento de la infección AAV. Se usa un ensayo para la expresión E1 basado en el aumento AAV. Brevemente, el método para identificar células que expresan E1 adenoviral, comprenden las etapas de infectar en cultivos separados una célula putativa que expresa E1 en una célula que contiene una secuencia no adenoviral, con ambos un virus adeno-asociado AAV que expresa un gen marcador y un AAV que expresa el ORF6 del gen E4 de adenovirus humano, por un tiempo adecuado. La actividad del gen marcador en las células resultantes se mide y aquellas células con un mayor grado medible en la actividad del marcador significativamente que

las células del control, se seleccionan como células que expresan E1 confirmadas. En el siguiente experimento, el gen marcador es el gen lacZ y la actividad del marcador es la aparición de tinción de azul.

Por ejemplo, las líneas celulares descritas anteriormente, así como también las células de control transfectadas (HeLa) se infectan con 100 genomas por célula de un vector AAV que apoya a un gen marcador, e. g., AV.LacZ [K. Fisher *et al.*, *J.* Virol., **70**: 520 (1996)] y un vector que expresa la región ORF6 de un humano 5 (AV.orf6). La secuencia de ADN del plásmido genera un nuevo virus adeno-asociado recombinante (rAAV) que contiene el transgen *LacZ* y Ad E4 ORF6, el cual es una apertura que lee el marco cuyo producto de expresión facilita la conversión de trenzado (ss) único a trenzado (ds) doble de el ADN genómico de rAAV. Estos vectores incubados en un medio que contiene 2% FCS y un 1% de Pen-Strep a 37°C por 4 horas, punto en el cual se adicionó un volumen igual de medio que contiene 10% de FCS. Se entenderá por alguien versado en el arte que el gen marcador (o gen reportero) puede ser empleado en el primer vector AAV de este ensayo, e. g., fosfatasa alcalina luciferasa, y otras. También se puede usar un ensayo anticuerpo-enzima para niveles cuantitavos de antígeno, donde el marcador expresa un antígeno. El ensayo no está limitado por la identidad del gen marcador. A las 20 o 24 horas pos-infección, las células son teñidas para la actividad LacZ usando métodos estándar. Después de 4 horas, las células se observan microscópicamente y las líneas celulares que son

significativamente más azules que A549 o los controles celulares HeLa son marcadas como positivas.

Ejemplo 9. Entrega De Transgen a Células Huésped.

Los adenovirus de chimpancé recombinantes que resultan como se escribieron en los ejemplos 4, 5, o 6 anteriormente, se emplean para entregar el transgen a un mamífero, preferiblemente una célula humana. Por ejemplo, siguiendo la purificación del virus recombinante, 293 células de riñón de embrión humano, son infectadas en un MOI de 50 partículas por célula. La expresión GFP se documentó 24 horas después de la infección.

A. Transferencia De Genes En Modelos de Ratón Vía Vectores Pan-6, Pan-7 y Pan-9.

Las eficiencias de transferencia de genes y perfil toxicológico de adenovirus de chimpancé recombinante, se compararon en hígado de ratón directamente a la transferencia del gen, a pulmón de ratón dirigido a la transferencia de gen y músculo de ratón dirigido a la transferencia del gen.

Los vectores adenovirales E1-suprimidos conteniendo LacZ bajo el control del promotor CMV fueron construidos utilizando técnicas descritas aquí para Ad5 humanos, Pan6 de chimpancé, Pan7 de chimpancé y Pan9 de chimpancé (C68).

Los vectores fueron entregados a ratones desnudos inmunodeficientes NCR (80 para cada estudio) como sigue. Para el estudio de hígado, 100 l (1x 10^{11} partículas, se inyectaron en la vena de la cola. Para el estudio del pulmón, 50 l (5 x 10^{10} partículas), se entregaron intratraquealmente). Para el estudio del músculo, 25 l (5 x 10^{10} partículas, se inyectaron en la tibia anterior). Los ratones fueron sacrificados a los 3, 7, 14 y 28 días después de la inyección del vector (5 animales por grupo en cada uno punto de cada tiempo). En cada necropsia, el hígado, pulmón y tejidos musculares, se cosecharon y se prepararon en criobloques y embebiéndolos en parafinas. Los criobloques fueron seccionados para tinción X-gal y las secciones de parafina, son teñidas H&E para análisis sistopático. En cada punto del tiempo, se llevó a cabo un sangrado terminal, las muestras de suero fueron sujetas a las pruebas de función de hígado.

Se observó en este experimento que los adenovirus de chimpancé Pan-6, Pan-7 y Pan-9, fueron menos eficientes que el huAd5 en la transferencia del gen al hígado y al pulmón. Sin embargo esto puede ser deseable en ciertas circunstancias, para reducir la toxicidad observada en el hígado por huAd5. La eficiencia de transferencia génica en el músculo, varía menos entre el serotipos.

B. *Estudio En Ratón Para La Posibilidad De La Administración De Vectores de Adenovirus Por intercambiando de Serotipos Entre Vectores Adhu5, Pan-6, Pan-7 y Pan-9.*

Se administró a los ratones vectores de LacZ (C57/B16; 4/grupo) basados en huAd5 Pan-6, Pan-7 y Pan-9 (H5.040CMVLacZ, Pan6.000CMVLacZ, Pan7.000CMVLacZ, Pan9.000CMVLacZ; 10^{11} partículas por inyección) por la vena de la cola. 30 días después, los ratones se administraron con vectores de adenovirus que expresan 1-anttripsina (H5.040CMVhA1AT, Pan6.000CMVhA1AT, 1x 10^{11} partículas, Pan7.000CMVhA1AT, Pan9.000CMVhA1AT; 10^{11} partículas por inyección). El éxito de la transducción por la readiministración del vector se monitoreó, midiendo el suero 1- antitripsina a los 3 y 7 días siguientes a la administración.

Se determinó la habilidad de los vectores de adenovirus basada en huAd5 Pan-6, Pan-7 y Pan-9 respectivamente para transducir los hígados de los ratones en presencia de anticuerpos neutralizantes para los otros serotipos. Los resultados son tabulados aquí.

1a. Injection	2a. Injection	Neutralización cruzada
Adhu5	Adhu5	Si (+ve control)
	Pan-6	No
	Pan-7	No
	Pan-9 (C68)	No
Pan – 6	Adhu5	No

	Pan-6	Si (+ve control)
	Pan-7	Si
	Pan-9 (C68)	No
Pan - 7	Adhu5	No
	Pan-6	Si
	Pan-7	Si (+ve control)
	Pan-9 (C68)	Si
Pan -9 (C68)	Adhu5	No
	Pan-6	No
	Pan-7	Si
	Pan-9 (C68)	Si (+ve control)

Capacidad de los vectores para transducir el hígado de múrido en presencia de anticuerpos de otros serotipos.

Así, la inmunización con huAd5 no se previene con la readministración ya sea de uno de los vectores de adenovirus de chimpancé Pan-6, Pan-7 o Pan-9 (C68). Este experimento también parece indicar que Pan-7 está entre Pan6- y Pan-9 en el espectro de relaciones antigénicas y reacciones cruzadas con ambos; sin embargo, Pan-6 y Pan-9, no se neutralizan uno al otro. Esto es un resultado sorprendente basado entre comparaciones de homología, el cual indica que Pan-6 es un poco distinto de Pan-7 y Pan-9. La evaluación de antisueros generados contra Pan-9, indica neutralización no cruzada contra Pan-6 pero alguna neutralización contra Pan-7, que argumenta que Pan-6 es diferente de los otros.

Ejemplo 10. Generación De Vector SV-25 E1-Suprimido Recombinante.

Se construyó un plásmido que contiene el genoma SV-25 completo, excepto por una supresión E1 hecha por ingeniería. En el sitio de la supresión E1 que reconoce sitios para las enzimas de restricción I-CeuI y PI-SceI los cuales podrían permitir la inserción de transgenes de un plásmido de disparo donde la cinta de expresión del transgen está flanqueada por estas dos enzimas que reconocen los sitios donde fueron insertados.

Se clonó un ligante sintético que contiene los sitios de restricción SwaI-SnaBI-SpeI-AflII-EcoRV-SwaI en pBR322 que fue cortado con EcoRI y NdeI. Estos se hizo recociendo ambos de los dos oligomeros sintéticos SV25T (5-AAT TTA AAT ACG TAG CGC ACT AGT CGC GCT AAG CGC GGA TAT CAT TTA AA-3`, SEQ ID NO: 38) y SV25B (5`-TAT TTA AAT GAT ATC CGC GCT TAA GCG CGA CTA GTG CGC TAC GTA TTT A-3`, SEQ ID NO: 39) e insertándolo en el digerido pBR322 con EcoRI y NdeI. Se clonó el extremo izquierdo (bp I a 1057, SEQ ID NO: 29) de Ad SV25 en el ligante anterior entre los sitios SnaBI y SpeI. Se clonó el extremo derecho (bp28059 a 31042, SEQ ID NO: 29) de Ad SV25 en el ligante entre los sitios AflII y EcoRI. El adenovirus E1 entonces se escindió entre el sitio EcoRI (pb 547) hasta XhoI (bp 2031) desde el extremo izquierdo clonado como sigue. Se insertó una cinta I-CeuI-PI-SceI generada por PCR desde el vector de disparo (Clontech) entre los sitios EcoRI y SpeI. El fragmento 10154 bp XhoI de Ad SV-25 (bp 2031 a 12185, SEQ ID NO: 29) entonces fue insertado en el sitio SpeI. El plásmido resultante fue digerido con HindIII y el constructo (pSV25) se completó por inserción del fragmento 18344 bp SV-25 HindIII (bp 11984 a 30328,

SEQ ID NO: 29) para generar un clon molecular completo del adenovirus SV25 E1-suprimido adecuado para la generación de los adenovirus recombinantes. Opcionalmente, un transgen deseado se inserta en los sitios entre I-CeuI y PI-SceI del plásmido vector pSV25 creado recientemente.

Para generar un vehículo AdSV25 que lleve un gen marcador, una cinta de expresión GFP (proteína verde fluorescente) previamente clonada en el plásmido disparo (Clontech) se escindió con las enzimas de restricción I-CeuI y PI-SceI y se ligó en pSV25 (u otro de los plásmidos Ad de chimpancé aquí descritos) digerido con las mismas enzimas. El plásmido resultante (pSV25GFP) ffue digerido con SwaI para separar el plásmido de columna y se transfectó en la línea celular HEK 293 que complementa E1. Cerca de 10 días después, se observó un efecto citopático que indica la presencia de virus replicados. El éxito en la generación de un vector adenoviral basado en Ad SV25 que expresa GFP se confirmó aplicando el sobrenadante del cultivo transfectado sobre cultivos celulares frescos. La presencia de células infectadas secundariamente, se determinó por observación verde fluorescente en una población de las células.

Ejemplo 11. Construcción De Vectores De Pan-5, Pan-6, Pan-7 y C68 E3-Suprimidos.

128
129

En orden a mejorar la capacidad de clonación de los vectores adenovirales, la región E3 puede ser suprimida, ya que esta región codifica genes que no se requieren para la propagación del virus en el cultivo. Hacia este extremo se hicieron, las versiones de E3-suprimidas de Pan-5, Pan-6, Pan-7 y C68 (un fragmento de 3.5 kb Nru-AvrII que contiene E31-9 se suprimió).

A. *Vector Basado En Pan5 E3- suprimido*

El plásmido pPan-5pkGFP E1-suprimido, se trató con endonucleasa Avr II para aislar ese fragmento de 5.8 kb que contiene la región E3 y recircular pPan-5-pkGFP con una supresión Avr II desde el constructo pPan-pkGFP-E3-Avr II. Subsiguientemente, el fragmento 5.8 kb Avr II, se subclonó en pSL-Pan5-E3-AVr II para una supresión adicional de la región E3 por digestión con Nru I. Esto conduce a un plásmido pSL-pan5-E3-suprimido. El constructo final pPan5-E3-pkGFP se produjo mediante la remoción de un fragmento 4.3 kb Avr II/Spe desde el plásmido pSL-pan5-E3-suprimido e insertando en pPan5-pkGFP-E3-Avr II en el sitio Avr II. En este constructo final, se logró una supresión de 3.1 kb en la región E3.

B. *Vector Basado en Pan6 E3-suprimido*

El clon molecular pPan6-pkGFP E1-suprimido se digirió con Sbf I y Not I para aislar un fragmento de 19.3 kb y se volvió a unir al sitio Sbf I. El constructo resultante pPan6-Sbf I-E3 se trató con Eco 47 III y Swa I, generando pPan6-E3. Finalmente el fragmento 21 kb Sbf I de la digestión Sbf I de pPan6-pkGFP se subclonó en pPan6-E3 para crear pPan6-E3-pkGFP con una supresión de 4 kb en E3.

C. *Vectores Pan7 y Pan9 E3-Suprimidos.*

La misma estrategia se utilizó para alcanzar las supresiones E3 en ambos vectores. Primero, un fragmento de 5.8 kb de Avr II que abarca la región E3 se subclonó pSL-1180, seguido por su presión de E3 mediante digestión con Nru I. Los plásmidos resultantes se trataron con Spe I y Avr II para obtener fragmentos de 4.4 kb y clonarlos entre pPan7-pkGFP y pPan9-pkGFP en los sitios Avr II para reemplazar el E3 original que contiene fragmentos Avr II, respectivamente. Los constructos finales pPan7-E3-pkGFP y pPan9-E3-pkGFP tienen supresiones E3 de 3.5 kb.

Ejemplo 12 – Construcción De Vector Pan-7 E3-E4-suprimidos.

Aunque la supresión de la región E1 de los adenovirus (primera generación de vectores de adenovirus) le suministra a ellos, una replicación incompetente, la expresión de los genes de columna del vector adenoviral no es totalmente abolida. La supresión de la región E4 atenúa considerablemente la expresión este gen

residual y puede conferir una ventaja de seguridad. Se ha construido un vector Pan-7 E4-suprimido que contiene una supresión de 2.5 kb (se suprimió un fragmento PvuII-AgeI que contiene E4ORFI-ORF7). Se generó un alto título estándar usando una línea celular basada en HEK 293, la cual en adición a E1, expresa un gen E4 esencial (orf 6).

1. *Supresión E4 En El Clon Molecular De Pan7.*

Se suprimió un fragmento de 19 kb Xba I desde pPan7-GFP para crear pPan7-Xba I desde el cual se suprimió un fragmento 2.5 kb E4 por digestión parcial con Age I y Pvu II, que resulta en un plásmido pPan7-Xba I.pPan7-E4-pkGFP que fue generado desde pPan7-Xba-I-E4 en dos etapas de clonación secuencial, adicionando los fragmentos de 19 kb Xba I y de 15 kb I-Ceu I/Mlu I, los cuales venían del constructo pPan7-pkGFP

2. *Introducción De Supresiones E3 y E4 En El Vector Pan9.*

Un plásmido de 11 kb pPan9-EcoRI, que contiene la región E4 se creó mediante una recuperación de un fragmento EcoRI de 11 kb desde pPan9pkGFP después de la digestión de EcoRI y autoligación. La región E4 fue suprimida desde este constructo por digestión/llenado en Age I y digestión parcial de Pvu II y auto-ligación para generar pPan9-EcoR I-E4. Un fragmento EcoR I de A 23 kb se aisló desde pPan9-pkGFP e insertó en pPan9-pkGFP e insertó pPan9-EcoRI-E4

en el sitio EcoRI, seguido de la adición del fragmento Avr II de 5.8 kb desde pPan9-pkGFP, para formar el producto final pPan9-E3-E4-pkGFP. Comparado con el tamaño del genoma de Pan9 tipo salvaje, este vector E1-E3-E4-suprimido, podría tener una capacidad transgénica superior a 8 kb.

3. *Introducción De Cintas de Minigen Con Genes De Interés Que Incluyen Genes Reporteros, Glicoproteínas y Proteínas Nucleares de Ebo En Clones Moleculares de Vectores Pan.*

Se utilizó una clonación directa altamente eficiente y un procedimiento de selección verde/blanco para crear clones moleculares de virus recombinantes. Brevemente, los genes de interés se clonaron en p-disparoGFP (pShuttlepkGFP) por tamizado de colonias blancas para recombinantes. Subsiguientemente, las cintas del minigen se transfirieron en los plásmidos de columna de adenovirus de chimpancé, pPanX-pkGFP con varias supresiones, fácilmente por inmersión con la cinta pkGFP en los sitios I-Ceu I y PI-Sce I y tamizando unas pocas colonias blancas para los recombinantes correctos.

4. *Rescate De Clones Moleculares De Vectores Pan Con Múltiples Supresiones En Regiones Tempranas y Propagación De Virus.*

Para rescatar los clones moleculares E1-E3-suprimidos de vectores de adenuvirus de chimpancé, los clones se linealizaron con enzimas de restricción apropiadas y se transfectaron en células 293 regulares. Una vez observado un efecto citopático completo (CPE) en las células transfectadas, el lisado crudo, se cosechó y se expandió en células 293 para infección a gran escala. Los virus se purificaron por el método de sedimentación con CsCl.

Para los vectores E1-E4 y E1-E3-E4-suprimidos, células 10-3, una línea celular basada en 293 E1-E4-complementada se usó para el rescate y la propagación de vectores. La expresión del gen E4ORF6 en las células E-3 se indujo por adición de 150 M de $ZnSO_4$ al medio del cultivo.

Ejemplo 13 – Vacunación Con Vectores De Adenovirus Que Expresan EboZ GP Tipo Salvaje y Variante.

Los vectores AdHu5 o AdC7 que expresan quimeras envolturas de Ebola se produjeron por experimentos de inmunización *in vivo* en ratones C57BL/6. Los virus recombinantes con diferentes columnas virales, se crearon por el método de creación molecular en el cual las cintas minigen, se insertaron en el lugar de las supresiones E1. Los clones moleculares de todos los virus recombinantes, se rescataron y se dejaron crecer en células 293 para purificación a gran escala usando el método de sedimentación con CsCl.

Cinco variantes EboZ codificadas por AdHu5 o AdPan7 (C7) se seleccionaron y produjeron para evaluar sus inmunogenicidades relativas que siguen a una inyección Ad intramuscular. El wt Ebo, una variante Ebo soluble Ebo 1, Ebo 2, Ebo 3, Ebo 4, Ebo 5S, Ebo 6S, Ebo 7S y Ebo 8S se evaluaron en los estudios de vacuna inicial. Para los datos resumidos en la siguiente tabla, el número de partículas virales (por mL o total) producidas y amplificadas desde células 293 infectadas, se estableció por lectura espectrofotométrica.

Tabla: Producción de vector adenoviral Adhu5 ó AdC7 que codifica la variante EboZ.

	HuAd5		AdC7	
	Título (VP x 10^{12}/ml)	Producción total (VP x 10^{12})	Título (VP x 10^{12}/ml)	Producción total (VP x 10^{12})
Ebo wt	2.6	12	4.3	43
EboS	4.9	49	4.6	55
Ebo 2	2.1	9	5.8	93
Ebo 3	1.7	8	5.3	95
Ebo 4	3	12	4.1	62

El vector se administró intramuscularmente (10^{11} copias de genomas/célula) en ratones C57BL/6 y la presencia de virus anticuerpo neutralizantes (VNA0 se

evalúo 28 días después como una primera medida de una respuesta inmune generada contra la glicoproteína que envuelve el Ebola. El VNA se define aquí como el anticuerpo del suero capaz de inhibir la transducción de células HeLa mediada por vectores basados en HIV seudotipados con la envoltura Ebola tipo salvaje.

El VNA para los seudotipos EboZ se detectó con la producción AdPan7 (C7) de más altos títulos que AdHu5. El EboZ 3 que sacó el más alto VNA en términos de los transgenes puestos como blanco. En la siguiente tabla se proporciona los datos resumidos de los títulos de anticuerpo neutralizante para seudotipos HIV-EboZ-GFP (dilución recíproca). (N=5 animales/grupo).

	Títulos: VNA		
	EboZ tipo salvaje	EboZs	EboZ 3
AdHu5	12	16	12
AdC7	44	12	140

Ejemplo 14 – Expresión Pan-7 Mediada De Proteína Ebola.

Se han iniciado estudios con ratón para evaluar vectores Pan-7 que expresan proteínas de envoltura de Ebola y antígenos nucleares Ebola. Estos están directamente orientados a la evaluación de anticuerpos neutralizantes en ratones C57B1/6 inyectados intramuscularmente (IM) con Adhu5 o Pan-7 que expresan cada uno de 4 constructos env Ebola.

A. *Evaluación De CTL Desde Ratones C57B18/6 Inyectados IM Con Adhu5 o Pan-7 Que Expresan Los Constructor env Ebola.*

1. Experimento de exploración en ratones con el virus Ebola.

Las respuestas de anticuerpos neutralizantes (NAB) a las envolturas Ebola se analizaron mirando en sueros de ratón inmunizados la neutralización mediada de un vector (HIV) lentiviral seudotipado con los varios constructos (eEbo, NTD2, NTD3, NTD4) de la glicoproteína que envuelve el Ebola. Los ratones C57BL/6 o BALB/c recibieron una única inyección intramuscular de 5×0^{10} partículas de raton C7 (Ad Pan-7) que codifica la variante de la envoltura de Ebola. El anticuerpo neutralizante se evalúo 30 después de la vacunación. Brevemente, el vector HIV seudotipado de Ebola Zaire seudotipo HIV que codifica para la -ga lactosidasa (EboZ-HIV-LacZ) se incubó por 2 horas a 37°C con diferentes diluciones del suero de ratón inactivado por calentamiento. Siguiendo la incubación con suero, se usó entonces el EboZ-HIV-LacZ para infectar las células HeLa por 16 horas a 37°C. La infectividad se reveló mediante tinción de la X-gal de las células HeLa

136
137

tranasducidas positivas para -g alactosidasa. El título de neutralización representa la dilución recíproca del suero en donde se observa una disminución del 50% en el número de células Pan-7 azul positivo de - g alactosidasa. Los sueros se recogieron a los 30 post-inmunización, la cual consistió en una única administración intramuscular (I.M.) de 5 x 10^{10} partículas por animal. El anticuerpo neutralizante para Ebola seudotipo HIV podía de ser detectado de todos los grupos con títulos de anticuerpo en el intervalo entre 20 para Ad-EboZ, (Adhu5 que expresa EboZ), Ad-NTD3 (Adhu5 que expresa NTD3) y C7-sEbo (Ad Pan-7 que expresa EboZ soluble) hasta arriba de 130 para C7-NTD3 (Ad Pan-7 que expresa NTD3 soluble) y C7-NTD4 (Ad Pan-7 que expresa NTD3 soluble). La misma estrategia de inmunización en ratones BALB/c dió títulos de anticuerpo neutralizante bajos para Ad- C7-NTD2, y NTD4.

B. *Respuesta Inmune Celular.*

La respuesta inmune celular para la envoltura Ebola en los ratones C57BL/6 se evaluada 8 días después de una única administración intramuscular de 5x10^{10} partículas de C7-LacZ o la variante de la envoltura C7-Ebola por animal. Los ratones se vacunaron I.M. con 5 x 10^{10} partículas de C7 que codifica LacZ o la variante de envoltura Ebola. Los linfocitos esplénicos de ratones inmunizados, se recogieron a los 8 días de la vacunación y se estimularon in vitro con un células alimentadoras (linfocitos esplénicos de ratones no tratados infectados con el serotipo 5 de adenovirus humano que codifica para la envoltura Ebola tipo

salvaje, e irradiados). Se hicieron los ensayos CTL estándar a las 5 horas usando células C57 singénicas marcadas con ^{51}Cr transfectadas con un expresor de EboZ.

Se observó una respuesta positiva de linfocito T (CTL) citotóxico MHC restringido de todos los AdPan-7 que codifican para variantes de envoltura Ebola con una alta respuesta en los ratones inmunizados con NTD2, NTD3 o NTD4. Verdaderamente, las células efectoras de C7 que codifican variantes de envolturas Ebola de ratones inmunizados reconocieron las células blanco transfctada con EboZ y dieron una respuesta CTL revocada hasta del 30% de lisis. Menos del 5% de lisis se vio con las células efectoras naïve o de ratones control inmunizados con LacZ lo que confirma que la lisis fue específica para antígenos de la envoltura Ebola.

C. Estudios De Protección.

La mayoría de los medios directos de evaluación de C7 (Ad Pan-7) que codifican por las variantes EboZ como una vacuna exitosa en ratones, consistieron en evaluar la protección contra la pérdida de peso y la muerte que sigue a un reto letal con ratones adaptados al virus del Ebola Zaire. Los ratones BALB/c se inmunizaron con una dosis única de 5×10^{10} partículas por animal como se hizo previamente y los animales vacunados fueron retados con 200 veces la dosis LD_{50} de ratones adaptados al Ebola de Zaire, 21 días más tarde. Todos los ratones de

control (vehículo C7-LacZ) murieron entre el día 5 y el día 9 posterior al reto. En contraste, los ratones vacunados excepto uno (del grupo C7-sEbo), sobrevivieron al reto con el Ebola de Zaire.

Se observó pérdida de peso en los ratones vacunados con C7-sEbo del día 4 al día 7. Señales de enfermedad tales como erección del pelo y de un ligero a severo estado letargico también se observó en los ratones vacunados con C7-sEbo, NTD2 y NTD3 entre el día 4 y el día 7. Los ratones inmunizados C7-EboZ y C7-NTD4 no mostraron señales de enfermedad. Sobre todo, una dosis única de C7-EboZ y C7-NTD4 protegió completamente a los ratones inmunizados, de la enfermedad y muerte posiblemente debido a una inmunidad mediada por células T significante.

Todos los documentos citados anteriormente, se incorporan aquí por referencia. Numerosas modificaciones y variaciones de la presente invención se incluyen en el alcance la especificación identificada anteriormente y se espera sean obvios para alguien versado en el arte. Tales modificaciones y alteraciones para las composiciones y procesos de la presente invención, tales como la selección de diferentes minigenes o selección o dosificación de los vectores o moduladores inmunes se acepta están incluidas en el alcance de las reivindicaciones aquí anexadas.

REIVINDICACIONES

139
140

1. Una secuencia aislada de ácido nucleico de adenovirus de simio seleccionada del grupo consiste de:

(a) Pan5 que tiene la secuencia de ácidos nucleicos 1 a 36462 de SEQ ID NO: 1;

(b) Pan6 que tiene la secuencia de ácidos nucleicos 1 a 36604 de SEQ ID NO: 5;

(c) Pan7 que tiene la secuencia de ácidos nucleicos 1 a 36535 de SEQ ID NO: 9;

(d) SVI que tiene la secuencia de ácidos nucleicos 1 a 34264 de SEQ de ID NO: 24;

(e) SV25 que tiene la secuencia de ácidos nucleicos 1 a 31044 de SEQ de ID NO: 29;

(f) SV39 que tiene la secuencia de ácidos nucleicos 1 a 34115 de SEQ de ID NO: 34, y

(g) Una secuencia de ácido nucleico complementaria a la secuencia de cualquiera de (a) a (f).

2. Una secuencia aislada de ácido nucleico de serotipo de adenovirus de simio seleccionada de uno o más del grupo que consiste de:

(a) Secuencias (ITR) de repetición terminal invertida 5';

(b) La región E1a del adenovirus, o un fragmento de la misma seleccionada de entre las regiones 13S, 12S y 9S;

(c) La región E1b del adenovirus, o un fragmento de la misma seleccionada de entre el grupo consistente de las regiones T pequeña, T grande, IX, y IVa2;

(d) La región E2b;

(e) La región L1, o un fragmento de la misma seleccionado de entre el grupo consistente de la proteína de 28,1 kD, polimerasa, agnoproteína, proteína 52/55 kD, y proteína IIIa;

(f) La región L2, o un fragmento de la misma seleccionado del grupo que consiste de las proteínas penton, VII, VI y Mu.

(g) La región L3, o un fragmento de la misma seleccionada del grupo que consiste de la VI, exón, o endoproteasa;

(h) La proteína 2 a;

(i) La región L4, o un fragmento de la misma seleccionado del grupo que consiste de la proteína de 100 kD, la homológa de 33 kD, y VIII;

(j) La región E3, o un fragmento de la misma seleccionado del grupo que consiste de E3 ORF1, E3 ORF2, E3 ORF3, E3 ORF4, E3 ORF5, E3 ORF6, E3 ORF7 E3 ORF8, y E3 ORF9;

(k) La región L5, o un fragmento de la misma seleccionado del grupo de una proteína fibra;

(l) La región E4, o un fragmento de la mismaza seleccionado del grupo que consiste de E4 ORF7, E4 ORF6, E4 ORF4, E4 ORF3, E4 ORF2, y E4 ORF1; y

(m) El 3' ITR, de cualquiera de Pan5 SEQ ID NO: 1; Pan6 SEQ ID NO: 5; Pan7 SEQ ID NO: 9; SVI SEQ ID NO: 24; SV25 SEQ ID NO: 29; y SV39 SEQ ID NO: 34, o la secuencia complementaria de cualquiera de (a) a (m).

3. Una proteína de adenovirus de simio codificada por la secuencia de ácido nucleico de acuerdo con la reivindicación 2.

4. Una molécula de ácido nucleico que comprende una secuencia adenoviral de simio heteróloga de acuerdo con la reivindicación 2.

5. La molécula de ácido nucleico de acuerdo con la reivindicación 4, en donde dicha secuencia adenoviral de simio codifica un producto génico adenoviral y esta operativamente unida a secuencias de control regulador las cuales expresan directamente el producto génico adenoviral en una célula huésped.

6. La molécula de ácido nucleico de acuerdo con la reivindicación 4 o 5, en donde dicha secuencia adenoviral de simio comprende la región E1 a de Pan5 SEQ ID NO: 1; Pan6 SEQ ID NO: 5; Pan7 SEQ ID NO: 9; SVI SEQ ID NO: 24; SV25 SEQ ID NO: 29; y SV39 SEQ ID NO: 34.

7. Una composición farmacéutica que comprende la molécula de ácido nucleico de acuerdo con la reivindicación 6 y un vehículo compatible fisiológicamente.

8. Una proteína cápside adenoviral de simio aislada seleccionada del grupo que consiste de:

(a) Una proteína exón de Pan5 SEQ ID NO: 3, Pan6 SEQ ID NO: 7, Pan7 SEQ ID NO: 11, SV1 SEQ ID NO: 26, SV25 SEQ ID NO: 31 o SV39 SEQ ID NO: 36, o un fragmento de la misma,

(b) Una proteína pentón de Pan5 SEQ ID NO: 2, Pan6 SEQ ID NO: 6, Pan7 SEQ ID NO: 10, SV1 SEQ ID NO: 25, SV25 SEQ ID NO: 30 o SV39 SEQ ID NO: 35;
(c) Una proteína fibra de Pan5 SEQ ID NO: 4, Pan6 SEQ ID NO: 8, Pan7 SEQ ID NO: 12, SV1 SEQ ID NO: 27 y SEQ ID NO: 28, SV25 SEQ ID NO: 32, y SEQ ID NO: 33 o SV39 SEQ ID NO: 37, o un fragmento de la misma.

9. Un serotipo de adenovirus artificial que comprende una proteína cápside de acuerdo con la reivindicación 8 o un fragmento de la misma.

10. El serotipo de adenovirus artificial de acuerdo con la reivindicación 9, en donde dicha cápside comprende un fragmento de la proteína exón seleccionada del grupo que consiste de Pan5 SEQ ID NO: 15, Pan7 SEQ ID NO: 16 y Pan7 SEQ ID NO: 17.

11. El serotipo de adenovirus artificial de acuerdo con la reivindicación 9 o 10, en donde dicha cápside comprende un fragmento de la proteína fibra seleccionada del grupo que consiste de Pan6 SEQ ID NO: 19, Pan7 SEQ ID NO: 20 y Pan5 SEQ ID NO: 21.

12. Una molécula de ácido nucleico que comprende una secuencia heteróloga que codifica una proteína de acuerdo con las reivindicaciones 3 u 8 o un serotipo de adenovirus artificial de acuerdo con cualquiera de las reivindicaciones 9 a 11.

13. Un vector recombinante que comprende una secuencia de adenovirus de simio de acuerdo con la reivindicación 2 o una molécula de ácido nucleico de acuerdo con la reivindicación 4 o 12 y un gen heterólogo unido operativamente a secuencias las cuales expresan directamente dicho gen en una célula huésped.

14. El vector recombinante de acuerdo con la reivindicación 13, que además comprende los elementos cis del adenovirus 5' y 3' necesarios para la replicación y encapsidación.

15. El vector recombinante de acuerdo con la reivindicación 13 o la reivindicación 14, en donde dicho vector es un virus.

16. El vector recombinante de acuerdo con cualquiera de las reivindicaciones 13 a 15 en donde dicho vector carece de todo o una parte del gen E1.

17. Un virus recombinante que comprende una proteína cápside de simio de acuerdo con la reivindicación 3 o un serotipo de adenovirus artificial de acuerdo con cualquiera de las reivindicaciones 9 a 11.

18. Una célula huésped que comprende una molécula de ácido nucleico de acuerdo con cualquiera de las reivindicaciones 4 a 6 o 12, un vector recombinante de acuerdo con cualquiera de las reivindicaciones 13 a 16, o un virus recombinante de acuerdo con la reivindicación 17.

145

19. La célula huésped de acuerdo con la reivindicación 18, en donde dicha célula huésped está transformada establemente con la molécula de ácido nucleico o el vector recombinante.

20. La célula huésped de acuerdo con la reivindicación 18 o la reivindicación 19, donde dicha célula huésped expresa uno o más productos génicos adenovirales de dicha molécula de ácido nucleico o vector recombinante, dichos productos génicos seleccionados del grupo que consiste de E1a, E1b, E2a, y E4 ORF6.

21. La célula huésped de acuerdo con cualquiera de las reivindicaciones 18 a 20, en donde dicha célula huésped está transformada establemente con una molécula de ácido nucleico que comprende las repeticiones de las terminales invertidas de adenovirus de simio.

22. Una composición que comprende un vector recombinante en un vehículo aceptable farmacéuticamente, dicho vector comprende una secuencia de adenovirus de simio de acuerdo con cualquiera de las reivindicaciones 1 o 2 y un gen heterólogo seleccionado unido operativamente a secuencias reguladoras las cuales expresan directamente dicho gen en una célula huésped.

23. Un método para la entrega de un gen heterólogo a una célula mamífera que comprende la introducción en dicha célula de una cantidad efectiva del vector del

145
146

de cualquiera de las reivindicaciones 13 a 16 o un virus de acuerdo con la reivindicación 17.

24. Un método para la administración repetida de un gen heterólogo a un mamífero que comprende las etapas de:

(a) Introducir en dicho mamífero un primer vector el cual comprende el gen heterólogo;

(b) Introducir en dicho mamífero un segundo vector el cual comprende el gen heterólogo;

Donde al menos el primer vector virus o el segundo vector virus es un virus de acuerdo con la reivindicación 17 y en donde el primero y segundo vector recombinantes son diferentes.

25. Un método para la producción de un producto génico seleccionado que comprende infectar una célula mamífera con el vector de cualquiera de las reivindicaciones 13 a 16 o un virus de acuerdo con la reivindicación 17, cultivando dichas células bajo condiciones adecuadas y recuperando de dicho cultivo celular el producto génico expresado.

26. Un método para la producir una respuesta inmune en un huésped mamífero contra un agente infeccioso que comprende la administración de dicho huésped en una cantidad efectiva del adenovirus recombinante de la reivindicación 17, en donde dicho gen heterólogo codifica una antígeno del agente infectivo.

146
147

27. El método de acuerdo con la reivindicación 26, que comprende la etapa de iniciación del huésped con una vacuna de ADN que comprende el gen heterólogo previamente a la administración del adenovirus recombinante.

28. Una composición que comprende una proteína cápside de adenovirus de simio de acuerdo con la reivindicación 8 unida a una molécula heteróloga para entregar a una célula huésped seleccionada.

29. Un método, para apuntar a una célula que actúa como blanco la cual tiene un receptor adenoviral, que comprende la entrega a un sujeto de una composición de acuerdo con la reivindicación 28.

LISTA DE SECUENCIAS

<110> The Trustees of the University of Pennsylvania
Wilson, James M.
Gao, Guangping
Roy, Soumitra

<120> ÁCIDO NUCLÉICO DE ADENOVIRUS DE SIMIO Y SECUENCIAS DE AMINOÁCIDOS, VECTORES QUE CONTIENEN LOS MISMOS Y MÉTODOS DE USO.

<130> UPN-02677PCT

<150> US 60/331,951
<151> 2001-11-21

<150> US 60/366,798
<151> 2002-03-22

<160> 39

<170> Patente en versión 3.1

<210> 1
<211> 36462
<212> DNA
<213> chimpanzee adenovirus serotype Pan5

<220>
<221> CDS
<222> (13898)..(15490)
<223> L2 Penton

<220>
<221> CDS
<222> (18315)..(21116)
<223> L3 Hexon

<220>
<221> CDS
<222> (32035)..(33372)
<223> L5 Fiber

<400> 1
catcatcaat aatatacctc aaacttttgg tgcgcgttaa tatgcaaatg aggtatttga 60

atttggggat gcggggcggt gattggctgc gggagcggcg accgttaggg gcggggcggg 120

tgacgttttg atgacgtggc cgtgaggcgg agccggtttg caagttctcg tgggaaaagt 180

gacgtcaaac gaggtgtggt ttgaacacgg aaatactcaa ttttcccgcg ctctctgaca 240

ggaaatgagg tgtttctggg cggatgcaag tgaaaacggg ccattttcgc gcgaaaactg 300

aatgaggaag tgaaaatctg agtaattccg cgtttatggc agggaggagt atttgccgag 360

ggccgagtag actttgaccg attacgtggg ggtttcgatt accgtatttt tcacctaaat 420

ttccgcgtac ggtgtcaaag tccggtgttt ttacgtaggt gtcagctgat cgccagggta 480

tttaaacctg cgctctctag tcaagaggcc actcttgagt gccagcgagt agagttttct 540

cctccgcgcc gcgagtcaga tctacacttt gaaagatgag gcacctgaga gacctgcccg 600

gtaatgtttt cctggctact gggaacgaga ttctggaact ggtggtggac gccatgatgg 660

gtgacgaccc tccggagccc cctaccccat ttgaagcgcc ttcgctgtac gatttgtatg 720

atctggaggt ggatgtgccc gagaacgacc ccaacgagga ggcggtgaat gatttgttta 780

gcgatgccgc gctgctggct gccgagcagg ctaatacgga ctctggctca gacagcgatt 840

cctctctcca taccccgaga cccggcagag gtgagaaaaa gatccccgag cttaaagggg 900

aagagctcga cctgcgctgc tatgaggaat gcttgcctcc gagcgatgat gaggaggacg 960

aggaggcgat tcgagctgca gcgaaccagg gagtgaaaac agcgagcgag ggctttagcc 1020

tggactgtcc tactctgccc ggacacggct gtaagtcttg tgaatttcat cgcatgaata 1080

ctggagataa gaatgtgatg tgtgccctgt gctatatgag agcttacaac cattgtgttt 1140

acagtaagtg tgattaactt tagctgggga ggcagagggt gactgggtgc tgactggttt 1200

atttatgtat atgtttttta tgtgtaggtc ccgtctctga cgtagatgag acccccacta 1260

cagagtgcat ttcatcaccc ccagaaattg gcgaggaacc gcccgaagat attattcata 1320

gaccagttgc agtgagagtc accgggcgta gagcagctgt ggagagtttg gatgacttgc 1380

tacagggtgg ggatgaacct ttggacttgt gtacccggaa acgccccagg cactaagtgc 1440

cacacatgtg tgtttactta aggtgatgtc agtatttata gggtgtggag tgcaataaaa 1500

tccgtgttga ctttaagtgc gtggtttatg actcaggggt ggggactgtg ggtatataag 1560

caggtgcaga cctgtgtggt cagttcagag caggactcat ggagatctgg acagtcttgg 1620

aagactttca ccagactaga cagctgctag agaactcatc ggagggagtc tcttacctgt 1680

ggagattctg cttcggtggg cctctagcta agctagtcta tagggccaag caggattata 1740

aggatcaatt tgaggatatt ttgagagagt gtcctggtat ttttgactct ctcaacttgg 1800

gccatcagtc tcactttaac cagagtattc tgagagccct tgacttttct actcctggca 1860

gaactaccgc cgcggtagcc ttttttgcct ttatccttga caaatggagt caagaaaccc 1920

atttcagcag ggattaccgt ctggactgct tagcagtagc tttgtggaga acatggaggt 1980

gccagcgcct gaatgcaatc tccggctact tgccagtaca gccggtagac acgctgagga 2040

tcctgagtct ccagtcaccc caggaacacc aacgccgcca gcagccgcag caggagcagc 2100

agcaagagga ggaccgagaa gagaacctga gagccggtct ggaccctccg gtggcggagg 2160

aggaggagta gctgacttgt ttcccgagct gcgccgggtg ctgactaggt cttccagtgg 2220

acgggagagg gggattaagc gggagaggca tgaggagact agccacagaa ctgaactgac 2280

tgtcagtctg atgagtcgca ggcgcccaga atcggtgtgg tggcatgagg tgcagtcgca 2340

ggggatagat gaggtctcag tgatgcatga gaaatattcc ctagaacaag tcaagacttg 2400

ttggttggag cccgaggatg attgggaggt agccatcagg aattatgcca agctggctct 2460

gaggccagac aagaagtaca agattaccaa actgattaat atcagaaatt cctgctacat 2520

ttcagggaat ggggccgagg tggagatcag tacccaggag agggtggcct tcagatgctg 2580

catgatgaat atgtacccgg gggtggtggg catggaggga gtcaccttta tgaacgcgag 2640

gttcaggggt gatgggtata atggggtggt ctttatggcc aacaccaagc tgacagtgca 2700

cggatgctcc ttctttggct tcaataacat gtgcattgag gcctggggca gtgtttcagt 2760

gaggggatgc agtttttcag ccaactggat gggggtcgtg ggcagaacca agagcatggt 2820

gtcagtgaag aaatgcctgt tcgagaggtg ccacctgggg gtgatgagcg agggcgaagc 2880

caaagtcaaa cactgcgcct ctaccgagac gggctgcttt gtactgatca agggcaatgc 2940

caaagtcaag cataatatga tctgtggggc ctcggatgag cgcggctacc agatgctgac 3000

ctgcgccggt gggaacagcc atatgctagc caccgtgcat gtggcctcgc acccccgcaa 3060

gacatggccc gagttcgagc acaacgtcat gacccgctgc aatgtgcacc tggggtcccg 3120

ccgaggcatg ttcatgccct accagtgcaa catgcaattt gtgaaggtgc tgctggagcc 3180

cgatgccatg tccagagtga gcctgacggg ggtgtttgac atgaatgtgg agctgtggaa 3240

aattctgaga tatgatgaat ccaagaccag gtgccgggcc tgcgaatgcg gaggcaagca 3300

cgccaggctt cagcccgtgt gtgtggaggt gacggaggac ctgcgacccg atcatttggt 3360

gttgtcctgc aacgggacgg agttcggctc cagcggggaa gaatctgact agagtgagta 3420

gtgtttggga ctgggtggga gcctgcatga tgggcagaat gactaaaatc tgtgtttttc 3480

tgcgcagcag catgagcgga agcgcctcct ttgagggagg ggtattcagc ccttatctga 3540

cggggcgtct cccctcctgg gcgggagtgc gtcagaatgt gatgggatcc acggtggacg 3600

gccggcccgt gcagcccgcg aactcttcaa ccctgaccta cgcgaccctg agctcctcgt 3660

ccgtggacgc agctgccgcc gcagctgctg cttccgccgc cagcgccgtg cgcggaatgg 3720

ccctgggcgc cggctactac agctctctgg tggccaactc gagttccacc aataatcccg 3780

ccagcctgaa cgaggagaag ctgctgctgc tgatggccca gctcgaggcc ctgacccagc 3840

gcctgggcga gctgacccag caggtggctc agctgcaggc ggagacgcgg gccgcggttg 3900

ccacggtgaa aaccaaataa aaaatgaatc aataaataaa cggagacggt tgttgatttt 3960

aacacagagt cttgaatctt tatttgattt ttcgcgcgcg gtaggccctg gaccaccggt 4020

ctcgatcatt gagcacccgg tggatctttt ccaggacccg gtagaggtgg gcttggatgt 4080

tgaggtacat gggcatgagc ccgtcccggg ggtggaggta gctccattgc agggcctcgt 4140

gctcgggggt ggtgttgtaa atcacccagt catagcaggg gcgcagggcg tggtgctgca 4200

cgatgtcctt gaggaggaga ctgatggcca cgggcagccc cttggtgtag gtgttgacga 4260

acctgttgag ctgggaggga tgcatgcggg gggagatgag atgcatcttg gcctggatct 4320

tgagattggc gatgttcccg cccagatccc gccgggggtt catgttgtgc aggaccacca 4380

gcacggtgta tccggtgcac ttggggaatt tgtcatgcaa cttggaaggg aaggcgtgaa 4440

agaatttgga gacgcccttg tgaccgccca ggttttccat gcactcatcc atgatgatgg 4500

cgatgggccc gtgggcggcg gcttgggcaa agacgtttcg ggggtcggac acatcgtagt 4560

tgtggtcctg ggtgagctcg tcataggcca ttttaatgaa tttggggcgg agggtgcccg 4620

actgggggac gaaggtgccc tcgatcccgg gggcgtagtt gccctcgcag atctgcatct 4680

cccaggcctt gagctcggag gggggatca tgtccacctg cggggcgatg aaaaaaacgg 4740

tttccggggc gggggagatg agctgggccg aaagcaggtt ccggagcagc tgggacttgc 4800

cgcagccggt ggggccgtag atgaccccga tgaccggctg caggtggtag ttgagggaga 4860

gacagctgcc gtcctcgcgg aggaggggg ccacctcgtt catcatctcg cgcacatgca 4920

tgttctcgcg cacgagttcc gccaggaggc gctcgccccc aagcgagagg agctcttgca 4980

gcgaggcgaa gtttttcagc ggcttgagcc cgtcggccat gggcattttg gagagggtct 5040

gttgcaagag ttccagacgg tcccagagct cggtgatgtg ctctagggca tctcgatcca 5100

gcagacctcc tcgtttcgcg ggttggggcg actgcgggag tagggcacca ggcgatgggc 5160

gtccagcgag gccagggtcc ggtccttcca ggggcgcagg gtccgcgtca gcgtggtctc 5220

cgtcacggtg aaggggtgcg cgccgggctg ggcgcttgcg agggtgcgct tcaggctcat 5280

ccggctggtc gagaaccgct cccggtcggc gccctgcgcg tcggccaggt agcaattgag 5340

catgagttcg tagttgagcg cctcggccgc gtggcccttg gcgcggagct taccttttgga 5400

agtgtgtccg cagacgggac agaggaggga cttgagggcg tagagcttgg gggcgaggaa 5460

gacggactcg ggggcgtagg cgtccgcgcc gcagctggcg cagacggtct cgcactccac 5520

gagccaggtg aggtctggcc ggtcggggtc aaaaacgagg tttcctccgt gctttttgat 5580

gcgtttctta cctctggtct ccatgagctc gtgtccccgc tgggtgacaa agaggctgtc 5640

cgtgtccccg tagaccgact ttatgggccg gtcctcgagc ggggtgccgc ggtcctcgtc 5700

gtagaggaac cccgcccact ccgagacgaa ggcccgggtc caggccagca cgaaggaggc 5760

cacgtgggag gggtagcggt cgttgtccac cagcgggtcc accttctcca gggtatgcaa 5820

gcacatgtcc ccctcgtcca catccaggaa ggtgattggc ttgtaagtgt aggccacgtg 5880

accgggggtc ccggccgggg gggtataaaa gggggcgggc ccctgctcgt cctcactgtc 5940

ttccggatcg ctgtccagga gcgccagctg ttggggtagg tattccctct cgaaggcggg 6000

catgacctcg gcactcaggt tgtcagtttc tagaaacgag gaggatttga tattgacggt 6060

gccgttggag acgcctttca tgagcccctc gtccatctgg tcagaaaaga cgatcttttt 6120

gttgtcgagc ttggtggcga aggagccgta gagggcgttg gagagcagct tggcgatgga 6180

gcgcatggtc tggttctttt ccttgtcggc gcgctccttg gcggcgatgt tgagctgcac 6240

gtactcgcgc gccacgcact tccattcggg gaagacggtg gtgagcttgt cgggcacgat 6300

tctgacccgc cagccgcggt tgtgcagggt gatgaggtcc acgctggtgg ccacctcgcc 6360

gcgcaggggc tcgttggtcc agcagaggcg cccgcccttg cgcgagcaga agggggggcag 6420

cgggtccagc atgagctcgt cgggggggtc ggcgtccacg gtgaagatgc cgggcaggag 6480

ctcggggtcg aagtagctga tgcaggtgcc cagatcgtcc agcgccgctt gccagtcgcg 6540

cacggccagc gcgcgctcgt aggggctgag gggcgtgccc cagggcatgg ggtgcgtgag 6600

cgcggaggcg tacatgccgc agatgtcgta gacgtagagg ggctcctcga ggacgccgat 6660

gtaggtgggg tagcagcgcc ccccgcggat gctggcgcgc acgtagtcgt acagctcgtg 6720

cgagggcgcg aggagcccgg tgccgaggtt ggagcgctgc ggcttttcgg cgcggtagac 6780

gatctggcgg aagatggcgt gggagttgga ggagatggtg ggcctctgga agatgttgaa 6840

gtgggcgtgg ggcagtccga ccgagtccct gatgaagtgg gcgtaggagt cctgcagctt 6900

ggcgacgagc tcggcggtga cgaggacgtc cagggcgcag tagtcgaggg tctcttggat 6960

gatgtcgtac ttgagctggc ccttctgctt ccacagctcg cggttgagaa ggaactcttc 7020

gcggtccttc cagtactctt cgaggggggaa cccgtcctga tcggcacggt aagagcccac 7080

catgtagaac tggttgacgg ccttgtaggc gcagcagccc ttctccacgg ggagggcgta 7140

agcttgcgcg gccttgcgca gggaggtgtg ggtgagggcg aaggtgtcgc gcaccatgac 7200

cttgaggaac tggtgcttga agtcgaggtc gtcgcagccg ccctgctccc agagctggaa 7260

gtccgtgcgc ttcttgtagg cggggttggg caaagcgaaa gtaacatcgt tgaagaggat 7320

cttgcccgcg cggggcatga agttgcgagt gatgcggaaa ggctggggca cctcggcccg 7380

gttgttgatg acctgggcgg cgaggacgat ctcgtcgaag ccgttgatgt tgtgcccgac 7440

gatgtagagt tccacgaatc gcgggcggcc cttgacgtgg ggcagcttct tgagctcgtc 7500

gtaggtgagc tcggcggggt cgctgaggcc gtgctgctcg agggcccagt cggcgaggtg 7560

ggggttggcg ccgaggaagg aagtccagag atccacggcc agggcggtct gcaagcggtc 7620

ccggtactga cggaactgct ggcccacggc catttttttcg ggggtgacgc agtagaaggt 7680

gcgggggtcg ccgtgccagc ggtcccactt gagctggagg gcgaggtcgt gggcgagctc 7740

gacgagcggc gggtccccgg agagtttcat gaccagcatg aaggggacga gctgcttgcc 7800

gaaggacccc atccaggtgt aggtttccac gtcgtaggtg aggaagagcc tttcggtgcg 7860

aggatgcgag ccgatgggga agaactggat ctcctgccac cagttggagg aatggctgtt 7920

gatgtgatgg aagtagaaat gccgacggcg cgccgagcac tcgtgcttgt gtttatacaa 7980

gcgtccgcag tgctcgcaac gctgcacggg atgcacgtgc tgcacgagct gtacctgggt 8040

tcctttgacg aggaatttca gtgggcagtg gagcgctggc ggctgcatct ggtgctgtac 8100

tacgtcctgg ccatcggcgt ggccatcgtc tgcctcgatg gtggtcatgc tgacgaggcc 8160

gcgcgggagg caggtccaga cctcggctcg gacgggtcgg agagcgagga cgagggcgcg 8220

caggccggag ctgtccaggg tcctgagacg ctgcggagtc aggtcagtgg gcagcggcgg 8280

cgcgcggttg acttgcagga gcttttccag ggcgcgcggg aggtccagat ggtacttgat 8340

ctccacggcg ccgttggtgg cgacgtccac ggcttgcagg gtcccgtgcc cctggggcgc 8400

caccaccgtg ccccgtttct tcttgggtgc tggcggcggc ggctccatgc ttagaagcgg 8460

cggcgaggac gcgcgccggg cggcaggggc ggctcggggc ccggaggcag gggcggcagg 8520

ggcacgtcgg cgccgcgcgc gggcaggttc tggtactgcg cccggagaag actggcgtga 8580

gcgacgacgc gacggttgac gtcctggatc tgacgcctct gggtgaaggc cacgggaccc 8640

gtgagtttga acctgaaaga gagttcgaca gaatcaatct cggtatcgtt gacggcggcc 8700

tgccgcagga tctcttgcac gtcgcccgag ttgtcctggt aggcgatctc ggtcatgaac 8760

tgctcgatct cctcctcctg aaggtctccg cgaccggcgc gctcgacggt ggccgcgagg 8820

tcgttggaga tgcggcccat gagctgcgag aaggcgttca tgccggcctc gttccagacg 8880

TS3
1541

```
cggctgtaga ccacggctcc gtcggggtcg cgcgcgcgca tgaccacctg ggcgaggttg   8940
agctcgacgt ggcgcgtgaa gaccgcgtag ttgcagaggc gctggtagag gtagttgagc   9000
gtggtggcga tgtgctcggt gacgaagaag tacatgatcc agcggcggag cggcatctcg   9060
ctgacgtcgc ccagggcttc caagcgctcc atggcctcgt agaagtccac ggcgaagttg   9120
aaaaactggg agttgcgcgc cgagacggtc aactcctcct ccagaagacg gatgagctcg   9180
gcgatggtgg cgcgcacctc gcgctcgaag gccccggggg gctcctcttc ttccatctcc   9240
tcctcctctt ccatctcctc cactaacatc tcttctactt cctcctcagg aggcggcggc   9300
gggggagggg ccctgcgtcg ccggcggcgc acgggcagac ggtcgatgaa gcgctcgatg   9360
gtctccccgc gccggcgacg catggtctcg gtgacggcgc gcccgtcctc gcggggccgc   9420
agcgtgaaga cgccgccgcg catctccagg tggccgccgg gggggtctcc gttgggcagg   9480
gagagggcgc tgacgatgca tcttatcaat tggcccgtag ggactccgcg caaggacctg   9540
agcgtctcga gatccacggg atccgaaaac cgctgaacga aggcttcgag ccagtcgcag   9600
tcgcaaggta ggctgagccc ggtttcttgt tcttcgggta tttggtcggg aggcgggcgg   9660
gcgatgctgc tggtgatgaa gttgaagtag gcggtcctga gacggcggat ggtggcgagg   9720
agcaccaggt ccttgggccc ggcttgctgg atgcgcagac ggtcggccat gccccaggcg   9780
tggtcctgac acctggcgag gtccttgtag tagtcctgca tgagccgctc cacgggcacc   9840
tcctcctcgc ccgcgcggcc gtgcatgcgc gtgagcccga acccgcgctg cggctggacg   9900
agcgccaggt cggcgacgac gcgctcggcg aggatggcct gctggatctg ggtgagggtg   9960
gtctggaagt cgtcgaagtc gacgaagcgg tggtaggctc cggtgttgat ggtgtaggag  10020
cagttggcca tgacggacca gttgacggtc tggtggccgg ggcgcacgag ctcgtggtac  10080
ttgaggcgcg agtaggcgcg cgtgtcgaag atgtagtcgt tgcaggtgcg cacgaggtac  10140
tggtatccga cgaggaagtg cggcggcggc tggcggtaga gcggccatcg ctcggtggcg  10200
ggggcgccgg gcgcgaggtc ctcgagcatg aggcggtggt agccgtagat gtacctggac  10260
atccaggtga tgccggcggc ggtggtggag gcgcgcggga actcgcggac gcggttccag  10320
atgttgcgca gcggcaggaa gtagttcatg gtggccgcgg tctggcccgt gaggcgcgcg  10380
cagtcgtgga tgctctagac atacgggcaa aaacgaaagc ggtcagcggc tcgactccgt  10440
ggcctggagg ctaagcgaac gggttgggct gcgcgtgtac cccggttcga gtccctgctc  10500
gaatcaggct ggagccgcag ctaacgtggt actggcactc ccgtctcgac ccaagcctgc  10560
taacgaaacc tccaggatac ggaggcgggt cgttttggcc attttcgtca ggccggaaat  10620
```

gaaactagta agcgcggaaa gcggccgtcc gcgatggctc gctgccgtag tctggagaaa 10680

gaatcgccag ggttgcgttg cggtgtgccc cggttcgagc ctcagcgctc ggcgccggcc 10740

ggattccgcg gctaacgtgg gcgtggctgc cccgtcgttt ccaagacccc ttagccagcc 10800

gacttctcca gttacggagc gagcccctct ttttcttgtg tttttgccag atgcatcccg 10860

tactgcggca gatgcgcccc caccctccac cacaaccgcc cctaccgcag cagcagcaac 10920

agccggcgct tctgcccccg ccccagcagc agcagccagc cactaccgcg gcggccgccg 10980

tgagcggagc cggcgttcag tatgacctgg ccttggaaga gggcgagggg ctggcgcggc 11040

tgggggcgtc gtcgccggag cggcacccgc gcgtgcagat gaaaagggac gctcgcgagg 11100

cctacgtgcc caagcagaac ctgttcagag acaggagcgg cgaggagccc gaggagatgc 11160

gcgcctcccg cttccacgcg gggcgggagc tgcggcgcgg cctggaccga aagcgggtgc 11220

tgagggacga ggatttcgag gcggacgagc tgacggggat cagccccgcg cgcgcgcacg 11280

tggccgcggc caacctggtc acggcgtacg agcagaccgt gaaggaggag agcaacttcc 11340

aaaaatcctt caacaaccac gtgcgcacgc tgatcgcgcg cgaggaggtg accctgggcc 11400

tgatgcacct gtgggacctg ctggaggcca tcgtgcagaa ccccacgagc aagccgctga 11460

cggcgcagct gtttctggtg gtgcagcaca gtcgggacaa cgagacgttc agggaggcgc 11520

tgctgaatat caccgagccc gagggccgct ggctcctgga cctggtgaac attctgcaga 11580

gcatcgtggt gcaggagcgc gggctgccgc tgtccgagaa gctggcggcc atcaacttct 11640

cggtgctgag cctgggcaag tactacgcta ggaagatcta caagaccccg tacgtgccca 11700

tagacaagga ggtgaagatc gacgggtttt acatgcgcat gaccctgaaa gtgctgaccc 11760

tgagcgacga tctgggggtg taccgcaacg acaggatgca ccgcgcggtg agcgccagcc 11820

gccggcgcga gctgagcgac caggagctga tgcacagcct gcagcgggcc ctgaccgggg 11880

ccgggaccga gggggagagc tactttgaca tgggcgcgga cctgcgctgg cagcctagcc 11940

gccgggcctt ggaagctgcc ggcggttccc cctacgtgga ggaggtggac gatgaggagg 12000

aggagggcga gtacctggaa gactgatggc gcgaccgtat ttttgctaga tgcagcaaca 12060

gccaccgccg cctcctgatc ccgcgatgcg ggcggcgctg cagagccagc cgtccggcat 12120

taactcctcg gacgattgga cccaggccat gcaacgcatc atggcgctga cgacccgcaa 12180

tcccgaagcc tttagacagc agcctcaggc caaccgactc tcggccatcc tggaggccgt 12240

ggtgccctcg cgctcgaacc ccacgcacga gaaggtgctg gccatcgtga acgcgctggt 12300

```
ggagaacaag gccatccgcg gcgacgaggc cgggctggtg tacaacgcgc tgctggagcg  12360

cgtggcccgc tacaacagca ccaacgtgca gacgaacctg gaccgcatgg tgaccgacgt  12420

gcgcgaggcg gtgtcgcagc gcgagcggtt ccaccgcgag tcgaacctgg gctccatggt  12480

ggcgctgaac gccttcctga gcacgcagcc cgccaacgtg ccccggggcc aggaggacta  12540

caccaacttc atcagcgcgc tgcggctgat ggtggccgag gtgccccaga gcgaggtgta  12600

ccagtcgggg ccggactact tcttccagac cagtcgccag ggcttgcaga ccgtgaacct  12660

gagccaggct ttcaagaact tgcagggact gtggggcgtg caggccccgg tcggggaccg  12720

cgcgacggtg tcgagcctgc tgacgccgaa ctcgcgcctg ctgctgctgc tggtggcgcc  12780

cttcacggac agcggcagcg tgagccgcga ctcgtacctg ggctacctgc ttaacctgta  12840

ccgcgaggcc atcgggcagg cgcacgtgga cgagcagacc taccaggaga tcacccacgt  12900

gagccgcgcg ctgggccagg aggacccggg caacctggag gccaccctga acttcctgct  12960

gaccaaccgg tcgcagaaga tcccgcccca gtacgcgctg agcaccgagg aggagcgcat  13020

cctgcgctac gtgcagcaga gcgtggggct gttcctgatg caggaggggg ccacgcccag  13080

cgccgcgctc gacatgaccg cgcgcaacat ggagcccagc atgtacgccc gcaaccgccc  13140

gttcatcaat aagctgatgg actacttgca tcgggcggcc gccatgaact cggactactt  13200

taccaacgcc atcttgaacc cgcactggct cccgccgccc gggttctaca cgggcgagta  13260

cgacatgccc gaccccaacg acgggttcct gtgggacgac gtggacagca gcgtgttctc  13320

gccgcgcccc accaccacca ccgtgtggaa gaaagagggc ggggaccggc ggccgtcctc  13380

ggcgctgtcc ggtcgcgcgg gtgctgccgc ggcggtgccc gaggccgcca gcccttcccc  13440

gagcctgccc ttttcgctga acagcgtgcg cagcagcgag ctgggtcggc tgacgcggcc  13500

gcgcctgctg ggcgaggagg agtacctgaa cgactccttg cttcggcccg agcgcgagaa  13560

gaacttcccc aataacggga tagagagcct ggtggacaag atgagccgct ggaagacgta  13620

cgcgcacgag cacagggacg agccccgagc tagcagcagc accggcgcca cccgtagacg  13680

ccagcggcac gacaggcagc ggggtctggt gtgggacgat gaggattccg ccgacgacag  13740

cagcgtgttg gacttgggtg ggagtggtgg tggtaacccg ttcgctcacc tgcgcccccg  13800

tatcgggcgc ctgatgtaag aatctgaaaa aataaaagac ggtactcacc aaggccatgg  13860

cgaccagcgt gcgttcttct ctgttgtttg tagtagt atg atg agg cgc gtg tac   13915
                                          Met Met Arg Arg Val Tyr
                                          1               5

ccg gag ggt cct cct ccc tcg tac gag agc gtg atg cag cag gcg gtg    13963
```

```
Pro Glu Gly Pro Pro Pro Ser Tyr Glu Ser Val Met Gln Gln Ala Val
            10                  15                  20

gcg gcg gcg atg cag ccc ccg ctg gag gcg cct tac gtg ccc ccg cgg    14011
Ala Ala Ala Met Gln Pro Pro Leu Glu Ala Pro Tyr Val Pro Pro Arg
        25                  30                  35

tac ctg gcg cct acg gag ggg cgg aac agc att cgt tac tcg gag ctg    14059
Tyr Leu Ala Pro Thr Glu Gly Arg Asn Ser Ile Arg Tyr Ser Glu Leu
    40                  45                  50

gca ccc ttg tac gat acc acc cgg ttg tac ctg gtg gac aac aag tcg    14107
Ala Pro Leu Tyr Asp Thr Thr Arg Leu Tyr Leu Val Asp Asn Lys Ser
55                  60                  65                  70

gcg gac atc gcc tcg ctg aac tac cag aac gac cac agc aac ttc ctg    14155
Ala Asp Ile Ala Ser Leu Asn Tyr Gln Asn Asp His Ser Asn Phe Leu
                75                  80                  85

acc acc gtg gtg cag aac aac gat ttc acc ccc acg gag gcc agc acc    14203
Thr Thr Val Val Gln Asn Asn Asp Phe Thr Pro Thr Glu Ala Ser Thr
            90                  95                  100

cag acc atc aac ttt gac gag cgc tcg cgg tgg ggc ggc cag ctg aaa    14251
Gln Thr Ile Asn Phe Asp Glu Arg Ser Arg Trp Gly Gly Gln Leu Lys
        105                 110                 115

acc atc atg cac acc aac atg ccc aac gtg aac gag ttc atg tac agc    14299
Thr Ile Met His Thr Asn Met Pro Asn Val Asn Glu Phe Met Tyr Ser
    120                 125                 130

aac aag ttc aag gcg cgg gtg atg gtc tcg cgc aag acc ccc aac ggg    14347
Asn Lys Phe Lys Ala Arg Val Met Val Ser Arg Lys Thr Pro Asn Gly
135                 140                 145                 150

gtc aca gta aca gat ggt agt cag gac gag ctg acc tac gag tgg gtg    14395
Val Thr Val Thr Asp Gly Ser Gln Asp Glu Leu Thr Tyr Glu Trp Val
                155                 160                 165

gag ttt gag ctg ccc gag ggc aac ttc tcg gtg acc atg acc atc gat    14443
Glu Phe Glu Leu Pro Glu Gly Asn Phe Ser Val Thr Met Thr Ile Asp
            170                 175                 180

ctg atg aac aac gcc atc atc gac aac tac ttg gcg gtg ggg cgg cag    14491
Leu Met Asn Asn Ala Ile Ile Asp Asn Tyr Leu Ala Val Gly Arg Gln
        185                 190                 195

aac ggg gtg ctg gag agc gac atc ggc gtg aag ttc gac acg cgc aac    14539
Asn Gly Val Leu Glu Ser Asp Ile Gly Val Lys Phe Asp Thr Arg Asn
    200                 205                 210

ttc cgg ctg ggc tgg gac ccc gtg acc gag ctg gtg atg ccg ggc gtg    14587
Phe Arg Leu Gly Trp Asp Pro Val Thr Glu Leu Val Met Pro Gly Val
215                 220                 225                 230

tac acc aac gag gcc ttc cac ccc gac atc gtc ctg ctg ccc ggc tgc    14635
Tyr Thr Asn Glu Ala Phe His Pro Asp Ile Val Leu Leu Pro Gly Cys
```

```
                235                 240                 245

ggc gtg gac ttc acc gag agc cgc ctc agc aac ctg ctg ggc atc cgc    14683
Gly Val Asp Phe Thr Glu Ser Arg Leu Ser Asn Leu Leu Gly Ile Arg
            250                 255                 260

aag cgg cag ccc ttc cag gag ggc ttc cag atc ctg tac gag gac ctg    14731
Lys Arg Gln Pro Phe Gln Glu Gly Phe Gln Ile Leu Tyr Glu Asp Leu
        265                 270                 275

gag ggg ggc aac atc ccc gcg ctg ctg gac gtg gac gcc tac gag aaa    14779
Glu Gly Gly Asn Ile Pro Ala Leu Leu Asp Val Asp Ala Tyr Glu Lys
    280                 285                 290

agc aag gag gat agc gcc gcc gcg gcg acc gca gcc gtg gcc acc gcc    14827
Ser Lys Glu Asp Ser Ala Ala Ala Ala Thr Ala Ala Val Ala Thr Ala
295                 300                 305                 310

tct acc gag gtg cgg ggc gat aat ttt gct agc gcc gcg aca ctg gca    14875
Ser Thr Glu Val Arg Gly Asp Asn Phe Ala Ser Ala Ala Thr Leu Ala
                315                 320                 325

gcg gcc gag gcg gct gaa acc gaa agt aag ata gtg atc cag ccg gtg    14923
Ala Ala Glu Ala Ala Glu Thr Glu Ser Lys Ile Val Ile Gln Pro Val
            330                 335                 340

gag aag gac agc aag gag agg agc tac aac gtg ctc gcg gac aag aaa    14971
Glu Lys Asp Ser Lys Glu Arg Ser Tyr Asn Val Leu Ala Asp Lys Lys
        345                 350                 355

aac acc gcc tac cgc agc tgg tac ctg gcc tac aac tac ggc gac ccc    15019
Asn Thr Ala Tyr Arg Ser Trp Tyr Leu Ala Tyr Asn Tyr Gly Asp Pro
    360                 365                 370

gag aag ggc gtg cgc tcc tgg acg ctg ctc acc acc tcg gac gtc acc    15067
Glu Lys Gly Val Arg Ser Trp Thr Leu Leu Thr Thr Ser Asp Val Thr
375                 380                 385                 390

tgc ggc gtg gag caa gtc tac tgg tcg ctg ccc gac atg atg caa gac    15115
Cys Gly Val Glu Gln Val Tyr Trp Ser Leu Pro Asp Met Met Gln Asp
                395                 400                 405

ccg gtc acc ttc cgc tcc acg cgt caa gtt agc aac tac ccg gtg gtg    15163
Pro Val Thr Phe Arg Ser Thr Arg Gln Val Ser Asn Tyr Pro Val Val
            410                 415                 420

ggc gcc gag ctc ctg ccc gtc tac tcc aag agc ttc ttc aac gag cag    15211
Gly Ala Glu Leu Leu Pro Val Tyr Ser Lys Ser Phe Phe Asn Glu Gln
        425                 430                 435

gcc gtc tac tcg cag cag ctg cgc gcc ttc acc tcg ctc acg cac gtc    15259
Ala Val Tyr Ser Gln Gln Leu Arg Ala Phe Thr Ser Leu Thr His Val
    440                 445                 450

ttc aac cgc ttc ccc gag aac cag atc ctc gtt cgc ccg ccc gcg ccc    15307
Phe Asn Arg Phe Pro Glu Asn Gln Ile Leu Val Arg Pro Pro Ala Pro
455                 460                 465                 470
```

```
acc att acc acc gtc agt gaa aac gtt cct gct ctc aca gat cac ggg    15355
Thr Ile Thr Thr Val Ser Glu Asn Val Pro Ala Leu Thr Asp His Gly
                475                 480                 485

acc ctg ccg ctg cgc agc agt atc cgg gga gtc cag cgc gtg acc gtc    15403
Thr Leu Pro Leu Arg Ser Ser Ile Arg Gly Val Gln Arg Val Thr Val
            490                 495                 500

act gac gcc aga cgc cgc acc tgc ccc tac gtc tac aag gcc ctg ggc    15451
Thr Asp Ala Arg Arg Arg Thr Cys Pro Tyr Val Tyr Lys Ala Leu Gly
        505                 510                 515

gta gtc gcg ccg cgc gtc ctc tcg agc cgc acc ttc taa aaaatgtcca     15500
Val Val Ala Pro Arg Val Leu Ser Ser Arg Thr Phe
    520                 525                 530

ttctcatctc gcccagtaat aacaccggtt ggggcctgcg cgcgcccagc aagatgtacg  15560

gaggcgctcg ccaacgctcc acgcaacacc ccgtgcgcgt gcgcgggcac ttccgcgctc  15620

cctggggcgc cctcaagggc cgcgtgcgct cgcgcaccac cgtcgacgac gtgatcgacc  15680

aggtggtggc cgacgcgcgc aactacacgc ccgccgccgc gcccgtctcc accgtggacg  15740

ccgtcatcga cagcgtggtg gccgacgcgc gccggtacgc ccgcgccaag agccggcggc  15800

ggcgcatcgc ccggcggcac cggagcaccc ccgccatgcg cgcggcgcga gccttgctgc  15860

gcagggccag gcgcacggga cgcagggcca tgctcagggc ggccagacgc gcggcctccg  15920

gcagcagcag cgccggcagg acccgcagac gcgcggccac ggcggcggcg gcggccatcg  15980

ccagcatgtc ccgcccgcgg cgcggcaacg tgtactgggt gcgcgacgcc gccaccggtg  16040

tgcgcgtgcc cgtgcgcacc cgccccccctc gcacttgaag atgctgactt cgcgatgttg  16100

atgtgtccca gcggcgagga ggatgtccaa gcgcaaattc aaggaagaga tgctccaggt  16160

catcgcgcct gagatctacg gcccggcggc ggtgaaggag gaaagaaagc cccgcaaact  16220

gaagcgggtc aaaaaggaca aaaaggagga ggaagatgtg gacggactgg tggagtttgt  16280

gcgcgagttc gccccccggc ggcgcgtgca gtggcgcggg cggaaagtga aaccggtgct  16340

gcgacccggc accacggtgg tcttcacgcc cggcgagcgt tccggctccg cctccaagcg  16400

ctcctacgac gaggtgtacg gggacgagga catcctcgag caggcggccg aacgtctggg  16460

cgagtttgct tacggcaagc gcagccgccc cgcgcccttg aaagaggagg cggtgtccat  16520

cccgctggac cacggcaacc ccacgccgag cctgaagccg gtgaccctgc agcaggtgct  16580

gcctggtgcg gcgccgcgcc ggggcttcaa gcgcgagggc ggcgaggatc tgtacccgac  16640

catgcagctg atggtgccca agcgccagaa gctggaggac gtgctggagc acatgaaggt  16700
```

```
ggaccccgag gtgcagcccg aggtcaaggt gcggcccatc aagcaggtgg ccccgggcct  16760
gggcgtgcag accgtggaca tcaagatccc cacggagccc atggaaacgc agaccgagcc  16820
cgtgaagccc agcaccagca ccatggaggt gcagacggat ccctggatgc cggcaccggc  16880
ttccaccacc cgccgaagac gcaagtacgg cgcggccagc ctgctgatgc ccaactacgc  16940
gctgcatcct tccatcatcc ccacgccggg ctaccgcggc acgcgcttct accgcggcta  17000
caccagcagc cgccgccgca agaccaccac ccgccgccgc cgtcgtcgca cccgccgcag  17060
cagcaccgcg acttccgccg ccgccctggt gcggagagtg taccgcagcg ggcgcgagcc  17120
tctgaccctg ccgcgcgcgc gctaccaccc gagcatcgcc atttaactac cgcctcctac  17180
ttgcagatat ggccctcaca tgccgcctcc gcgtccccat tacgggctac cgaggaagaa  17240
agccgcgccg tagaaggctg acggggaacg ggctgcgtcg ccatcaccac cggcggcggc  17300
gcgccatcag caagcggttg gggggaggct tcctgcccgc gctgatgccc atcatcgccg  17360
cggcgatcgg ggcgatcccc ggcatagctt ccgtggcggt gcaggcctct cagcgccact  17420
gagacacagc ttggaaaatt tgtaataaaa aatggactga cgctcctggt cctgtgatgt  17480
gtgtttttag atggaagaca tcaatttttc gtccctggca ccgcgacacg gcacgcggcc  17540
gtttatgggc acctggagcg acatcggcaa cagccaactg aacgggggcg ccttcaattg  17600
gagcagtctc tggagcgggc ttaagaattt cgggtccacg ctcaaaacct atggcaacaa  17660
ggcgtggaac agcagcacag ggcaggcgct gagggaaaag ctgaaagagc agaacttcca  17720
gcagaaggtg gtcgatggcc tggcctcggg catcaacggg gtggtggacc tggccaacca  17780
ggccgtgcag aaacagatca acagccgcct ggacgcggtc ccgcccgcgg ggtccgtgga  17840
gatgccccag gtggaggagg agctgcctcc cctggacaag cgcggcgaca agcgaccgcg  17900
tcccgacgcg gaggagacgc tgctgacgca cacggacgag ccgcccccgt acgaggaggc  17960
ggtgaaactg ggtctgccca ccacgcggcc cgtggcgcct ctggccaccg gggtgctgaa  18020
acccagcagc agcagcagcc agcccgcgac cctggacttg cctccgcctg cttcccgccc  18080
ctccacagtg gctaagcccc tgccgccggt ggccgtcgcg tcgcgcgccc cccgaggccg  18140
cccccaggcg aactggcaga gcactctgaa cagcatcgtg ggtctgggag tgcagagtgt  18200
gaagcgccgc cgctgctatt aaaagacact gtagcgctta acttgcttgt ctgtgtgtat  18260
atgtatgtcc gccgaccaga aggaggagga agaggcgcgt cgccgagttg caag atg     18317
                                                              Met

gcc acc cca tcg atg ctg ccc cag tgg gcg tac atg cac atc gcc gga     18365
```

```
Ala Thr Pro Ser Met Leu Pro Gln Trp Ala Tyr Met His Ile Ala Gly
            535                 540                 545

cag gac gct tcg gag tac ctg agt ccg ggt ctg gtg cag ttc gcc cgc    18413
Gln Asp Ala Ser Glu Tyr Leu Ser Pro Gly Leu Val Gln Phe Ala Arg
        550                 555                 560

gcc aca gac acc tac ttc agt ctg ggg aac aag ttt agg aac ccc acg    18461
Ala Thr Asp Thr Tyr Phe Ser Leu Gly Asn Lys Phe Arg Asn Pro Thr
    565                 570                 575

gtg gcg ccc acg cac gat gtg acc acc gac cgc agc cag cgg ctg acg    18509
Val Ala Pro Thr His Asp Val Thr Thr Asp Arg Ser Gln Arg Leu Thr
580                 585                 590                 595

ctg cgc ttc gtg ccc gtg gac cgc gag gac aac acc tac tcg tac aaa    18557
Leu Arg Phe Val Pro Val Asp Arg Glu Asp Asn Thr Tyr Ser Tyr Lys
                600                 605                 610

gtg cgc tac acg ctg gcc gtg ggc gac aac cgc gtg ctg gac atg gcc    18605
Val Arg Tyr Thr Leu Ala Val Gly Asp Asn Arg Val Leu Asp Met Ala
            615                 620                 625

agc acc tac ttt gac atc cgc ggc gtg ctg gat cgg ggc cct agc ttc    18653
Ser Thr Tyr Phe Asp Ile Arg Gly Val Leu Asp Arg Gly Pro Ser Phe
        630                 635                 640

aaa ccc tac tcc ggc acc gct tac aac agc ctg gct ccc aag gga gcg    18701
Lys Pro Tyr Ser Gly Thr Ala Tyr Asn Ser Leu Ala Pro Lys Gly Ala
    645                 650                 655

ccc aac act tgc cag tgg aca tat aaa gct gat ggt gat act ggt aca    18749
Pro Asn Thr Cys Gln Trp Thr Tyr Lys Ala Asp Gly Asp Thr Gly Thr
660                 665                 670                 675

gaa aaa acc tat aca tat gga aat gcg cct gtg caa ggc att agt att    18797
Glu Lys Thr Tyr Thr Tyr Gly Asn Ala Pro Val Gln Gly Ile Ser Ile
                680                 685                 690

aca aaa gat ggt att caa ctt gga act gac act gat gat cag ccc att    18845
Thr Lys Asp Gly Ile Gln Leu Gly Thr Asp Thr Asp Asp Gln Pro Ile
            695                 700                 705

tat gca gat aaa act tat caa cca gag cct caa gtg ggt gat gct gaa    18893
Tyr Ala Asp Lys Thr Tyr Gln Pro Glu Pro Gln Val Gly Asp Ala Glu
        710                 715                 720

tgg cat gac atc act ggt act gat gaa aaa tat gga ggc aga gct ctc    18941
Trp His Asp Ile Thr Gly Thr Asp Glu Lys Tyr Gly Gly Arg Ala Leu
    725                 730                 735

aag cct gac acc aaa atg aag ccc tgc tat ggt tct ttt gcc aag cct    18989
Lys Pro Asp Thr Lys Met Lys Pro Cys Tyr Gly Ser Phe Ala Lys Pro
740                 745                 750                 755

acc aat aaa gaa gga ggt cag gca aat gtg aaa acc gaa aca ggc ggt    19037
Thr Asn Lys Glu Gly Gly Gln Ala Asn Val Lys Thr Glu Thr Gly Gly
```

```
                760                 765                 770

acc aaa gaa tat gac att gac atg gca ttc ttc gat aat cga agt gca    19085
Thr Lys Glu Tyr Asp Ile Asp Met Ala Phe Phe Asp Asn Arg Ser Ala
            775                 780                 785

gct gcg gct ggc ctg gcc cca gaa att gtt ttg tat act gag aat gtg    19133
Ala Ala Ala Gly Leu Ala Pro Glu Ile Val Leu Tyr Thr Glu Asn Val
        790                 795                 800

gat ctg gaa act cca gat act cat att gta tac aag gcg ggc aca gat    19181
Asp Leu Glu Thr Pro Asp Thr His Ile Val Tyr Lys Ala Gly Thr Asp
    805                 810                 815

gac agc agc tct tct atc aat ttg ggt cag cag tcc atg ccc aac aga    19229
Asp Ser Ser Ser Ser Ile Asn Leu Gly Gln Gln Ser Met Pro Asn Arg
820                 825                 830                 835

ccc aac tac att ggc ttt aga gac aac ttt atc ggg ctc atg tac tac    19277
Pro Asn Tyr Ile Gly Phe Arg Asp Asn Phe Ile Gly Leu Met Tyr Tyr
                840                 845                 850

aac agc act ggc aac atg ggc gtg ctg gct ggt cag gcc tcc cag ctg    19325
Asn Ser Thr Gly Asn Met Gly Val Leu Ala Gly Gln Ala Ser Gln Leu
            855                 860                 865

aat gct gtg gtg gac ttg cag gac aga aac act gaa ctg tcc tac cag    19373
Asn Ala Val Val Asp Leu Gln Asp Arg Asn Thr Glu Leu Ser Tyr Gln
        870                 875                 880

ctc ttg ctt gac tct ctg ggc gac aga acc agg tat ttc agt atg tgg    19421
Leu Leu Leu Asp Ser Leu Gly Asp Arg Thr Arg Tyr Phe Ser Met Trp
    885                 890                 895

aat cag gcg gtg gac agc tat gac ccc gat gtg cgc att att gaa aat    19469
Asn Gln Ala Val Asp Ser Tyr Asp Pro Asp Val Arg Ile Ile Glu Asn
900                 905                 910                 915

cac ggt gtg gag gat gaa ctc cct aac tat tgc ttc ccc ctg gat gct    19517
His Gly Val Glu Asp Glu Leu Pro Asn Tyr Cys Phe Pro Leu Asp Ala
                920                 925                 930

gtg ggt aga act gat act tac cag gga att aag gcc aat ggt gct gat    19565
Val Gly Arg Thr Asp Thr Tyr Gln Gly Ile Lys Ala Asn Gly Ala Asp
            935                 940                 945

caa acc acc tgg acc aaa gat gat act gtt aat gat gct aat gaa ttg    19613
Gln Thr Thr Trp Thr Lys Asp Asp Thr Val Asn Asp Ala Asn Glu Leu
        950                 955                 960

ggc aag ggc aat cct ttc gcc atg gag atc aac atc cag gcc aac ctg    19661
Gly Lys Gly Asn Pro Phe Ala Met Glu Ile Asn Ile Gln Ala Asn Leu
    965                 970                 975

tgg cgg aac ttc ctc tac gcg aac gtg gcg ctg tac ctg ccc gac tcc    19709
Trp Arg Asn Phe Leu Tyr Ala Asn Val Ala Leu Tyr Leu Pro Asp Ser
980                 985                 990                 995
```

```
tac aag tac acg ccg  gcc aac atc acg ctg  ccg acc aac acc aac      19754
Tyr Lys Tyr Thr Pro  Ala Asn Ile Thr Leu  Pro Thr Asn Thr Asn
                1000                1005                1010

acc tac gat tac atg  aac ggc cgc gtg gtg  gcg ccc tcg ctg gtg      19799
Thr Tyr Asp Tyr Met  Asn Gly Arg Val Val  Ala Pro Ser Leu Val
                1015                1020                1025

gac gcc tac atc aac  atc ggg gcg cgc tgg  tcg ctg gac ccc atg      19844
Asp Ala Tyr Ile Asn  Ile Gly Ala Arg Trp  Ser Leu Asp Pro Met
                1030                1035                1040

gac aac gtc aac ccc  ttc aac cac cac cgc  aac gcg ggc ctg cgc      19889
Asp Asn Val Asn Pro  Phe Asn His His Arg  Asn Ala Gly Leu Arg
                1045                1050                1055

tac cgc tcc atg ctc  ctg ggc aac ggg cgc  tac gtg ccc ttc cac      19934
Tyr Arg Ser Met Leu  Leu Gly Asn Gly Arg  Tyr Val Pro Phe His
                1060                1065                1070

atc cag gtg ccc caa  aag ttc ttc gcc atc  aag agc ctc ctg ctc      19979
Ile Gln Val Pro Gln  Lys Phe Phe Ala Ile  Lys Ser Leu Leu Leu
                1075                1080                1085

ctg ccc ggg tcc tac  acc tac gag tgg aac  ttc cgc aag gac gtc      20024
Leu Pro Gly Ser Tyr  Thr Tyr Glu Trp Asn  Phe Arg Lys Asp Val
                1090                1095                1100

aac atg atc ctg cag  agc tcc ctc ggc aac  gac ctg cgc acg gac      20069
Asn Met Ile Leu Gln  Ser Ser Leu Gly Asn  Asp Leu Arg Thr Asp
                1105                1110                1115

ggg gcc tcc atc gcc  ttc acc agc atc aac  ctc tac gcc acc ttc      20114
Gly Ala Ser Ile Ala  Phe Thr Ser Ile Asn  Leu Tyr Ala Thr Phe
                1120                1125                1130

ttc ccc atg gcg cac  aac acc gcc tcc acg  ctc gag gcc atg ctg      20159
Phe Pro Met Ala His  Asn Thr Ala Ser Thr  Leu Glu Ala Met Leu
                1135                1140                1145

cgc aac gac acc aac  gac cag tcc ttc aac  gac tac ctc tcg gcg      20204
Arg Asn Asp Thr Asn  Asp Gln Ser Phe Asn  Asp Tyr Leu Ser Ala
                1150                1155                1160

gcc aac atg ctc tac  ccc atc ccg gcc aac  gcc acc aac gtg ccc      20249
Ala Asn Met Leu Tyr  Pro Ile Pro Ala Asn  Ala Thr Asn Val Pro
                1165                1170                1175

atc tcc atc ccc tcg  cgc aac tgg gcc gcc  ttc cgc gga tgg tcc      20294
Ile Ser Ile Pro Ser  Arg Asn Trp Ala Ala  Phe Arg Gly Trp Ser
                1180                1185                1190

ttc acg cgc ctc aag  acc cgc gag acg ccc  tcg ctc ggc tcc ggg      20339
Phe Thr Arg Leu Lys  Thr Arg Glu Thr Pro  Ser Leu Gly Ser Gly
                1195                1200                1205
```

```
ttc gac ccc tac ttc  gtc tac tcg ggc tcc  atc ccc tac ctc gac      20384
Phe Asp Pro Tyr Phe  Val Tyr Ser Gly Ser  Ile Pro Tyr Leu Asp
                1210                1215                1220

ggc acc ttc tac ctc  aac cac acc ttc aag  aag gtc tcc atc acc      20429
Gly Thr Phe Tyr Leu  Asn His Thr Phe Lys  Lys Val Ser Ile Thr
                1225                1230                1235

ttc gac tcc tcc gtc  agc tgg ccc ggc aac  gac cgc ctc ctg acg      20474
Phe Asp Ser Ser Val  Ser Trp Pro Gly Asn  Asp Arg Leu Leu Thr
                1240                1245                1250

ccc aac gag ttc gaa  atc aag cgc acc gtc  gac gga gag ggg tac      20519
Pro Asn Glu Phe Glu  Ile Lys Arg Thr Val  Asp Gly Glu Gly Tyr
                1255                1260                1265

aac gtg gcc cag tgc  aac atg acc aag gac  tgg ttc ctg gtc cag      20564
Asn Val Ala Gln Cys  Asn Met Thr Lys Asp  Trp Phe Leu Val Gln
                1270                1275                1280

atg ctg gcc cac tac  aac atc ggc tac cag  ggc ttc tac gtg ccc      20609
Met Leu Ala His Tyr  Asn Ile Gly Tyr Gln  Gly Phe Tyr Val Pro
                1285                1290                1295

gag ggc tac aag gac  cgc atg tac tcc ttc  ttc cgc aac ttc cag      20654
Glu Gly Tyr Lys Asp  Arg Met Tyr Ser Phe  Phe Arg Asn Phe Gln
                1300                1305                1310

ccc atg agc cgc cag  gtc gtg gac gag gtc  aac tac aag gac tac      20699
Pro Met Ser Arg Gln  Val Val Asp Glu Val  Asn Tyr Lys Asp Tyr
                1315                1320                1325

cag gcc gtc acc ctg  gcc tac cag cac aac  aac tcg ggc ttc gtc      20744
Gln Ala Val Thr Leu  Ala Tyr Gln His Asn  Asn Ser Gly Phe Val
                1330                1335                1340

ggc tac ctc gcg ccc  acc atg cgc cag gga  cag ccc tac ccc gcc      20789
Gly Tyr Leu Ala Pro  Thr Met Arg Gln Gly  Gln Pro Tyr Pro Ala
                1345                1350                1355

aac tac ccc tac ccg  ctc atc ggc aag agc  gcc gtc gcc agc gtc      20834
Asn Tyr Pro Tyr Pro  Leu Ile Gly Lys Ser  Ala Val Ala Ser Val
                1360                1365                1370

acc cag aaa aag ttc  ctc tgc gac cgg gtc  atg tgg cgc atc ccc      20879
Thr Gln Lys Lys Phe  Leu Cys Asp Arg Val  Met Trp Arg Ile Pro
                1375                1380                1385

ttc tcc agc aac ttc  atg tcc atg ggc gcg  ctc acc gac ctc ggc      20924
Phe Ser Ser Asn Phe  Met Ser Met Gly Ala  Leu Thr Asp Leu Gly
                1390                1395                1400

cag aac atg ctc tac  gcc aac tcc gcc cac  gcg cta gac atg aat      20969
Gln Asn Met Leu Tyr  Ala Asn Ser Ala His  Ala Leu Asp Met Asn
                1405                1410                1415

ttc gaa gtc gac ccc  atg gat gag tcc acc  ctt ctc tat gtt gtc      21014
```

```
Phe Glu Val Asp Pro  Met Asp Glu Ser Thr  Leu Leu Tyr Val Val
                1420                1425                1430

ttc gaa gtc ttc gac  gtc gtc cga gtg cac  cag ccc cac cgc ggc      21059
Phe Glu Val Phe Asp  Val Val Arg Val His  Gln Pro His Arg Gly
                1435                1440                1445

gtc atc gag gcc gtc  tac ctg cgc acg ccc  ttc tcg gcc ggc aac      21104
Val Ile Glu Ala Val  Tyr Leu Arg Thr Pro  Phe Ser Ala Gly Asn
                1450                1455                1460

gcc acc acc taa gccccgctct tgcttcttgc aagatgacgg cctgtgcggg        21156
Ala Thr Thr

ctccggcgag caggagctca gggccatcct ccgcgacctg ggctgcgggc cctgcttcct  21216

gggcaccttc gacaagcgct tcccgggatt catggccccg cacaagctgg cctgcgccat  21276

cgtcaacacg gccggccgcg agaccggggg cgagcactgg ctggccttcg cctggaaccc  21336

gcgctcccac acctgctacc tcttcgaccc cttcgggttc tcggacgagc gcctcaagca  21396

gatctaccag ttcgagtacg agggcctgct gcgccgcagc gccctggcca ccgaggaccg  21456

ctgcgtcacc ctggaaaagt ccacccagac cgtgcagggt ccgcgctcgg ccgcctgcgg  21516

gctcttctgc tgcatgttcc tgcacgcctt cgtgcactgg cccgaccgcc ccatggacaa  21576

gaaccccacc atgaacttgc tgacgggggt gcccaacggc atgctccagt cgccccaggt  21636

ggaacccacc ctgcgccgca accaggaggc gctctaccgc ttcctcaacg cccactccgc  21696

ctactttcgc tcccaccgcg cgcgcatcga gaaggccacc gccttcgacc gcatgaatca  21756

agacatgtaa accgtgtgtg tatgtgaatg ctttattcat aataaacagc acatgtttat  21816

gccacctttt ctgaggctct gactttattt agaaatcgaa ggggttctgc cggctctcgg  21876

cgtgccccgc gggcagggat acgttgcgga actggtactt gggcagccac ttgaactcgg  21936

ggatcagcag cttcggcacg gggaggtcgg ggaacgagtc gctccacagc ttgcgcgtga  21996

gttgcagggc gcccagcagg tcgggcgcgg agatcttgaa atcgcagttg ggacccgcgt  22056

tctgcgcgcg ggagttgcgg tacacggggt tgcagcactg gaacaccatc agggccgggt  22116

gcttcacgct cgccagcacc gtcgcgtcgg tgatgccctc cacgtccaga tcctcggcgt  22176

tggccatccc gaagggggtc atcttgcagg tctgccgccc catgctgggc acgcagccgg  22236

gcttgtggtt gcaatcgcag tgcaggggga tcagcatcat ctgggcctgc tcggagctca  22296

tgcccgggta catggccttc atgaaagcct ccagctggcg gaaggcctgc tgcgccttgc  22356

cgccctcggt gaagaagacc ccgcaggact tgctagagaa ctggttggtg gcgcagccgg  22416
```

cgtcgtgcac gcagcagcgc gcgtcgttgt tggccagctg caccacgctg cgcccccagc 22476

ggttctgggt gatcttggcc cggtcggggt tctccttcag cgcgcgctgc ccgttctcgc 22536

tcgccacatc catctcgatc gtgtgctcct tctggatcat cacggtcccg tgcaggcatc 22596

gcagcttgcc ctcggcctcg gtgcacccgt gcagccacag cgcgcagccg gtgcactccc 22656

agttcttgtg ggcgatctgg gagtgcgagt gcacgaagcc ctgcaggaag cggcccatca 22716

tcgtggtcag ggtcttgttg ctggtgaagg tcagcgggat gccgcggtgc tcctcgttca 22776

catacaggtg gcagatgcgg cggtacacct cgccctgctc gggcatcagc tggaaggcgg 22836

acttcaggtc gctctccacg cggtaccggt ccatcagcag cgtcatgact tccatgccct 22896

tctcccaggc cgagacgatc ggcaggctca gggggttctt caccgccgtt gtcatcttag 22956

tcgccgccgc tgaggtcagg gggtcgttct cgtccagggt ctcaaacact cgcttgccgt 23016

ccttctcggt gatgcgcacg gggggaaagc tgaagcccac ggccgccagc tcctcctcgg 23076

cctgcctttc gtcctcgctg tcctggctga tgtcttgcaa aggcacatgc ttggtcttgc 23136

ggggtttctt tttgggcggc agaggcggcg gcggagacgt gctgggcgag cgcgagttct 23196

cgctcaccac gactatttct tcttcttggc cgtcgtccga gaccacgcgg cggtaggcat 23256

gcctcttctg gggcagaggc ggaggcgacg ggctctcgcg gttcggcggg cggctggcag 23316

agcccccttcc gcgttcgggg gtgcgctcct ggcggcgctg ctctgactga cttcctccgc 23376

ggccggccat tgtgttctcc tagggagcaa caagcatgga gactcagcca tcgtcgccaa 23436

catcgccatc tgcccccgcc gccgccgacg agaaccagca gcagaatgaa agcttaaccg 23496

ccccgccgcc cagccccacc tccgacgccg ccgcggcccc agacatgcaa gagatggagg 23556

aatccatcga gattgacctg ggctacgtga cgcccgcgga gcacgaggag gagctggcag 23616

cgcgcttttc agccccggaa gagaaccacc aagagcagcc agagcaggaa gcagagagcg 23676

agcagcagca ggctgggctc gagcatggcg actacctgag cggggcagag gacgtgctca 23736

tcaagcatct ggcccgccaa tgcatcatcg tcaaggacgc gctgctcgac cgcgccgagg 23796

tgcccctcag cgtggcggag ctcagccgcg cctacgagcg caacctcttc tcgccgcgcg 23856

tgccccccaa gcgccagccc aacggcacct gcgagcccaa cccgcgcctc aacttctacc 23916

cggtcttcgc ggtgcccgag gccctggcca cctaccacct ctttttcaag aaccaaagga 23976

tccccgtctc ctgccgcgcc aaccgcaccc gcgccgacgc cctgctcaac ctgggtcccg 24036

gcgcccgcct acctgatatc gcctccttgg aagaggttcc caagatcttc gagggtctgg 24096

gcagcgacga gactcgggcc gcgaacgctc tgcaaggaag cggagaggag catgagcacc 24156

acagcgccct ggtggagttg gaaggcgaca acgcgcgcct ggcggtgctc aagcgcacgg 24216

tcgagctgac ccacttcgcc tacccggcgc tcaacctgcc ccccaaggtc atgagcgccg 24276

tcatggacca ggtgctcatc aagcgcgcct cgcccctctc ggatgaggac atgcaggacc 24336

ccgagagctc ggacgagggc aagcccgtgg tcagcgacga gcagctggcg cgctggctgg 24396

gagcgagtag caccccccag agcttggaag agcggcgcaa gctcatgatg gccgtggtcc 24456

tggtgaccgt ggagctggag tgtctgcgcc gcttcttcgc cgacgcagag accctgcgca 24516

aggtcgagga gaacctgcac tacctcttca ggcacgggtt tgtgcgccag gcctgcaaga 24576

tctccaacgt ggagctgacc aacctggtct cctacatggg catcctgcac gagaaccgcc 24636

tggggcagaa cgtgctgcac accaccctgc gcggggaggc ccgccgcgac tacatccgcg 24696

actgcgtcta cctgtacctc tgccacacct ggcagacggg catgggcgtg tggcagcagt 24756

gcctggagga gcagaacctg aaagagctct gcaagctcct gcagaagaac ctgaaggccc 24816

tgtggaccgg gttcgacgag cgcaccaccg cctcggacct ggccgacctc atcttccccg 24876

agcgcctgcg gctgacgctg cgcaacggac tgcccgactt tatgagtcaa agcatgttgc 24936

aaaactttcg ctctttcatc ctcgaacgct ccgggatcct gcccgccacc tgctccgcgc 24996

tgccctcgga cttcgtgccg ctgaccttcc gcgagtgccc cccgccgctc tggagccact 25056

gctacctgct gcgcctggcc aactacctgg cctaccactc ggacgtgatc gaggacgtca 25116

gcggcgaggg tctgctcgag tgccactgcc gctgcaacct ctgcacgccg caccgctccc 25176

tggcctgcaa cccccagctg ctgagcgaga cccagatcat cggcaccttc gagttgcaag 25236

gccccggcga gggcaagggg ggtctgaaac tcaccccggg gctgtggacc tcggcctact 25296

tgcgcaagtt cgtgcccgag gactaccatc ccttcgagat caggttctac gaggaccaat 25356

cccagccgcc caaggccgaa ctgtcggcct gcgtcatcac ccaggggcc atcctggccc 25416

aattgcaagc catccagaaa tcccgccaag aatttctgct gaaaaagggc cacggggtct 25476

acctggaccc ccagaccgga gaggagctca accccagctt cccccaggat gccccgagga 25536

agcagcaaga agctgaaagt ggagctgccg ccgccggagg atttggagga agactgggag 25596

agcagtcagg cagaggagga ggagatggaa gactgggaca gcactcaggc agaggaggac 25656

agcctgcaag acagtctgga agacgaggtg gaggaggagg cagaggaaga agcagccgcc 25716

gccagaccgt cgtcctcggc ggagaaagca agcagcacgg ataccatctc cgctccgggt 25776

cggggtcgcg gcgaccgggc ccacagtagg tgggacgaga ccgggcgctt cccgaacccc 25836

accacccaga ccggtaagaa ggagcggcag ggatacaagt cctggcgggg gcacaaaaac 25896

gccatcgtct cctgcttgca agcctgcggg ggcaacatct ccttcacccg ccgctacctg 25956

ctcttccacc gcggggtgaa cttccccccgc aacatcttgc attactaccg tcacctccac 26016

agcccctact actgtttcca agaagaggca gaaacccagc agcagcagaa aaccagcggc 26076

agcagcagct agaaaatcca cagcggcggc aggtggactg aggatcgcag cgaacgagcc 26136

ggcgcagacc cgggagctga ggaaccggat ctttcccacc ctctatgcca tcttccagca 26196

gagtcggggg caggagcagg aactgaaagt caagaaccgt tctctgcgct cgctcacccg 26256

cagttgtctg tatcacaaga gcgaagacca acttcagcgc actctcgagg acgccgaggc 26316

tctcttcaac aagtactgcg cgctcactct taaagagtag cccgcgcccg cccacacacg 26376

gaaaaaggcg ggaattacgt caccacctgc gcccttcgcc cgaccatcat catgagcaaa 26436

gagattccca cgccttacat gtggagctac cagccccaga tgggcctggc cgccggcgcc 26496

gcccaggact actccacccg catgaactgg ctcagcgccg ggcccgcgat gatctcacgg 26556

gtgaatgaca tccgcgcccg ccgaaaccag atactcctag aacagtcagc gatcaccgcc 26616

acgccccgcc atcaccttaa tccgcgtaat tggcccgccg ccctggtgta ccaggaaatt 26676

ccccagccca cgaccgtact acttccgcga gacgcccagg ccgaagtcca gctgactaac 26736

tcaggtgtcc agctggccgg cggcgccgcc ctgtgtcgtc accgccccgc tcagggtata 26796

aagcggctgg tgatccgagg cagaggcaca cagctcaacg acgaggtggt gagctcttcg 26856

ctgggtctgc gacctgacgg agtcttccaa ctcgccggat cggggagatc ttccttcacg 26916

cctcgtcagg ccgtcctgac tttggagagt tcgtcctcgc agccccgctc gggtggcatc 26976

ggcactctcc agttcgtgga ggagttcact ccctcggtct acttcaaccc cttctccggc 27036

tccccccggcc actacccgga cgagttcatc ccgaacttcg acgccatcag cgagtcggtg 27096

gacggctacg attgaatgtc ccatggtggc gcagctgacc tagctcggct tcgacacctg 27156

gaccactgcc gccgcttccg ctgcttcgct cgggatctcg ccgagtttgc ctactttgag 27216

ctgcccgagg agcaccctca gggcccggcc cacggagtgc ggatcatcgt cgaaggggggc 27276

ctcgactccc acctgcttcg gatcttcagc cagcgaccga tcctggtcga gcgcgagcaa 27336

ggacagaccc ttctgaccct gtactgcatc tgcaaccacc ccggcctgca tgaaagtctt 27396

tgttgtctgc tgtgtactga gtataataaa agctgagatc agcgactact ccggactcga 27456

ttgtggtgtt cctgctatca accggtccct gttcttcacc gggaacgaga ccgagctcca 27516

gcttcagtgt aagccccaca agaagtacct cacctggctg ttccagggct ccccgatcgc 27576

cgttgtcaac cactgcgaca acgacggagt cctgctgagc ggccccgcca accttacttt 27636

ttccacccgc agaagcaagc tccagctctt ccaacccttc ctccccggga cctatcagtg 27696

cgtctcggga ccctgccatc acaccttcca cctgatcccg aataccacag cgccgctccc 27756

cgctactaac aaccaaacta cccaccatcg ccaccgtcgc gacctttctg aatctaacac 27816

taccacccac accggaggtg agctccgagg tcgaccaacc tctgggattt actacggccc 27876

ctgggaggtg gtggggttaa tagcgctagg cctagttgtg ggtgggcttt tggctctctg 27936

ctacctatac ctcccttgct gttcgtactt agtggtgctg tgttgctggt ttaagaaatg 27996

gggaagatca ccctagtgag ctgcggtgcg ctggtggcgg tggtggtgtt ttcgattgtg 28056

ggactgggcg gcgcggctgt agtgaaggag aaggccgatc cctgcttgca tttcaatccc 28116

gacaattgcc agctgagttt tcagcccgat ggcaatcggt gcgcggtgct gatcaagtgc 28176

ggatgggaat gcgagaacgt gagaatcgag tacaataaca agactcggaa caatactctc 28236

gcgtccgtgt ggcagcccgg ggaccccgag tggtacaccg tctctgtccc cggtgctgac 28296

ggctccccgc gcaccgtgaa caatactttc atttttgcgc acatgtgcga cacggtcatg 28356

tggatgagca agcagtacga tatgtggccc cccacgaagg agaacatcgt ggtcttctcc 28416

atcgcttaca gcgcgtgcac ggcgctaatc accgctatcg tgtgcctgag cattcacatg 28476

ctcatcgcta ttcgccccag aaataatgcc gaaaaagaga aacagccata acacgttttt 28536

tcacacacct ttttcagacc atggcctctg ttaaattttt gcttttattt gccagtctca 28596

ttactgttat aagtaatgag aaactcacta tttacattgg cactaaccac actttagacg 28656

gaattccaaa atcctcatgg tattgctatt ttgatcaaga tccagactta actatagaac 28716

tgtgtggtaa caagggaaaa aatacaagca ttcatttaat taactttaat tgcggagaca 28776

atttgaaatt aattaatatc actaaagagt atggaggtat gtattactat gttgcagaaa 28836

ataacaacat gcagttttat gaagttactg taactaatcc caccacacct agaacaacaa 28896

caaccaccac cacaaaaact acacctgtta ccactatgca gctcactacc aataacattt 28956

ttgccatgcg tcaaatggtc aacaatagca ctcaacccac cccacccagt gaggaaattc 29016

ccaaatccat gattggcatt attgttgctg tagtggtgtg catgttgatc atcgccttgt 29076

gcatggtgta ctatgccttc tgctacagaa agcacagact gaacgacaag ctggaacact 29136

tactaagtgt tgaattttaa ttttttagaa ccatgaagat cctaggcctt ttaatttttt 29196

ctatcattac ctctgctcta tgcaattctg acaatgagga cgttactgtc gttgtcggaa 29256

ccaattatac actgaaaggt ccagcgaagg gtatgctttc gtggtattgc tggtttggaa 29316

ctgacgagca acagacagag ctctgcaatg ctcaaaaagg caaaacctca aattctaaaa 29376

tctctaatta tcaatgcaat ggcactgact tagtactgct caatgtcacg aaagcatatg 29436

ctggcagcta cacctgccct ggagatgata ctgagaacat gatttttac aaagtggaag 29496

tggttgatcc cactactcca cctccaccca ccacaactac tcacaccaca cacacagaac 29556

aaaccacagc agaggaggca gcaaagttag ccttgcaggt ccaagacagt tcatttgttg 29616

gcattacccc tacacctgat cagcggtgtc cggggctgct cgtcagcggc attgtcggtg 29676

tgctttcggg attagcagtc ataatcatct gcatgttcat ttttgcttgc tgctatagaa 29736

ggctttaccg acaaaaatca gacccactgc tgaacctcta tgtttaattt tttccagagc 29796

catgaaggca gttagcactc tagtttttg ttctttgatt ggcactgttt ttagtgttag 29856

ctttttgaaa caaatcaatg ttactgaggg ggaaaatgtg acactggtag gcgtagaggg 29916

tgctcaaaat accacctgga caaaattcca tctagatggg tggaaagaaa tttgcacctg 29976

gaatgtcagt acttatacat gtgaaggagt taatcttacc attgtcaatg tcagccaaat 30036

tcaaaagggt tggattaaag ggcaatctgt tagtgttagc aatagtgggt actataccca 30096

gcatactctt atctatgaca ttatagttat accactgcct acacctagcc cacctagcac 30156

taccacacag acaacccaca ctacacaaac aaccacatac agtacatcaa atcagcctac 30216

caccactaca acagcagagg ttgccagctc gtctggggtc cgagtggcat ttttgatgtt 30276

ggccccatct agcagtccca ctgctagtac caatgagcag actactgaat ttttgtccac 30336

tgtcgagagc cacaccacag ctacctcgag tgccttctct agcaccgcca atctatcctc 30396

gctttcctct acaccaatca gtcccgctac tactcctacc cccgctattc tccccactcc 30456

cctgaagcaa acagacggcg acatgcaatg gcagatcacc ctgctcattg tgatcgggtt 30516

ggtcatcctg gccgtgttgc tctactacat cttctgccgc cgcattccca acgcgcaccg 30576

caagccggcc tacaagccca tcgttgtcgg gcagccggag ccgcttcagg tggaaggggg 30636

tctaaggaat cttctcttct cttttacagt atggtgattg aattatgatt cctagacaaa 30696

tcttgatcac tattcttatc tgcctcctcc aagtctgtgc caccctcgct ctggtggcca 30756

acgccagtcc agactgtatt gggcccttcg cctcctacgt gctctttgcc ttcatcacct 30816

gcatctgctg ctgtagcata gtctgcctgc ttatcacctt cttccagttc attgactgga 30876

tctttgtgcg catcgcctac ctgcgccacc acccccagta ccgcgaccag cgagtggcgc 30936

ggctgctcag gatcctctga taagcatgcg ggctctgcta cttctcgcgc ttctgctgtt 30996

```
agtgctcccc cgtcccgtcg acccccggac ccccacccag tcccccgagg aggtccgcaa  31056

atgcaaattc caagaaccct ggaaattcct caaatgctac cgccaaaaat cagacatgca  31116

tcccagctgg atcatgatca ttgggatcgt gaacattctg gcctgcaccc tcatctcctt  31176

tgtgatttac ccctgctttg actttggttg gaactcgcca gaggcgctct atctcccgcc  31236

tgaacctgac acaccaccac agcaacctca ggcacacgca ctaccaccac caccacagcc  31296

taggccacaa tacatgccca tattagacta tgaggccgag ccacagcgac ccatgctccc  31356

cgctattagt tacttcaatc taaccggcgg agatgactga cccactggcc aacaacaacg  31416

tcaacgacct tctcctggac atggacggcc gcgcctcgga gcagcgactc gcccaacttc  31476

gcattcgcca gcagcaggag agagccgtca aggagctgca ggacggcata gccatccacc  31536

agtgcaagaa aggcatcttc tgcctggtga aacaggccaa gatctcctac gaggtcaccc  31596

agaccgacca tcgcctctcc tacgagctcc tgcagcagcg ccagaagttc acctgcctgg  31656

tcggagtcaa ccccatcgtc atcacccagc agtcgggcga taccaagggg tgcatccact  31716

gctcctgcga ctcccccgac tgcgtccaca ctctgatcaa gaccctctgc ggcctccgcg  31776

acctcctccc catgaactaa tcacccccct atccagtgaa ataaagatca tattgatgat  31836

ttgagtttaa taaaaataaa gaatcactta cttgaaatct gataccaggt ctctgtccat  31896

gttttctgcc aacaccactt cactcccctc ttcccagctc tggtactgca ggccccggcg  31956

ggctgcaaac ttcctccaca ccctgaaggg gatgtcaaat tcctcctgtc cctcaatctt  32016

cattttatct tctatcag atg tcc  aaa aag cgc gtc cgg  gtg gat gat gac  32067
                  Met Ser  Lys Lys Arg Val Arg  Val Asp Asp Asp
                      1465                1470

ttc  gac ccc gtc tac ccc  tac gat gca gac aac  gca ccg acc gtg     32112
Phe  Asp Pro Val Tyr Pro  Tyr Asp Ala Asp Asn  Ala Pro Thr Val
1475                 1480                1485

ccc  ttc atc aac ccc ccc  ttc gtc tct tca gat  gga ttc caa gag     32157
Pro  Phe Ile Asn Pro Pro  Phe Val Ser Ser Asp  Gly Phe Gln Glu
1490                 1495                1500

aag  ccc ctg ggg gtg ctg  tcc ctg cgt ctg gcc  gat ccc gtc acc     32202
Lys  Pro Leu Gly Val Leu  Ser Leu Arg Leu Ala  Asp Pro Val Thr
1505                 1510                1515

acc  aag aac ggg gaa atc  acc ctc aag ctg gga  gat ggg gtg gac     32247
Thr  Lys Asn Gly Glu Ile  Thr Leu Lys Leu Gly  Asp Gly Val Asp
1520                 1525                1530

ctc  gac tcc tcg gga aaa  ctc atc tcc aac acg  gcc acc aag gcc     32292
Leu  Asp Ser Ser Gly Lys  Leu Ile Ser Asn Thr  Ala Thr Lys Ala
```

```
1535                1540                1545

gcc gcc cct ctc agt ttt tcc aac aac acc att tcc ctt aac atg     32337
Ala Ala Pro Leu Ser Phe Ser Asn Asn Thr Ile Ser Leu Asn Met
1550                1555                1560

gat acc cct ttt tac aac aac aat gga aag tta ggc atg aaa gtc     32382
Asp Thr Pro Phe Tyr Asn Asn Asn Gly Lys Leu Gly Met Lys Val
1565                1570                1575

act gct cca ctg aag ata cta gac aca gac ttg cta aaa aca ctt     32427
Thr Ala Pro Leu Lys Ile Leu Asp Thr Asp Leu Leu Lys Thr Leu
1580                1585                1590

gtt gta gct tat gga caa ggt tta gga aca aac acc act ggt gcc     32472
Val Val Ala Tyr Gly Gln Gly Leu Gly Thr Asn Thr Thr Gly Ala
1595                1600                1605

ctt gtt gcc caa cta gca tcc cca ctt gct ttt gat agc aat agc     32517
Leu Val Ala Gln Leu Ala Ser Pro Leu Ala Phe Asp Ser Asn Ser
1610                1615                1620

aaa att gcc ctt aat tta ggc aat gga cca ttg aaa gtg gat gca     32562
Lys Ile Ala Leu Asn Leu Gly Asn Gly Pro Leu Lys Val Asp Ala
1625                1630                1635

aat aga ctg aac atc aat tgc aat aga gga ctc tat gtt act acc     32607
Asn Arg Leu Asn Ile Asn Cys Asn Arg Gly Leu Tyr Val Thr Thr
1640                1645                1650

aca aaa gat gca ctg gaa gcc aat ata agt tgg gct aat gct atg     32652
Thr Lys Asp Ala Leu Glu Ala Asn Ile Ser Trp Ala Asn Ala Met
1655                1660                1665

aca ttt ata gga aat gcc atg ggt gtc aat att gat aca caa aaa     32697
Thr Phe Ile Gly Asn Ala Met Gly Val Asn Ile Asp Thr Gln Lys
1670                1675                1680

ggc ttg caa ttt ggc acc act agt acc gtc gca gat gtt aaa aac     32742
Gly Leu Gln Phe Gly Thr Thr Ser Thr Val Ala Asp Val Lys Asn
1685                1690                1695

gct tac ccc ata caa atc aaa ctt gga gct ggt ctc aca ttt gac     32787
Ala Tyr Pro Ile Gln Ile Lys Leu Gly Ala Gly Leu Thr Phe Asp
1700                1705                1710

agc aca ggt gca att gtt gca tgg aac aaa gat gat gac aag ctt     32832
Ser Thr Gly Ala Ile Val Ala Trp Asn Lys Asp Asp Asp Lys Leu
1715                1720                1725

aca cta tgg acc aca gcc gac ccc tct cca aat tgt cac ata tat     32877
Thr Leu Trp Thr Thr Ala Asp Pro Ser Pro Asn Cys His Ile Tyr
1730                1735                1740

tct gaa aag gat gct aag ctt aca ctt tgc ttg aca aag tgt ggc     32922
Ser Glu Lys Asp Ala Lys Leu Thr Leu Cys Leu Thr Lys Cys Gly
1745                1750                1755
```

```
agt  cag att ctg ggc act  gtt tcc ctc ata gct  gtt gat act ggc     32967
Ser  Gln Ile Leu Gly Thr  Val Ser Leu Ile Ala  Val Asp Thr Gly
1760                 1765                 1770

agt  tta aat ccc ata aca  gga aca gta acc act  gct ctt gtc tca     33012
Ser  Leu Asn Pro Ile Thr  Gly Thr Val Thr Thr  Ala Leu Val Ser
1775                 1780                 1785

ctt  aaa ttc gat gca aat  gga gtt ttg caa agc  agc tca aca cta     33057
Leu  Lys Phe Asp Ala Asn  Gly Val Leu Gln Ser  Ser Ser Thr Leu
1790                 1795                 1800

gac  tca gac tat tgg aat  ttc aga cag gga gat  gtt aca cct gct     33102
Asp  Ser Asp Tyr Trp Asn  Phe Arg Gln Gly Asp  Val Thr Pro Ala
1805                 1810                 1815

gaa  gcc tat act aat gct  ata ggt ttc atg ccc  aat cta aaa gca     33147
Glu  Ala Tyr Thr Asn Ala  Ile Gly Phe Met Pro  Asn Leu Lys Ala
1820                 1825                 1830

tac  cct aaa aac aca agt  gga gct gca aaa agt  cac att gtt ggg     33192
Tyr  Pro Lys Asn Thr Ser  Gly Ala Ala Lys Ser  His Ile Val Gly
1835                 1840                 1845

aaa  gtg tac cta cat ggg  gat aca ggc aaa cca  ctg gac ctc att     33237
Lys  Val Tyr Leu His Gly  Asp Thr Gly Lys Pro  Leu Asp Leu Ile
1850                 1855                 1860

att  act ttc aat gaa aca  agt gat gaa tct tgc  act tac tgt att     33282
Ile  Thr Phe Asn Glu Thr  Ser Asp Glu Ser Cys  Thr Tyr Cys Ile
1865                 1870                 1875

aac  ttt caa tgg cag tgg  ggg gct gat caa tat  aaa aat gaa aca     33327
Asn  Phe Gln Trp Gln Trp  Gly Ala Asp Gln Tyr  Lys Asn Glu Thr
1880                 1885                 1890

ctt  gcc gtc agt tca ttc  acc ttt tcc tat att  gct aaa gaa taa     33372
Leu  Ala Val Ser Ser Phe  Thr Phe Ser Tyr Ile  Ala Lys Glu
1895                 1900                 1905

accccactct gtaccccatc tctgtctatg gaaaaaactc tgaaacacaa aataaaataa  33432

agttcaagtg ttttattgat tcaacagttt tacaggattc gagcagttat ttttcctcca  33492

ccctcccagg acatggaata caccaccctc tccccccgca cagccttgaa catctgaatg  33552

ccattggtga tggacatgct tttggtctcc acgttccaca cagtttcaga gcgagccagt  33612

ctcgggtcgg tcagggagat gaaaccctcc gggcactccc gcatctgcac ctcacagctc  33672

aacagctgag gattgtcctc ggtggtcggg atcacggtta tctggaagaa gcagaagagc  33732

ggcggtggga atcatagtcc gcgaacggga tcggccggtg gtgtcgcatc aggccccgca  33792

gcagtcgctg tcgccgccgc tccgtcaagc tgctgctcag ggggtccggg tccagggact  33852
```

ccctcagcat gatgcccacg gccctcagca tcagtcgtct ggtgcggcgg gcgcagcagc 33912

gcatgcggat ctcgctcagg tcgctgcagt acgtgcaaca caggaccacc aggttgttca 33972

acagtccata gttcaacacg ctccagccga aactcatcgc gggaaggatg ctacccacgt 34032

ggccgtcgta ccagatcctc aggtaaatca agtggcgccc cctccagaac acgctgccca 34092

tgtacatgat ctccttgggc atgtggcggt tcaccacctc ccggtaccac atcaccctct 34152

ggttgaacat gcagccccgg atgatcctgc ggaaccacag ggccagcacc gccccgcccg 34212

ccatgcagcg aagagacccc gggtcccgac aatggcaatg gaggacccac cgctcgtacc 34272

cgtggatcat ctgggagctg aacaagtcta tgttggcaca gcacaggcat atgctcatgc 34332

atctcttcag cactctcagc tcctcggggg tcaaaaccat atcccagggc acggggaact 34392

cttgcaggac agcgaacccc gcagaacagg gcaatcctcg cacataactt acattgtgca 34452

tggacagggt atcgcaatca ggcagcaccg ggtgatcctc caccagagaa gcgcgggtct 34512

cggtctcctc acagcgtggt aaggggccg gccgatacgg gtgatggcgg gacgcggctg 34572

atcgtgttcg cgaccgtgtt atgatgcagt tgctttcgga cattttcgta cttgctgtag 34632

cagaacctgg tccgggcgct gcacaccgat cgccggcggc ggtcccggcg cttggaacgc 34692

tcggtgttga agttgtaaaa cagccactct ctcagaccgt gcagcagatc tagggcctca 34752

ggagtgatga agatcccatc atgcctgatg gctctaatca catcgaccac cgtggaatgg 34812

gccagaccca gccagatgat gcaattttgt tgggtttcgg tgacggcggg ggagggaaga 34872

acaggaagaa ccatgattaa cttttaatcc aaacggtctc ggagcacttc aaaatgaaga 34932

tcgcggagat ggcacctctc gcccccgctg tgttggtgga aaataacagc caggtcaaag 34992

gtgatacggt tctcgagatg ttccacggtg gcttccagca aagcctccac gcgcacatcc 35052

agaaacaaga caatagcgaa agcgggaggg ttctctaatt cctcaatcat catgttacac 35112

tcctgcacca tccccagata attttcattt ttccagcctt gaatgattcg aactagttcc 35172

tgaggtaaat ccaagccagc catgataaag agctcgcgca gagcgccctc caccggcatt 35232

cttaagcaca ccctcataat tccaagatat tctgctcctg gttcacctgc agcagattga 35292

caagcggaat atcaaaatct ctgccgcgat ccctaagctc ctccctcagc aataactgta 35352

agtactcttt catatcctct ccgaaatttt tagccatagg accaccagga ataagattag 35412

ggcaagccac agtacagata aaccgaagtc ctccccagtg agcattgcca aatgcaagac 35472

tgctataagc atgctggcta gacccggtga tatcttccag ataactggac agaaaatcgc 35532

ccaggcaatt tttaagaaaa tcaacaaaag aaaaatcctc caggtgcacg tttagagcct 35592

cgggaacaac gatggagtaa atgcaagcgg tgcgttccag catggttagt tagctgatct 35652

gtagaaaaaa acaaaaatga acattaaacc atgctagcct ggcgaacagg tgggtaaatc 35712

gttctctcca gcaccaggca ggccacgggg tctccggcac gaccctcgta aaaattgtcg 35772

ctatgattga aaaccatcac agagagacgt tcccggtggc cggcgtgaat gattcgacaa 35832

gatgaataca cccccggaac attggcgtcc gcgagtgaaa aaaagcgccc aaggaagcaa 35892

taaggcacta caatgctcag tctcaagtcc agcaaagcga tgccatgcgg atgaagcaca 35952

aaattctcag gtgcgtacaa aatgtaatta ctccctctcct gcacaggcag caaagccccc 36012

gatccctcca ggtacacata caaagcctca gcgtccatag cttaccgagc agcagcacac 36072

aacaggcgca agagtcagag aaaggctgag ctctaacctg tccacccgct ctctgctcaa 36132

tatatagccc agatctacac tgacgtaaag gccaaagtct aaaaataccc gccaaataat 36192

cacacacgcc cagcacacgc ccagaaaccg gtgacacact caaaaaaata cgcgcacttc 36252

ctcaaacgcc caaactgccg tcatttccgg gttcccacgc tacgtcatca aaattcgact 36312

ttcaaattcc gtcgaccgtt aaaaacgtcg cccgccccgc ccctaacggt cgccgctccc 36372

gcagccaatc accgccccgc atccccaaat tcaaatacct catttgcata ttaacgcgca 36432

ccaaaagttt gaggtatatt attgatgatg 36462

<210> 2
<211> 530
<212> PRT
<213> chimpanzee adenovirus serotype Pan5

<400> 2

Met Met Arg Arg Val Tyr Pro Glu Gly Pro Pro Pro Ser Tyr Glu Ser
1 5 10 15

Val Met Gln Gln Ala Val Ala Ala Ala Met Gln Pro Pro Leu Glu Ala
20 25 30

Pro Tyr Val Pro Pro Arg Tyr Leu Ala Pro Thr Glu Gly Arg Asn Ser
35 40 45

Ile Arg Tyr Ser Glu Leu Ala Pro Leu Tyr Asp Thr Thr Arg Leu Tyr
50 55 60

Leu Val Asp Asn Lys Ser Ala Asp Ile Ala Ser Leu Asn Tyr Gln Asn
65 70 75 80

```
Asp His Ser Asn Phe Leu Thr Thr Val Val Gln Asn Asn Asp Phe Thr
                85                  90                  95

Pro Thr Glu Ala Ser Thr Gln Thr Ile Asn Phe Asp Glu Arg Ser Arg
            100                 105                 110

Trp Gly Gly Gln Leu Lys Thr Ile Met His Thr Asn Met Pro Asn Val
        115                 120                 125

Asn Glu Phe Met Tyr Ser Asn Lys Phe Lys Ala Arg Val Met Val Ser
    130                 135                 140

Arg Lys Thr Pro Asn Gly Val Thr Val Thr Asp Gly Ser Gln Asp Glu
145                 150                 155                 160

Leu Thr Tyr Glu Trp Val Glu Phe Glu Leu Pro Glu Gly Asn Phe Ser
                165                 170                 175

Val Thr Met Thr Ile Asp Leu Met Asn Asn Ala Ile Ile Asp Asn Tyr
            180                 185                 190

Leu Ala Val Gly Arg Gln Asn Gly Val Leu Glu Ser Asp Ile Gly Val
        195                 200                 205

Lys Phe Asp Thr Arg Asn Phe Arg Leu Gly Trp Asp Pro Val Thr Glu
    210                 215                 220

Leu Val Met Pro Gly Val Tyr Thr Asn Glu Ala Phe His Pro Asp Ile
225                 230                 235                 240

Val Leu Leu Pro Gly Cys Gly Val Asp Phe Thr Glu Ser Arg Leu Ser
                245                 250                 255

Asn Leu Leu Gly Ile Arg Lys Arg Gln Pro Phe Gln Glu Gly Phe Gln
            260                 265                 270

Ile Leu Tyr Glu Asp Leu Glu Gly Gly Asn Ile Pro Ala Leu Leu Asp
        275                 280                 285

Val Asp Ala Tyr Glu Lys Ser Lys Glu Asp Ser Ala Ala Ala Ala Thr
    290                 295                 300
```

```
Ala Ala Val Ala Thr Ala Ser Thr Glu Val Arg Gly Asp Asn Phe Ala
305                 310                 315                 320

Ser Ala Ala Thr Leu Ala Ala Ala Glu Ala Ala Glu Thr Glu Ser Lys
                325                 330                 335

Ile Val Ile Gln Pro Val Glu Lys Asp Ser Lys Glu Arg Ser Tyr Asn
            340                 345                 350

Val Leu Ala Asp Lys Lys Asn Thr Ala Tyr Arg Ser Trp Tyr Leu Ala
        355                 360                 365

Tyr Asn Tyr Gly Asp Pro Glu Lys Gly Val Arg Ser Trp Thr Leu Leu
    370                 375                 380

Thr Thr Ser Asp Val Thr Cys Gly Val Glu Gln Val Tyr Trp Ser Leu
385                 390                 395                 400

Pro Asp Met Met Gln Asp Pro Val Thr Phe Arg Ser Thr Arg Gln Val
                405                 410                 415

Ser Asn Tyr Pro Val Val Gly Ala Glu Leu Leu Pro Val Tyr Ser Lys
            420                 425                 430

Ser Phe Phe Asn Glu Gln Ala Val Tyr Ser Gln Gln Leu Arg Ala Phe
        435                 440                 445

Thr Ser Leu Thr His Val Phe Asn Arg Phe Pro Glu Asn Gln Ile Leu
    450                 455                 460

Val Arg Pro Pro Ala Pro Thr Ile Thr Thr Val Ser Glu Asn Val Pro
465                 470                 475                 480

Ala Leu Thr Asp His Gly Thr Leu Pro Leu Arg Ser Ser Ile Arg Gly
                485                 490                 495

Val Gln Arg Val Thr Val Thr Asp Ala Arg Arg Arg Thr Cys Pro Tyr
            500                 505                 510

Val Tyr Lys Ala Leu Gly Val Val Ala Pro Arg Val Leu Ser Ser Arg
        515                 520                 525
```

Thr Phe
530

<210> 3
<211> 933
<212> PRT
<213> chimpanzee adenovirus serotype Pan5

<400> 3

Met Ala Thr Pro Ser Met Leu Pro Gln Trp Ala Tyr Met His Ile Ala
1 5 10 15

Gly Gln Asp Ala Ser Glu Tyr Leu Ser Pro Gly Leu Val Gln Phe Ala
20 25 30

Arg Ala Thr Asp Thr Tyr Phe Ser Leu Gly Asn Lys Phe Arg Asn Pro
35 40 45

Thr Val Ala Pro Thr His Asp Val Thr Thr Asp Arg Ser Gln Arg Leu
50 55 60

Thr Leu Arg Phe Val Pro Val Asp Arg Glu Asp Asn Thr Tyr Ser Tyr
65 70 75 80

Lys Val Arg Tyr Thr Leu Ala Val Gly Asp Asn Arg Val Leu Asp Met
85 90 95

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

Phe Lys Pro Tyr Ser Gly Thr Ala Tyr Asn Ser Leu Ala Pro Lys Gly
115 120 125

Ala Pro Asn Thr Cys Gln Trp Thr Tyr Lys Ala Asp Gly Asp Thr Gly
130 135 140

Thr Glu Lys Thr Tyr Thr Tyr Gly Asn Ala Pro Val Gln Gly Ile Ser
145 150 155 160

Ile Thr Lys Asp Gly Ile Gln Leu Gly Thr Asp Thr Asp Asp Gln Pro
165 170 175

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

```
Glu Trp His Asp Ile Thr Gly Thr Asp Glu Lys Tyr Gly Gly Arg Ala
        195                 200                 205

Leu Lys Pro Asp Thr Lys Met Lys Pro Cys Tyr Gly Ser Phe Ala Lys
    210                 215                 220

Pro Thr Asn Lys Glu Gly Gly Gln Ala Asn Val Lys Thr Glu Thr Gly
225                 230                 235                 240

Gly Thr Lys Glu Tyr Asp Ile Asp Met Ala Phe Phe Asp Asn Arg Ser
                245                 250                 255

Ala Ala Ala Ala Gly Leu Ala Pro Glu Ile Val Leu Tyr Thr Glu Asn
            260                 265                 270

Val Asp Leu Glu Thr Pro Asp Thr His Ile Val Tyr Lys Ala Gly Thr
        275                 280                 285

Asp Asp Ser Ser Ser Ser Ile Asn Leu Gly Gln Gln Ser Met Pro Asn
    290                 295                 300

Arg Pro Asn Tyr Ile Gly Phe Arg Asp Asn Phe Ile Gly Leu Met Tyr
305                 310                 315                 320

Tyr Asn Ser Thr Gly Asn Met Gly Val Leu Ala Gly Gln Ala Ser Gln
                325                 330                 335

Leu Asn Ala Val Val Asp Leu Gln Asp Arg Asn Thr Glu Leu Ser Tyr
            340                 345                 350

Gln Leu Leu Leu Asp Ser Leu Gly Asp Arg Thr Arg Tyr Phe Ser Met
        355                 360                 365

Trp Asn Gln Ala Val Asp Ser Tyr Asp Pro Asp Val Arg Ile Ile Glu
    370                 375                 380

Asn His Gly Val Glu Asp Glu Leu Pro Asn Tyr Cys Phe Pro Leu Asp
385                 390                 395                 400

Ala Val Gly Arg Thr Asp Thr Tyr Gln Gly Ile Lys Ala Asn Gly Ala
                405                 410                 415
```

Asp Gln Thr Thr Trp Thr Lys Asp Asp Thr Val Asn Asp Ala Asn Glu
420 425 430

Leu Gly Lys Gly Asn Pro Phe Ala Met Glu Ile Asn Ile Gln Ala Asn
435 440 445

Leu Trp Arg Asn Phe Leu Tyr Ala Asn Val Ala Leu Tyr Leu Pro Asp
450 455 460

Ser Tyr Lys Tyr Thr Pro Ala Asn Ile Thr Leu Pro Thr Asn Thr Asn
465 470 475 480

Thr Tyr Asp Tyr Met Asn Gly Arg Val Val Ala Pro Ser Leu Val Asp
485 490 495

Ala Tyr Ile Asn Ile Gly Ala Arg Trp Ser Leu Asp Pro Met Asp Asn
500 505 510

Val Asn Pro Phe Asn His His Arg Asn Ala Gly Leu Arg Tyr Arg Ser
515 520 525

Met Leu Leu Gly Asn Gly Arg Tyr Val Pro Phe His Ile Gln Val Pro
530 535 540

Gln Lys Phe Phe Ala Ile Lys Ser Leu Leu Leu Leu Pro Gly Ser Tyr
545 550 555 560

Thr Tyr Glu Trp Asn Phe Arg Lys Asp Val Asn Met Ile Leu Gln Ser
565 570 575

Ser Leu Gly Asn Asp Leu Arg Thr Asp Gly Ala Ser Ile Ala Phe Thr
580 585 590

Ser Ile Asn Leu Tyr Ala Thr Phe Phe Pro Met Ala His Asn Thr Ala
595 600 605

Ser Thr Leu Glu Ala Met Leu Arg Asn Asp Thr Asn Asp Gln Ser Phe
610 615 620

Asn Asp Tyr Leu Ser Ala Ala Asn Met Leu Tyr Pro Ile Pro Ala Asn
625 630 635 640

```
Ala Thr Asn Val Pro Ile Ser Ile Pro Ser Arg Asn Trp Ala Ala Phe
                645                 650                 655

Arg Gly Trp Ser Phe Thr Arg Leu Lys Thr Arg Glu Thr Pro Ser Leu
            660                 665                 670

Gly Ser Gly Phe Asp Pro Tyr Phe Val Tyr Ser Gly Ser Ile Pro Tyr
        675                 680                 685

Leu Asp Gly Thr Phe Tyr Leu Asn His Thr Phe Lys Lys Val Ser Ile
    690                 695                 700

Thr Phe Asp Ser Ser Val Ser Trp Pro Gly Asn Asp Arg Leu Leu Thr
705                 710                 715                 720

Pro Asn Glu Phe Glu Ile Lys Arg Thr Val Asp Gly Glu Gly Tyr Asn
                725                 730                 735

Val Ala Gln Cys Asn Met Thr Lys Asp Trp Phe Leu Val Gln Met Leu
            740                 745                 750

Ala His Tyr Asn Ile Gly Tyr Gln Gly Phe Tyr Val Pro Glu Gly Tyr
        755                 760                 765

Lys Asp Arg Met Tyr Ser Phe Phe Arg Asn Phe Gln Pro Met Ser Arg
    770                 775                 780

Gln Val Val Asp Glu Val Asn Tyr Lys Asp Tyr Gln Ala Val Thr Leu
785                 790                 795                 800

Ala Tyr Gln His Asn Asn Ser Gly Phe Val Gly Tyr Leu Ala Pro Thr
                805                 810                 815

Met Arg Gln Gly Gln Pro Tyr Pro Ala Asn Tyr Pro Tyr Pro Leu Ile
            820                 825                 830

Gly Lys Ser Ala Val Ala Ser Val Thr Gln Lys Lys Phe Leu Cys Asp
        835                 840                 845

Arg Val Met Trp Arg Ile Pro Phe Ser Ser Asn Phe Met Ser Met Gly
    850                 855                 860

Ala Leu Thr Asp Leu Gly Gln Asn Met Leu Tyr Ala Asn Ser Ala His
```

865 870 875 880

Ala Leu Asp Met Asn Phe Glu Val Asp Pro Met Asp Glu Ser Thr Leu
885 890 895

Leu Tyr Val Val Phe Glu Val Phe Asp Val Val Arg Val His Gln Pro
900 905 910

His Arg Gly Val Ile Glu Ala Val Tyr Leu Arg Thr Pro Phe Ser Ala
915 920 925

Gly Asn Ala Thr Thr
930

<210> 4
<211> 445
<212> PRT
<213> chimpanzee adenovirus serotype Pan5

<400> 4

Met Ser Lys Lys Arg Val Arg Val Asp Asp Asp Phe Asp Pro Val Tyr
1 5 10 15

Pro Tyr Asp Ala Asp Asn Ala Pro Thr Val Pro Phe Ile Asn Pro Pro
20 25 30

Phe Val Ser Ser Asp Gly Phe Gln Glu Lys Pro Leu Gly Val Leu Ser
35 40 45

Leu Arg Leu Ala Asp Pro Val Thr Thr Lys Asn Gly Glu Ile Thr Leu
50 55 60

Lys Leu Gly Asp Gly Val Asp Leu Asp Ser Ser Gly Lys Leu Ile Ser
65 70 75 80

Asn Thr Ala Thr Lys Ala Ala Ala Pro Leu Ser Phe Ser Asn Asn Thr
85 90 95

Ile Ser Leu Asn Met Asp Thr Pro Phe Tyr Asn Asn Asn Gly Lys Leu
100 105 110

Gly Met Lys Val Thr Ala Pro Leu Lys Ile Leu Asp Thr Asp Leu Leu
115 120 125

```
Lys Thr Leu Val Val Ala Tyr Gly Gln Gly Leu Gly Thr Asn Thr Thr
    130                 135                 140

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

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

Asn Arg Leu Asn Ile Asn Cys Asn Arg Gly Leu Tyr Val Thr Thr Thr
            180                 185                 190

Lys Asp Ala Leu Glu Ala Asn Ile Ser Trp Ala Asn Ala Met Thr Phe
        195                 200                 205

Ile Gly Asn Ala Met Gly Val Asn Ile Asp Thr Gln Lys Gly Leu Gln
    210                 215                 220

Phe Gly Thr Thr Ser Thr Val Ala Asp Val Lys Asn Ala Tyr Pro Ile
225                 230                 235                 240

Gln Ile Lys Leu Gly Ala Gly Leu Thr Phe Asp Ser Thr Gly Ala Ile
                245                 250                 255

Val Ala Trp Asn Lys Asp Asp Asp Lys Leu Thr Leu Trp Thr Thr Ala
            260                 265                 270

Asp Pro Ser Pro Asn Cys His Ile Tyr Ser Glu Lys Asp Ala Lys Leu
        275                 280                 285

Thr Leu Cys Leu Thr Lys Cys Gly Ser Gln Ile Leu Gly Thr Val Ser
    290                 295                 300

Leu Ile Ala Val Asp Thr Gly Ser Leu Asn Pro Ile Thr Gly Thr Val
305                 310                 315                 320

Thr Thr Ala Leu Val Ser Leu Lys Phe Asp Ala Asn Gly Val Leu Gln
                325                 330                 335

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

Val Thr Pro Ala Glu Ala Tyr Thr Asn Ala Ile Gly Phe Met Pro Asn
355 360 365

Leu Lys Ala Tyr Pro Lys Asn Thr Ser Gly Ala Ala Lys Ser His Ile
370 375 380

Val Gly Lys Val Tyr Leu His Gly Asp Thr Gly Lys Pro Leu Asp Leu
385 390 395 400

Ile Ile Thr Phe Asn Glu Thr Ser Asp Glu Ser Cys Thr Tyr Cys Ile
405 410 415

Asn Phe Gln Trp Gln Trp Gly Ala Asp Gln Tyr Lys Asn Glu Thr Leu
420 425 430

Ala Val Ser Ser Phe Thr Phe Ser Tyr Ile Ala Lys Glu
435 440 445

<210> 5
<211> 36604
<212> DNA
<213> chimpanzee adenovirus serotype Pan6

<220>
<221> CDS
<222> (13878)..(15467)
<223> L2 Penton

<220>
<221> CDS
<222> (18284)..(21112)
<223> L3 Hexon

<220>
<221> CDS
<222> (32162)..(33493)
<223> L5 Fiber

<400> 5
catcatcaat aatatacctc aaacttttgg tgcgcgttaa tatgcaaatg agctgtttga 60

atttggggag ggaggaaggt gattggctgc gggagcggcg accgttaggg gcggggcggg 120

tgacgttttg atgacgtggc tatgaggcgg agccggtttg caagttctcg tgggaaaagt 180

gacgtcaaac gaggtgtggt ttgaacacgg aaatactcaa ttttcccgcg ctctctgaca 240

ggaaatgagg tgtttctggg cggatgcaag tgaaaacggg ccattttcgc gcgaaaactg 300

aatgaggaag tgaaaatctg agtaatttcg cgtttatggc agggaggagt atttgccgag 360

ggccgagtag actttgaccg attacgtggg ggtttcgatt accgtatttt tcacctaaat 420

ttccgcgtac ggtgtcaaag tccggtgttt ttacgtaggc gtcagctgat cgccagggta 480

tttaaacctg cgctctctag tcaagaggcc actcttgagt gccagcgagt agagttttct 540

cctccgcgcc gcgagtcaga tctacacttt gaaagatgag gcacctgaga gacctgcccg 600

gtaatgtttt cctggctact gggaacgaga ttctggaatt ggtggtggac gccatgatgg 660

gtgacgaccc tccagagccc cctaccccat ttgaggcgcc ttcgctgtac gatttgtatg 720

atctggaggt ggatgtgccc gagagcgacc ctaacgagga ggcggtgaat gatttgttta 780

gcgatgccgc gctgctggct gccgagcagg ctaatacgga ctctggctca gacagcgatt 840

cctctctcca taccccgaga cccggcagag gtgagaaaaa gatccccgag cttaaagggg 900

aagagctcga cctgcgctgc tatgaggaat gcttgcctcc gagcgatgat gaggaggacg 960

aggaggcgat tcgagctgcg gtgaaccagg gagtgaaaac tgcgggcgag agctttagcc 1020

tggactgtcc tactctgccc ggacacggct gtaagtcttg tgaatttcat cgcatgaata 1080

ctggagataa gaatgtgatg tgtgccctgt gctatatgag agcttacaac cattgtgttt 1140

acagtaagtg tgattaactt tagttgggaa ggcagagggt gactgggtgc tgactggttt 1200

atttatgtat atgtttttt atgtgtaggt cccgtctctg acgtagatga gacccccact 1260

tcagagtgca tttcatcacc cccagaaatt ggcgaggaac cgcccgaaga tattattcat 1320

agaccagttg cagtgagagt caccgggcgg agagcagctg tggagagttt ggatgacttg 1380

ctacagggtg gggatgaacc tttggacttg tgtacccgga aacgccccag gcactaagtg 1440

ccacacatgt gtgtttactt aaggtgatgt cagtatttat agggtgtgga gtgcaataaa 1500

atccgtgttg actttaagtg cgtgttttat gactcagggg tggggactgt gggtatataa 1560

gcaggtgcag acctgtgtgg tcagttcaga gcaggactca tggagatctg gactgtcttg 1620

gaagactttc accagactag acagttgcta gagaactcat cggagggagt ctcttacctg 1680

tggagattct gcttcggtgg gcctctagct aagctagtct atagggccaa acaggattat 1740

aaggaacaat ttgaggatat tttgagagag tgtcctggta tttttgactc tctcaacttg 1800

ggccatcagt ctcactttaa ccagagtatt ctgagagccc ttgacttttc tactcctggc 1860

agaactaccg ccgcggtagc ctttttttgcc tttattcttg acaaatggag tcaagaaacc 1920

catttcagca gggattaccg tctggactgc ttagcagtag ctttgtggag aacatggagg 1980

tgccagcgcc tgaatgcaat ctccggctac ttgccagtac agccggtaga cacgctgagg 2040

atcctgagtc tccagtcacc ccaggaacac caacgccgcc agcagccgca gcaggagcag 2100

cagcaagagg aggaccgaga agagaacccg agagccggtc tggaccctcc ggtggcggag 2160

gaggaggagt agctgacttg tttcccgagc tgcgccgggt gctgactagg tcttccagtg 2220

gacgggagag ggggattaag cgggagaggc atgaggagac tagccacaga actgaactga 2280

ctgtcagtct gatgagccgc aggcgcccag aatcggtgtg gtggcatgag gtgcagtcgc 2340

aggggataga tgaggtctcg gtgatgcatg agaaatattc cctagaacaa gtcaagactt 2400

gttggttgga gcccgaggat gattgggagg tagccatcag gaattatgcc aagctggctc 2460

tgaagccaga caagaagtac aagattacca aactgattaa tatcagaaat tcctgctaca 2520

tttcagggaa tggggccgag gtggagatca gtacccagga gagggtggcc ttcagatgtt 2580

gtatgatgaa tatgtacccg ggggtggtgg gcatggaggg agtcaccttt atgaacacga 2640

ggttcagggg tgatgggtat aatggggtgg tctttatggc caacaccaag ctgacagtgc 2700

acggatgctc cttctttggc ttcaataaca tgtgcatcga ggcctggggc agtgtttcag 2760

tgaggggatg cagcttttca gccaactgga tgggggtcgt gggcagaacc aagagcaagg 2820

tgtcagtgaa gaaatgcctg ttcgagaggt gccacctggg ggtgatgagc gagggcgaag 2880

ccaaagtcaa acactgcgcc tctaccgaga cgggctgctt tgtgctgatc aagggcaatg 2940

cccaagtcaa gcataacatg atctgtgggg cctcggatga gcgcggctac cagatgctga 3000

cctgcgccgg tgggaacagc catatgctgg ccaccgtgca tgtggcctcg caccccgca 3060

agacatggcc cgagttcgag cacaacgtca tgacccgctg caatgtgcac ctgggctccc 3120

gccgaggcat gttcatgccc taccagtgca acatgcaatt tgtgaaggtg ctgctggagc 3180

ccgatgccat gtccagagtg agcctgacgg gggtgtttga catgaatgtg gagctgtgga 3240

aaattctgag atatgatgaa tccaagacca ggtgccgggc ctgcgaatgc ggaggcaagc 3300

acgccaggct tcagcccgtg tgtgtggagg tgacggagga cctgcgaccc gatcatttgg 3360

tgttgtcctg caacgggacg gagttcggct ccagcgggga agaatctgac tagagtgagt 3420

agtgtttggg gctgggtgtg agcctgcatg aggggcagaa tgactaaaat ctgtggtttt 3480

ctgtgtgttg cagcagcatg agcggaagcg cctcctttga gggaggggta ttcagccctt 3540

atctgacggg gcgtctcccc tcctgggcgg gagtgcgtca gaatgtgatg ggatccacgg 3600

tggacggccg gcccgtgcag cccgcgaact cttcaaccct gacctacgcg accctgagct 3660

cctcgtccgt ggacgcagct gccgccgcag ctgctgcttc cgccgccagc gccgtgcgcg 3720

gaatggccct gggcgccggc tactacagct ctctggtggc caactcgagt tccaccaata 3780

atcccgccag cctgaacgag gagaagctgc tgctgctgat ggcccagctc gaggccctga 3840

cccagcgcct gggcgagctg acccagcagg tggctcagct gcaggcggag acgcgggccg 3900

cggttgccac ggtgaaaacc aaataaaaaa tgaatcaata aataaacgga gacggttgtt 3960

gattttaaca cagagtcttg aatctttatt tgatttttcg cgcgcggtag gccctggacc 4020

accggtctcg atcattgagc acccggtgga tcttttccag gacccggtag aggtgggctt 4080

ggatgttgag gtacatgggc atgagcccgt cccgggggtg gaggtagctc cattgcaggg 4140

cctcgtgctc ggggatggtg ttgtaaatca cccagtcata gcaggggcgc agggcgtggt 4200

gctgcacgat gtccttgagg aggagactga tggccacggg cagccccttg gtgtaggtgt 4260

tgacgaacct gttgagctgg gagggatgca tgcgggggga gatgagatgc atcttggcct 4320

ggatcttgag attggcgatg ttcccgccca gatcccgccg ggggttcatg ttgtgcagga 4380

ccaccagcac ggtgtatccg gtgcacttgg ggaatttgtc atgcaacttg gaagggaagg 4440

cgtgaaagaa tttggagacg cccttgtgac cgcccaggtt ttccatgcac tcatccatga 4500

tgatggcgat gggcccgtgg gcggcggcct gggcaaagac gtttcggggg tcggacacat 4560

cgtagttgtg gtcctgggtg agctcgtcat aggccatttt aatgaatttg gggcggaggg 4620

tgcccgactg ggggacgaag gtgccctcga tcccgggggc gtagttgccc tcgcagatct 4680

gcatctccca ggccttgagc tcggaggggg ggatcatgtc cacctgcggg gcgatgaaaa 4740

aaacggtttc cggggcgggg gagatgagct gggccgaaag caggttccgg agcagctggg 4800

acttgccgca accggtgggg ccgtagatga ccccgatgac cggctgcagg tggtagttga 4860

gggagagaca gctgccgtcc tcgcggagga ggggggccac ctcgttcatc atctcgcgca 4920

catgcatgtt ctcgcgcacg agttccgcca ggaggcgctc gccccccagc gagaggagct 4980

cttgcagcga ggcgaagttt ttcagcggct tgagtccgtc ggccatgggc attttggaga 5040

gggtctgttg caagagttcc agacggtccc agagctcggt gatgtgctct agggcatctc 5100

gatccagcag acctcctcgt ttcgcgggtt ggggcgactg cgggagtagg gcaccaggcg 5160

atgggcgtcc agcgaggcca gggtccggtc cttccagggc cgcagggtcc gcgtcagcgt 5220

ggtctccgtc acggtgaagg ggtgcgcgcc gggctgggcg cttgcgaggg tgcgcttcag 5280

gctcatccgg ctggtcgaga accgctcccg gtcggcgccc tgcgcgtcgg ccaggtagca 5340

attgagcatg agttcgtagt tgagcgcctc ggccgcgtgg cccttggcgc ggagcttacc 5400

tttggaagtg tgtccgcaga cgggacagag gagggacttg agggcgtaga gcttgggggc 5460

gaggaagacg gactcggggg cgtaggcgtc cgcgccgcag ctggcgcaga cggtctcgca 5520

ctccacgagc caggtgaggt cggggcggtt ggggtcaaaa acgaggtttc ctccgtgctt 5580

tttgatgcgt ttcttacctc tggtctccat gagctcgtgt ccccgctggg tgacaaagag 5640

gctgtccgtg tccccgtaga ccgactttat gggccggtcc tcgagcgggg tgccgcggtc 5700

ctcgtcgtag aggaaccccg cccactccga gacgaaggcc cgggtccagg ccagcacgaa 5760

ggaggccacg tgggaggggt agcggtcgtt gtccaccagc gggtccacct tctccagggt 5820

atgcaagcac atgtcccccc cgtccacatc caggaaggtg attggcttgt aagtgtaggc 5880

cacgtgaccg ggggtcccgg ccgggggggt ataaaagggg gcgggccccc gctcgtcctc 5940

actgtcttcc ggatcgctgt ccaggagcgc cagctgttgg ggtaggtatt ccctctcgaa 6000

ggcgggcatg acctcggcac tcaggttgtc agtttctaga aacgaggagg atttgatatt 6060

gacggtgccg ttggagacgc ctttcatgag ccccctcgtcc atttggtcag aaaagacgat 6120

ctttttgttg tcgagcttgg tggcgaagga gccgtagagg gcgttggaga gcagcttggc 6180

gatggagcgc atggtctggt tcttttcctt gtcggcgcgc tccttggcgg cgatgttgag 6240

ctgcacgtac tcgcgcgcca cgcacttcca ttcggggaag acggtggtga gctcgtcggg 6300

cacgattctg acccgccagc cgcggttgtg cagggtgatg aggtccacgc tggtggccac 6360

ctcgccgcgc aggggctcgt tggtccagca gaggcgcccg cccttgcgcg agcagaaggg 6420

gggcagcggg tccagcatga gctcgtcggg ggggtcggcg tccacggtga agatgccggg 6480

caggagctcg gggtcgaagt agctgatgca ggtgcccaga ttgtccagcg ccgcttgcca 6540

gtcgcgcacg gccagcgcgc gctcgtaggg gctgaggggc gtgccccagg gcatggggtg 6600

cgtgagcgcg gaggcgtaca tgccgcagat gtcgtagacg tagaggggct cctcgaggac 6660

gccgatgtag gtggggtagc agcgcccccc gcggatgctg gcgcgcacgt agtcgtacag 6720

ctcgtgcgag ggcgcgagga gccccgtgcc gaggttggag cgttgcggct tttcggcgcg 6780

gtagacgatc tggcggaaga tggcgtggga gttggaggag atggtgggcc tttggaagat 6840

gttgaagtgg gcgtggggca ggccgaccga gtccctgatg aagtgggcgt aggagtcctg 6900

cagcttggcg acgagctcgg cggtgacgag gacgtccagg gcgcagtagt cgagggtctc 6960

ttggatgatg tcatacttga gctggccctt ctgcttccac agctcgcggt tgagaaggaa 7020

ctcttcgcgg tccttccagt actcttcgag ggggaacccg tcctgatcgg cacggtaaga 7080

gcccaccatg tagaactggt tgacggcctt gtaggcgcag cagcccttct ccacggggag 7140

ggcgtaagct tgcgcggcct tgcgcaggga ggtgtgggtg agggcgaagg tgtcgcgcac 7200

catgaccttg aggaactggt gcttgaagtc gaggtcgtcg cagccgccct gctcccagag 7260

ttggaagtcc gtgcgcttct gtaggcgggg gttaggcaaa gcgaaagtaa catcgttgaa 7320

gaggatcttg cccgcgcggg gcatgaagtt gcgagtgatg cggaaaggct ggggcacctc 7380

ggcccggttg ttgatgacct gggcggcgag gacgatctcg tcgaagccgt tgatgttgtg 7440

cccgacgatg tagagttcca cgaatcgcgg gcggcccttg acgtggggca gcttcttgag 7500

ctcgtcgtag gtgagctcgg cggggtcgct gagcccgtgc tgctcgaggg cccagtcggc 7560

gacgtggggg ttggcgctga ggaaggaagt ccagagatcc acggccaggg cggtctgcaa 7620

gcggtcccgg tactgacgga actgttggcc cacggccatt ttttcggggg tgacgcagta 7680

gaaggtgcgg gggtcgccgt gccagcggtc ccacttgagc tggagggcga ggtcgtgggc 7740

gagctcgacg agcggcgggt ccccggagag tttcatgacc agcatgaagg ggacgagctg 7800

cttgccgaag gaccccatcc aggtgtaggt ttccacatcg taggtgagga agagcctttc 7860

ggtgcgagga tgcgagccga tggggaagaa ctggatctcc tgccaccagt tggaggaatg 7920

gctgttgatg tgatggaagt agaaatgccg acggcgcgcc gagcactcgt gcttgtgttt 7980

atacaagcgt ccgcagtgct cgcaacgctg cacgggatgc acgtgctgca cgagctgtac 8040

ctgggttcct ttggcgagga atttcagtgg gcagtggagc gctggcggct gcatctcgtg 8100

ctgtactacg tcttggccat cggcgtggcc atcgtctgcc tcgatggtgg tcatgctgac 8160

gagcccgcgc gggaggcagg tccagacctc ggctcggacg ggtcggagag cgaggacgag 8220

ggcgcgcagg ccggagctgt ccagggtcct gagacgctgc ggagtcaggt cagtgggcag 8280

cggcggcgcg cggttgactt gcaggagctt ttccagggcg cgcgggaggt ccagatggta 8340

cttgatctcc acggcgccgt tggtggctac gtccacggct tgcagggtgc cgtgcccctg 8400

gggcgccacc accgtgcccc gtttcttctt gggcgctgct tccatgtcgg tcagaagcgg 8460

cggcgaggac gcgcgccggg cggcaggggc ggctcggggc ccggaggcag gggcggcagg 8520

ggcacgtcgg cgccgcgcgc gggcaggttc tggtactgcg cccggagaag actggcgtga 8580

gcgacgacgc gacggttgac gtcctggatc tgacgcctct gggtgaaggc cacgggaccc 8640

gtgagtttga acctgaaaga gagttcgaca gaatcaatct cggtatcgtt gacggcggcc 8700

tgccgcagga tctcttgcac gtcgcccgag ttgtcctggt aggcgatctc ggtcatgaac 8760

tgctcgatct cctcctcctg aaggtctccg cggccggcgc gctcgacggt ggccgcgagg 8820

tcgttggaga tgcggcccat gagctgcgag aaggcgttca tgccggcctc gttccagacg 8880

cggctgtaga ccacggctcc gtcggggtcg cgcgcgcgca tgaccacctg ggcgaggttg 8940

agctcgacgt ggcgcgtgaa gaccgcgtag ttgcagaggc gctggtagag gtagttgagc 9000

gtggtggcga tgtgctcggt gacgaagaag tacatgatcc agcggcggag cggcatctcg 9060

ctgacgtcgc ccagggcttc caagcgttcc atggcctcgt agaagtccac ggcgaagttg 9120

aaaaactggg agttgcgcgc cgagacggtc aactcctcct ccagaagacg gatgagctcg 9180

gcgatggtgg cgcgcacctc gcgctcgaag gccccggggg gctcctcttc catctcctcc 9240

tcttcctcct ccactaacat ctcttctact tcctcctcag gaggcggtgg cgggggaggg 9300

gccctgcgtc gccggcggcg cacgggcaga cggtcgatga agcgctcgat ggtctccccg 9360

cgccggcgac gcatggtctc ggtgacggcg cgcccgtcct cgcggggccg cagcatgaag 9420

acgccgccgc gcatctccag gtggccgccg gggggtctc cgttgggcag ggagagggcg 9480

ctgacgatgc atcttatcaa ttgacccgta gggactccgc gcaaggacct gagcgtctcg 9540

agatccacgg gatccgaaaa ccgctgaacg aaggcttcga gccagtcgca gtcgcaaggt 9600

aggctgagcc cggtttcttg ttcttcgggt atttggtcgg gaggcgggcg ggcgatgctg 9660

ctggtgatga agttgaagta ggcggtcctg agacggcgga tggtggcgag gagcaccagg 9720

tccttgggcc cggcttgctg gatgcgcaga cggtcggcca tgccccaggc gtggtcctga 9780

cacctggcga ggtccttgta gtagtcctgc atgagccgct ccacgggcac ctcctcctcg 9840

cccgcgcggc cgtgcatgcg cgtgagcccg aacccgcgct gcggctggac gagcgccagg 9900

tcggcgacga cgcgctcggt gaggatggcc tgctggatct gggtgagggt ggtctggaag 9960

tcgtcgaagt cgacgaagcg gtggtaggct ccggtgttga tggtgtagga gcagttggcc 10020

atgacggacc agttgacggt ctggtggccg ggtcgcacga gctcgtggta cttgaggcgc 10080

gagtaggcgc gcgtgtcgaa gatgtagtcg ttgcaggcgc gcacgaggta ctggtatccg 10140

acgaggaagt gcggcggcgg ctggcggtag agcggccatc gctcggtggc gggggcgccg 10200

ggcgcgaggt cctcgagcat gaggcggtgg tagccgtaga tgtacctgga catccaggtg 10260

atgccggcgg cggtggtgga ggcgcgcggg aactcgcgga cgcggttcca gatgttgcgc 10320

agcggcagga agtagttcat ggtggccgcg gtctggcccg tgaggcgcgc gcagtcgtgg 10380

atgctctaga catacgggca aaaacgaaag cggtcagcgg ctcgactccg tggcctggag 10440

gctaagcgaa cgggttgggc tgcgcgtgta ccccggttcg aatctcgaat caggctggag 10500

ccgcagctaa cgtggtactg gcactcccgt ctcgacccaa gcctgctaac gaaacctcca 10560

ggatacggag gcgggtcgtt ttttggcctt ggtcgctggt catgaaaaac tagtaagcgc 10620

ggaaagcggc cgcccgcgat ggctcgctgc cgtagtctgg agaaagaatc gccagggttg 10680

cgttgcggtg tgccccggtt cgagcctcag cgctcggcgc cggccggatt ccgcggctaa 10740

cgtgggcgtg gctgccccgt cgtttccaag acccccttagc cagccgactt ctccagttac 10800

ggagcgagcc cctcttttttt tttcttgtgt ttttgccaga tgcatcccgt actgcggcag 10860

atgcgccccc accctccacc acaaccgccc ctaccgcagc agcagcaaca gccggcgctt 10920

ctgcccccgc cccagcagca gccagccact accgcggcgg ccgccgtgag cggagccggc 10980

gttcagtatg acctggcctt ggaagagggc gaggggctgg cgcggctggg ggcgtcgtcg 11040

ccggagcggc acccgcgcgt gcagatgaaa agggacgctc gcgaggccta cgtgcccaag 11100

cagaacctgt tcagagacag gagcggcgag gagcccgagg agatgcgcgc ctcccgcttc 11160

cacgcggggc gggagctgcg gcgcggcctg gaccgaaagc gggtgctgag ggacgaggat 11220

ttcgaggcgg acgagctgac ggggatcagc cccgcgcgcg cgcacgtggc cgcggccaac 11280

ctggtcacgg cgtacgagca gaccgtgaag gaggagagca acttccaaaa atccttcaac 11340

aaccacgtgc gcacgctgat cgcgcgcgag gaggtgaccc tgggcctgat gcacctgtgg 11400

gacctgctgg aggccatcgt gcagaacccc acgagcaagc cgctgacggc gcagctgttt 11460

ctggtggtgc agcacagtcg ggacaacgag acgttcaggg aggcgctgct gaatatcacc 11520

gagcccgagg gccgctggct cctggacctg gtgaacattt tgcagagcat cgtggtgcag 11580

gagcgcgggc tgccgctgtc cgagaagctg gcggccatca acttctcggt gctgagtctg 11640

ggcaagtact acgctaggaa gatctacaag accccgtacg tgcccataga caaggaggtg 11700

aagatcgacg ggttttacat gcgcatgacc ctgaaagtgc tgaccctgag cgacgatctg 11760

ggggtgtacc gcaacgacag gatgcaccgc gcggtgagcg ccagccgccg gcgcgagctg 11820

agcgaccagg agctgatgca cagcctgcag cgggccctga ccggggccgg gaccgagggg 11880

gagagctact ttgacatggg cgcggacctg cgctggcagc ccagccgccg ggccttggaa 11940

gctgccggcg gttccccctа cgtggaggag gtggacgatg aggaggagga gggcgagtac 12000

ctggaagact gatggcgcga ccgtattttt gctagatgca gcaacagcca ccgccgccgc 12060

ctcctgatcc cgcgatgcgg gcggcgctgc agagccagcc gtccggcatt aactcctcgg 12120

acgattggac ccaggccatg caacgcatca tggcgctgac gacccgcaat cccgaagcct 12180

ttagacagca gcctcaggcc aaccggctct cggccatcct ggaggccgtg gtgccctcgc 12240

```
gctcgaaccc cacgcacgag aaggtgctgg ccatcgtgaa cgcgctggtg gagaacaagg  12300

ccatccgcgg tgacgaggcc gggctggtgt acaacgcgct gctggagcgc gtggcccgct  12360

acaacagcac caacgtgcag acgaacctgg accgcatggt gaccgacgtg cgcgaggcgg  12420

tgtcgcagcg cgagcggttc caccgcgagt cgaacctggg ctccatggtg gcgctgaacg  12480

ccttcctgag cacgcagccc gccaacgtgc cccggggcca ggaggactac accaacttca  12540

tcagcgcgct gcggctgatg gtggccgagg tgccccagag cgaggtgtac cagtcggggc  12600

cggactactt cttccagacc agtcgccagg gcttgcagac cgtgaacctg agccaggctt  12660

tcaagaactt gcagggactg tggggcgtgc aggccccggt cggggaccgc gcgacggtgt  12720

cgagcctgct gacgccgaac tcgcgcctgc tgctgctgct ggtggcgccc ttcacggaca  12780

gcggcagcgt gagccgcgac tcgtacctgg gctacctgct taacctgtac cgcgaggcca  12840

tcggacaggc gcacgtggac gagcagacct accaggagat cacccacgtg agccgcgcgc  12900

tgggccagga ggacccgggc aacctggagg ccaccctgaa cttcctgctg accaaccggt  12960

cgcagaagat cccgccccag tacgcgctga gcaccgagga ggagcgcatc ctgcgctacg  13020

tgcagcagag cgtggggctg ttcctgatgc aggaggggcc cacgcccagc gcggcgctcg  13080

acatgaccgc gcgcaacatg gagcccagca tgtacgcccg caaccgcccg ttcatcaata  13140

agctgatgga ctacttgcat cgggcggccg ccatgaactc ggactacttt accaacgcca  13200

tcttgaaccc gcactggctc ccgccgcccg ggttctacac gggcgagtac gacatgcccg  13260

accccaacga cgggttcctg tgggacgacg tggacagcag cgtgttctcg ccgcgtccag  13320

gaaccaatgc cgtgtggaag aaagagggcg gggaccggcg gccgtcctcg gcgctgtccg  13380

gtcgcgcggg tgctgccgcg gcggtgcccg aggccgccag ccccttcccg agcctgccct  13440

tttcgctgaa cagcgtgcgc agcagcgagc tgggtcggct gacgcgaccg cgcctgctgg  13500

gcgaggagga gtacctgaac gactccttgt tgaggcccga gcgcgagaag aacttcccca  13560

ataacgggat agagagcctg gtggacaaga tgagccgctg gaagacgtac gcgcacgagc  13620

acagggacga gccccgagct agcagcgcag gcacccgtag acgccagcgg cacgacaggc  13680

agcggggact ggtgtgggac gatgaggatt ccgccgacga cagcagcgtg ttggacttgg  13740

gtgggagtgg tggtaacccg ttcgctcacc tgcgcccccg tatcgggcgc ctgatgtaag  13800

aatctgaaaa aataaaagac ggtactcacc aaggccatgg cgaccagcgt gcgttcttct  13860

ctgttgtttg tagtagt atg atg agg cgc gtg tac ccg gag ggt cct cct      13910
                 Met Met Arg Arg Val Tyr Pro Glu Gly Pro Pro
                 1               5                   10
```

```
ccc tcg tac gag agc gtg atg cag cag gcg gtg gcg gcg gcg atg cag    13958
Pro Ser Tyr Glu Ser Val Met Gln Gln Ala Val Ala Ala Ala Met Gln
            15                  20                  25

ccc ccg ctg gag gcg cct tac gtg ccc ccg cgg tac ctg gcg cct acg    14006
Pro Pro Leu Glu Ala Pro Tyr Val Pro Pro Arg Tyr Leu Ala Pro Thr
        30                  35                  40

gag ggg cgg aac agc att cgt tac tcg gag ctg gca ccc ttg tac gat    14054
Glu Gly Arg Asn Ser Ile Arg Tyr Ser Glu Leu Ala Pro Leu Tyr Asp
    45                  50                  55

acc acc cgg ttg tac ctg gtg gac aac aag tcg gca gac atc gcc tcg    14102
Thr Thr Arg Leu Tyr Leu Val Asp Asn Lys Ser Ala Asp Ile Ala Ser
60                  65                  70                  75

ctg aac tac cag aac gac cac agc aac ttc ctg acc acc gtg gtg cag    14150
Leu Asn Tyr Gln Asn Asp His Ser Asn Phe Leu Thr Thr Val Val Gln
                80                  85                  90

aac aac gat ttc acc ccc acg gag gcc agc acc cag acc atc aac ttt    14198
Asn Asn Asp Phe Thr Pro Thr Glu Ala Ser Thr Gln Thr Ile Asn Phe
            95                  100                 105

gac gag cgc tcg cgg tgg ggc ggc cag ctg aaa acc atc atg cac acc    14246
Asp Glu Arg Ser Arg Trp Gly Gly Gln Leu Lys Thr Ile Met His Thr
        110                 115                 120

aac atg ccc aac gtg aac gag ttc atg tac agc aac aag ttc aag gcg    14294
Asn Met Pro Asn Val Asn Glu Phe Met Tyr Ser Asn Lys Phe Lys Ala
    125                 130                 135

cgg gtg atg gtc tcg cgc aag acc ccc aac ggg gtg gat gat gat tat    14342
Arg Val Met Val Ser Arg Lys Thr Pro Asn Gly Val Asp Asp Asp Tyr
140                 145                 150                 155

gat ggt agt cag gac gag ctg acc tac gag tgg gtg gag ttt gag ctg    14390
Asp Gly Ser Gln Asp Glu Leu Thr Tyr Glu Trp Val Glu Phe Glu Leu
                160                 165                 170

ccc gag ggc aac ttc tcg gtg acc atg acc atc gat ctg atg aac aac    14438
Pro Glu Gly Asn Phe Ser Val Thr Met Thr Ile Asp Leu Met Asn Asn
            175                 180                 185

gcc atc atc gac aac tac ttg gcg gtg ggg cgg cag aac ggg gtg ctg    14486
Ala Ile Ile Asp Asn Tyr Leu Ala Val Gly Arg Gln Asn Gly Val Leu
        190                 195                 200

gag agc gac atc ggc gtg aag ttc gac acg cgc aac ttc cgg ctg ggc    14534
Glu Ser Asp Ile Gly Val Lys Phe Asp Thr Arg Asn Phe Arg Leu Gly
    205                 210                 215

tgg gac ccc gtg acc gag ctg gtg atg ccg ggc gtg tac acc aac gag    14582
Trp Asp Pro Val Thr Glu Leu Val Met Pro Gly Val Tyr Thr Asn Glu
220                 225                 230                 235
```

```
gcc ttc cac ccc gac atc gtc ctg ctg ccc ggc tgc ggc gtg gac ttc    14630
Ala Phe His Pro Asp Ile Val Leu Leu Pro Gly Cys Gly Val Asp Phe
                240                 245                 250

acc gag agc cgc ctc agc aac ctg ctg ggc atc cgc aag cgg cag ccc    14678
Thr Glu Ser Arg Leu Ser Asn Leu Leu Gly Ile Arg Lys Arg Gln Pro
            255                 260                 265

ttc cag gag ggc ttc cag atc ctg tac gag gac ctg gag ggg ggc aac    14726
Phe Gln Glu Gly Phe Gln Ile Leu Tyr Glu Asp Leu Glu Gly Gly Asn
        270                 275                 280

atc ccc gcg ctc ttg gat gtc gaa gcc tac gag aaa agc aag gag gat    14774
Ile Pro Ala Leu Leu Asp Val Glu Ala Tyr Glu Lys Ser Lys Glu Asp
    285                 290                 295

agc acc gcc gcg gcg acc gca gcc gtg gcc acc gcc tct acc gag gtg    14822
Ser Thr Ala Ala Ala Thr Ala Ala Val Ala Thr Ala Ser Thr Glu Val
300                 305                 310                 315

cgg ggc gat aat ttt gct agc gct gcg gca gcg gcc gag gcg gct gaa    14870
Arg Gly Asp Asn Phe Ala Ser Ala Ala Ala Ala Ala Glu Ala Ala Glu
                320                 325                 330

acc gaa agt aag ata gtc atc cag ccg gtg gag aag gac agc aag gac    14918
Thr Glu Ser Lys Ile Val Ile Gln Pro Val Glu Lys Asp Ser Lys Asp
            335                 340                 345

agg agc tac aac gtg ctc gcg gac aag aaa aac acc gcc tac cgc agc    14966
Arg Ser Tyr Asn Val Leu Ala Asp Lys Lys Asn Thr Ala Tyr Arg Ser
        350                 355                 360

tgg tac ctg gcc tac aac tac ggc gac ccc gag aag ggc gtg cgc tcc    15014
Trp Tyr Leu Ala Tyr Asn Tyr Gly Asp Pro Glu Lys Gly Val Arg Ser
    365                 370                 375

tgg acg ctg ctc acc acc tcg gac gtc acc tgc ggc gtg gag caa gtc    15062
Trp Thr Leu Leu Thr Thr Ser Asp Val Thr Cys Gly Val Glu Gln Val
380                 385                 390                 395

tac tgg tcg ctg ccc gac atg atg caa gac ccg gtc acc ttc cgc tcc    15110
Tyr Trp Ser Leu Pro Asp Met Met Gln Asp Pro Val Thr Phe Arg Ser
                400                 405                 410

acg cgt caa gtt agc aac tac ccg gtg gtg ggc gcc gag ctc ctg ccc    15158
Thr Arg Gln Val Ser Asn Tyr Pro Val Val Gly Ala Glu Leu Leu Pro
            415                 420                 425

gtc tac tcc aag agc ttc ttc aac gag cag gcc gtc tac tcg cag cag    15206
Val Tyr Ser Lys Ser Phe Phe Asn Glu Gln Ala Val Tyr Ser Gln Gln
        430                 435                 440

ctg cgc gcc ttc acc tcg ctc acg cac gtc ttc aac cgc ttc ccc gag    15254
Leu Arg Ala Phe Thr Ser Leu Thr His Val Phe Asn Arg Phe Pro Glu
    445                 450                 455

aac cag atc ctc gtc cgc ccg ccc gcg ccc acc att acc acc gtc agt    15302
```

```
Asn Gln Ile Leu Val Arg Pro Pro Ala Pro Thr Ile Thr Thr Val Ser
460                 465                 470                 475

gaa aac gtt cct gct ctc aca gat cac ggg acc ctg ccg ctg cgc agc    15350
Glu Asn Val Pro Ala Leu Thr Asp His Gly Thr Leu Pro Leu Arg Ser
                480                 485                 490

agt atc cgg gga gtc cag cgc gtg acc gtc act gac gcc aga cgc cgc    15398
Ser Ile Arg Gly Val Gln Arg Val Thr Val Thr Asp Ala Arg Arg Arg
            495                 500                 505

acc tgc ccc tac gtc tac aag gcc ctg ggc gta gtc gcg ccg cgc gtc    15446
Thr Cys Pro Tyr Val Tyr Lys Ala Leu Gly Val Val Ala Pro Arg Val
        510                 515                 520

ctc tcg agc cgc acc ttc taa aaaatgtcca ttctcatctc gcccagtaat       15497
Leu Ser Ser Arg Thr Phe
    525

aacaccggtt ggggcctgcg cgcgcccagc aagatgtacg gaggcgctcg ccaacgctcc  15557

acgcaacacc ccgtgcgcgt gcgcgggcac ttccgcgctc cctggggcgc cctcaagggc  15617

cgcgtgcgct cgcgcaccac cgtcgacgac gtgatcgacc aggtggtggc cgacgcgcgc  15677

aactacacgc ccgccgccgc gcccgtctcc accgtggacg ccgtcatcga cagcgtggtg  15737

gccgacgcgc gccggtacgc ccgcaccaag agccggcggc ggcgcatcgc ccggcggcac  15797

cggagcaccc ccgccatgcg cgcggcgcga gccttgctgc gcagggccag gcgcacggga  15857

cgcagggcca tgctcagggc ggccagacgc gcggcctccg gcagcagcag cgccggcagg  15917

acccgcagac gcgcggccac ggcggcggcg gcggccatcg ccagcatgtc ccgcccgcgg  15977

cgcggcaacg tgtactgggt gcgcgacgcc gccaccggtg tgcgcgtgcc cgtgcgcacc  16037

cgcccccctc gcacttgaag atgctgactt cgcgatgttg atgtgtccca gcggcgagga  16097

ggatgtccaa gcgcaaatac aaggaagaga tgctccaggt catcgcgcct gagatctacg  16157

gccccgcggc ggcggtgaag gaggaaagaa agccccgcaa actgaagcgg gtcaaaaagg  16217

acaaaaagga ggaggaagat gacggactgg tggagtttgt gcgcgagttc gccccccggc  16277

ggcgcgtgca gtggcgcggg cggaaagtga aaccggtgct gcggcccggc accacggtgg  16337

tcttcacgcc cggcgagcgt tccggctccg cctccaagcg ctcctacgac gaggtgtacg  16397

gggacgagga catcctcgag caggcggtcg agcgtctggg cgagtttgcg tacggcaagc  16457

gcagccgccc cgcgcccttg aaagaggagg cggtgtccat cccgctggac cacggcaacc  16517

ccacgccgag cctgaagccg gtgaccctgc agcaggtgct accgagcgcg gcgccgcgcc  16577

ggggcttcaa gcgcgagggc ggcgaggatc tgtacccgac catgcagctg atggtgccca  16637
```

```
agcgccagaa gctggaggac gtgctggagc acatgaaggt ggaccccgag gtgcagcccg  16697

aggtcaaggt gcggcccatc aagcaggtgg ccccgggcct gggcgtgcag accgtggaca  16757

tcaagatccc cacggagccc atggaaacgc agaccgagcc cgtgaagccc agcaccagca  16817

ccatggaggt gcagacggat ccctggatgc cagcaccagc ttccaccagc actcgccgaa  16877

gacgcaagta cggcgcggcc agcctgctga tgcccaacta cgcgctgcat ccttccatca  16937

tccccacgcc gggctaccgc ggcacgcgct tctaccgcgg ctacaccagc agccgccgcc  16997

gcaagaccac cacccgccgc cgtcgtcgca gccgccgcag cagcaccgcg acttccgcct  17057

tggtgcggag agtgtatcgc agcgggcgcg agcctctgac cctgccgcgc gcgcgctacc  17117

acccgagcat cgccatttaa ctaccgcctc ctacttgcag atatggccct cacatgccgc  17177

ctccgcgtcc ccattacggg ctaccgagga agaaagccgc gccgtagaag gctgacgggg  17237

aacgggctgc gtcgccatca ccaccggcgg cggcgcgcca tcagcaagcg gttgggggga  17297

ggcttcctgc ccgcgctgat ccccatcatc gccgcggcga tcggggcgat ccccggcata  17357

gcttccgtgg cggtgcaggc ctctcagcgc cactgagaca caaaaaagca tggatttgta  17417

ataaaaaaaa aaatggactg acgctcctgg tcctgtgatg tgtgttttta gatggaagac  17477

atcaattttt cgtccctggc accgcgacac ggcacgcggc cgtttatggg cacctggagc  17537

gacatcggca acagccaact gaacgggggc gccttcaatt ggagcagtct ctggagcggg  17597

cttaagaatt tcgggtccac gctcaaaacc tatggcaaca aggcgtggaa cagcagcaca  17657

gggcaggcgc tgagggaaaa gctgaaagaa cagaacttcc agcagaaggt ggttgatggc  17717

ctggcctcag gcatcaacgg ggtggttgac ctggccaacc aggccgtgca gaaacagatc  17777

aacagccgcc tggacgcggt cccgcccgcg gggtccgtgg agatgcccca ggtggaggag  17837

gagctgcctc ccctggacaa gcgcggcgac aagcgaccgc gtcccgacgc ggaggagacg  17897

ctgctgacgc acacggacga gccgcccccg tacgaggagg cggtgaaact gggcctgccc  17957

accacgcggc ccgtggcgcc tctggccacc ggagtgctga aacccagcag cagccagccc  18017

gcgaccctgg acttgcctcc gcctcgcccc tccacagtgg ctaagccccc gccgccggtg  18077

gccgtcgcgt cgcgcgcccc ccgaggccgc ccccaggcga actggcagag cactctgaac  18137

agcatcgtgg gtctgggagt gcagagtgtg aagcgccgcc gctgctatta aaagacactg  18197

tagcgcttaa cttgcttgtc tgtgtgtata tgtatgtccg ccgaccagaa ggaggagtgt  18257

gaagaggcgc gtcgccgagt tgcaag atg gcc acc cca tcg atg ctg ccc cag   18310
                              Met Ala Thr Pro Ser Met Leu Pro Gln
                              530                 535
```

```
tgg gcg tac atg cac atc gcc gga cag gac gct tcg gag tac ctg agt    18358
Trp Ala Tyr Met His Ile Ala Gly Gln Asp Ala Ser Glu Tyr Leu Ser
    540                 545                 550

ccg ggt ctg gtg cag ttc gcc cgc gcc aca gac acc tac ttc agt ctg    18406
Pro Gly Leu Val Gln Phe Ala Arg Ala Thr Asp Thr Tyr Phe Ser Leu
555                 560                 565                 570

ggg aac aag ttt agg aac ccc acg gtg gcg ccc acg cac gat gtg acc    18454
Gly Asn Lys Phe Arg Asn Pro Thr Val Ala Pro Thr His Asp Val Thr
                575                 580                 585

acc gac cgc agc cag cgg ctg acg ctg cgc ttc gtg ccc gtg gac cgc    18502
Thr Asp Arg Ser Gln Arg Leu Thr Leu Arg Phe Val Pro Val Asp Arg
            590                 595                 600

gag gac aac acc tac tcg tac aaa gtg cgc tac acg ctg gcc gtg ggc    18550
Glu Asp Asn Thr Tyr Ser Tyr Lys Val Arg Tyr Thr Leu Ala Val Gly
        605                 610                 615

gac aac cgc gtg ctg gac atg gcc agc acc tac ttt gac atc cgc ggc    18598
Asp Asn Arg Val Leu Asp Met Ala Ser Thr Tyr Phe Asp Ile Arg Gly
    620                 625                 630

gtg ctg gac cgg ggc cct agc ttc aaa ccc tac tct ggc acc gcc tac    18646
Val Leu Asp Arg Gly Pro Ser Phe Lys Pro Tyr Ser Gly Thr Ala Tyr
635                 640                 645                 650

aac agc cta gct ccc aag gga gct ccc aat tcc agc cag tgg gag caa    18694
Asn Ser Leu Ala Pro Lys Gly Ala Pro Asn Ser Ser Gln Trp Glu Gln
                655                 660                 665

gca aaa aca ggc aat ggg gga act atg gaa aca cac aca tat ggt gtg    18742
Ala Lys Thr Gly Asn Gly Gly Thr Met Glu Thr His Thr Tyr Gly Val
            670                 675                 680

gcc cca atg ggc gga gag aat att aca aaa gat ggt ctt caa att gga    18790
Ala Pro Met Gly Gly Glu Asn Ile Thr Lys Asp Gly Leu Gln Ile Gly
        685                 690                 695

act gac gtt aca gcg aat cag aat aaa cca att tat gcc gac aaa aca    18838
Thr Asp Val Thr Ala Asn Gln Asn Lys Pro Ile Tyr Ala Asp Lys Thr
    700                 705                 710

ttt caa cca gaa ccg caa gta gga gaa gaa aat tgg caa gaa act gaa    18886
Phe Gln Pro Glu Pro Gln Val Gly Glu Glu Asn Trp Gln Glu Thr Glu
715                 720                 725                 730

aac ttt tat ggc ggt aga gct ctt aaa aaa gac aca aac atg aaa cct    18934
Asn Phe Tyr Gly Gly Arg Ala Leu Lys Lys Asp Thr Asn Met Lys Pro
                735                 740                 745

tgc tat ggc tcc tat gct aga ccc acc aat gaa aaa gga ggt caa gct    18982
Cys Tyr Gly Ser Tyr Ala Arg Pro Thr Asn Glu Lys Gly Gly Gln Ala
            750                 755                 760
```

```
aaa ctt aaa gtt gga gat gat gga gtt cca acc aaa gaa ttc gac ata    19030
Lys Leu Lys Val Gly Asp Asp Gly Val Pro Thr Lys Glu Phe Asp Ile
        765                 770                 775

gac ctg gct ttc ttt gat act ccc ggt ggc acc gtg aac ggt caa gac    19078
Asp Leu Ala Phe Phe Asp Thr Pro Gly Gly Thr Val Asn Gly Gln Asp
    780                 785                 790

gag tat aaa gca gac att gtc atg tat acc gaa aac acg tat ttg gaa    19126
Glu Tyr Lys Ala Asp Ile Val Met Tyr Thr Glu Asn Thr Tyr Leu Glu
795                 800                 805                 810

act cca gac acg cat gtg gta tac aaa cca ggc aag gat gat gca agt    19174
Thr Pro Asp Thr His Val Val Tyr Lys Pro Gly Lys Asp Asp Ala Ser
                815                 820                 825

tct gaa att aac ctg gtt cag cag tct atg ccc aac aga ccc aac tac    19222
Ser Glu Ile Asn Leu Val Gln Gln Ser Met Pro Asn Arg Pro Asn Tyr
            830                 835                 840

att ggg ttc agg gac aac ttt atc ggt ctt atg tac tac aac agc act    19270
Ile Gly Phe Arg Asp Asn Phe Ile Gly Leu Met Tyr Tyr Asn Ser Thr
        845                 850                 855

ggc aat atg ggt gtg ctt gct ggt cag gcc tcc cag ctg aat gct gtg    19318
Gly Asn Met Gly Val Leu Ala Gly Gln Ala Ser Gln Leu Asn Ala Val
    860                 865                 870

gtt gat ttg caa gac aga aac acc gag ctg tcc tac cag ctc ttg ctt    19366
Val Asp Leu Gln Asp Arg Asn Thr Glu Leu Ser Tyr Gln Leu Leu Leu
875                 880                 885                 890

gac tct ttg ggt gac aga acc cgg tat ttc agt atg tgg aac cag gcg    19414
Asp Ser Leu Gly Asp Arg Thr Arg Tyr Phe Ser Met Trp Asn Gln Ala
                895                 900                 905

gtg gac agt tat gac ccc gat gtg cgc atc atc gaa aac cat ggt gtg    19462
Val Asp Ser Tyr Asp Pro Asp Val Arg Ile Ile Glu Asn His Gly Val
            910                 915                 920

gag gat gaa ttg cca aac tat tgc ttc ccc ttg gac ggc tct ggc act    19510
Glu Asp Glu Leu Pro Asn Tyr Cys Phe Pro Leu Asp Gly Ser Gly Thr
        925                 930                 935

aac gcc gca tac caa ggt gtg aaa gta aaa gat ggt caa gat ggt gat    19558
Asn Ala Ala Tyr Gln Gly Val Lys Val Lys Asp Gly Gln Asp Gly Asp
    940                 945                 950

gtt gag agt gaa tgg gaa aat gac gat act gtt gca gct cga aat caa    19606
Val Glu Ser Glu Trp Glu Asn Asp Asp Thr Val Ala Ala Arg Asn Gln
955                 960                 965                 970

tta tgt aaa ggt aac att ttc gcc atg gag att aat ctc cag gct aac    19654
Leu Cys Lys Gly Asn Ile Phe Ala Met Glu Ile Asn Leu Gln Ala Asn
                975                 980                 985

ctg tgg aga agt ttc ctc tac tcg aac gtg gcc ctg tac ctg  ccc gac    19702
```

```
Leu Trp Arg Ser Phe Leu Tyr Ser Asn Val Ala Leu Tyr Leu  Pro Asp
            990                 995                 1000

tcc tac aag  tac acg ccg acc aac  gtc acg ctg ccg acc  aac acc     19747
Ser Tyr Lys  Tyr Thr Pro Thr Asn  Val Thr Leu Pro Thr  Asn Thr
        1005                1010                1015

aac acc tac  gat tac atg aat ggc  aga gtg aca cct ccc  tcg ctg     19792
Asn Thr Tyr  Asp Tyr Met Asn Gly  Arg Val Thr Pro Pro  Ser Leu
        1020                1025                1030

gta gac gcc  tac ctc aac atc ggg  gcg cgc tgg tcg ctg  gac ccc     19837
Val Asp Ala  Tyr Leu Asn Ile Gly  Ala Arg Trp Ser Leu  Asp Pro
        1035                1040                1045

atg gac aac  gtc aac ccc ttc aac  cac cac cgc aac gcg  ggc ctg     19882
Met Asp Asn  Val Asn Pro Phe Asn  His His Arg Asn Ala  Gly Leu
        1050                1055                1060

cgc tac cgc  tcc atg ctc ctg ggc  aac ggg cgc tac gtg  ccc ttc     19927
Arg Tyr Arg  Ser Met Leu Leu Gly  Asn Gly Arg Tyr Val  Pro Phe
        1065                1070                1075

cac atc cag  gtg ccc caa aag ttt  ttc gcc atc aag agc  ctc ctg     19972
His Ile Gln  Val Pro Gln Lys Phe  Phe Ala Ile Lys Ser  Leu Leu
        1080                1085                1090

ctc ctg ccc  ggg tcc tac acc tac  gag tgg aac ttc cgc  aag gac     20017
Leu Leu Pro  Gly Ser Tyr Thr Tyr  Glu Trp Asn Phe Arg  Lys Asp
        1095                1100                1105

gtc aac atg  atc ctg cag agc tcc  cta ggc aac gac ctg  cgc acg     20062
Val Asn Met  Ile Leu Gln Ser Ser  Leu Gly Asn Asp Leu  Arg Thr
        1110                1115                1120

gac ggg gcc  tcc atc gcc ttc acc  agc atc aac ctc tac  gcc acc     20107
Asp Gly Ala  Ser Ile Ala Phe Thr  Ser Ile Asn Leu Tyr  Ala Thr
        1125                1130                1135

ttc ttc ccc  atg gcg cac aac acc  gcc tcc acg ctc gag  gcc atg     20152
Phe Phe Pro  Met Ala His Asn Thr  Ala Ser Thr Leu Glu  Ala Met
        1140                1145                1150

ctg cgc aac  gac acc aac gac cag  tcc ttc aac gac tac  ctc tcg     20197
Leu Arg Asn  Asp Thr Asn Asp Gln  Ser Phe Asn Asp Tyr  Leu Ser
        1155                1160                1165

gcg gcc aac  atg ctc tac ccc atc  ccg gcc aac gcc acc  aac gtg     20242
Ala Ala Asn  Met Leu Tyr Pro Ile  Pro Ala Asn Ala Thr  Asn Val
        1170                1175                1180

ccc atc tcc  atc ccc tcg cgc aac  tgg gcc gcc ttc cgc  gga tgg     20287
Pro Ile Ser  Ile Pro Ser Arg Asn  Trp Ala Ala Phe Arg  Gly Trp
        1185                1190                1195

tcc ttc acg  cgc ctg aag acc cgc  gag acg ccc tcg ctc  ggc tcc     20332
Ser Phe Thr  Arg Leu Lys Thr Arg  Glu Thr Pro Ser Leu  Gly Ser
```

```
        1200                1205                1210

ggg ttc gac  ccc tac ttc gtc tac  tcg ggc tcc atc ccc  tac cta     20377
Gly Phe Asp  Pro Tyr Phe Val Tyr  Ser Gly Ser Ile Pro  Tyr Leu
        1215                1220                1225

gac ggc acc  ttc tac ctc aac cac  acc ttc aag aag gtc  tcc atc     20422
Asp Gly Thr  Phe Tyr Leu Asn His  Thr Phe Lys Lys Val  Ser Ile
        1230                1235                1240

acc ttc gac  tcc tcc gtc agc tgg  ccc ggc aac gac cgc  ctc ctg     20467
Thr Phe Asp  Ser Ser Val Ser Trp  Pro Gly Asn Asp Arg  Leu Leu
        1245                1250                1255

acg ccc aac  gag ttc gaa atc aag  cgc acc gtc gac gga  gag gga     20512
Thr Pro Asn  Glu Phe Glu Ile Lys  Arg Thr Val Asp Gly  Glu Gly
        1260                1265                1270

tac aac gtg  gcc cag tgc aac atg  acc aag gac tgg ttc  ctg gtc     20557
Tyr Asn Val  Ala Gln Cys Asn Met  Thr Lys Asp Trp Phe  Leu Val
        1275                1280                1285

cag atg ctg  gcc cac tac aac atc  ggc tac cag ggc ttc  tac gtg     20602
Gln Met Leu  Ala His Tyr Asn Ile  Gly Tyr Gln Gly Phe  Tyr Val
        1290                1295                1300

ccc gag ggc  tac aag gac cgc atg  tac tcc ttc ttc cgc  aac ttc     20647
Pro Glu Gly  Tyr Lys Asp Arg Met  Tyr Ser Phe Phe Arg  Asn Phe
        1305                1310                1315

cag ccc atg  agc cgc cag gtc gtg  gac gag gtc aac tac  aag gac     20692
Gln Pro Met  Ser Arg Gln Val Val  Asp Glu Val Asn Tyr  Lys Asp
        1320                1325                1330

tac cag gcc  gtc acc ctg gcc tac  cag cac aac aac tcg  ggc ttc     20737
Tyr Gln Ala  Val Thr Leu Ala Tyr  Gln His Asn Asn Ser  Gly Phe
        1335                1340                1345

gtc ggc tac  ctc gcg ccc acc atg  cgc cag ggc cag ccc  tac ccc     20782
Val Gly Tyr  Leu Ala Pro Thr Met  Arg Gln Gly Gln Pro  Tyr Pro
        1350                1355                1360

gcc aac tac  ccc tac ccg ctc atc  ggc aag agc gcc gtc  gcc agc     20827
Ala Asn Tyr  Pro Tyr Pro Leu Ile  Gly Lys Ser Ala Val  Ala Ser
        1365                1370                1375

gtc acc cag  aaa aag ttc ctc tgc  gac cgg gtc atg tgg  cgc atc     20872
Val Thr Gln  Lys Lys Phe Leu Cys  Asp Arg Val Met Trp  Arg Ile
        1380                1385                1390

ccc ttc tcc  agc aac ttc atg tcc  atg ggc gcg ctc acc  gac ctc     20917
Pro Phe Ser  Ser Asn Phe Met Ser  Met Gly Ala Leu Thr  Asp Leu
        1395                1400                1405

ggc cag aac  atg ctc tac gcc aac  tcc gcc cac gcg cta  gac atg     20962
Gly Gln Asn  Met Leu Tyr Ala Asn  Ser Ala His Ala Leu  Asp Met
        1410                1415                1420
```

```
aat ttc gaa  gtc gac ccc atg gat  gag tcc acc ctt ctc  tat gtt     21007
Asn Phe Glu  Val Asp Pro Met Asp  Glu Ser Thr Leu Leu  Tyr Val
        1425                 1430                 1435

gtc ttc gaa  gtc ttc gac gtc gtc  cga gtg cac cag ccc  cac cgc     21052
Val Phe Glu  Val Phe Asp Val Val  Arg Val His Gln Pro  His Arg
        1440                 1445                 1450

ggc gtc atc  gaa gcc gtc tac ctg  cgc acg ccc ttc tcg  gcc ggc     21097
Gly Val Ile  Glu Ala Val Tyr Leu  Arg Thr Pro Phe Ser  Ala Gly
        1455                 1460                 1465

aac gcc acc  acc taa gccgctcttg cttcttgcaa gatgacggcg ggctccggcg   21152
Asn Ala Thr  Thr
        1470

agcaggagct cagggccatc ctccgcgacc tgggctgcgg gccctgcttc ctgggcacct  21212

tcgacaagcg cttccctgga ttcatggccc cgcacaagct ggcctgcgcc atcgtgaaca  21272

cggccggccg cgagaccggg ggcgagcact ggctggcctt cgcctggaac ccgcgctccc  21332

acacatgcta cctcttcgac cccttcgggt tctcggacga gcgcctcaag cagatctacc  21392

agttcgagta cgagggcctg ctgcgtcgca gcgccctggc caccgaggac cgctgcgtca  21452

ccctggaaaa gtccacccag accgtgcagg gtccgcgctc ggccgcctgc gggctcttct  21512

gctgcatgtt cctgcacgcc ttcgtgcact ggcccgaccg ccccatggac aagaacccca  21572

ccatgaactt actgacgggg gtgcccaacg gcatgctcca gtcgccccag gtggaaccca  21632

ccctgcgccg caaccaggaa gcgctctacc gcttcctcaa tgcccactcc gcctactttc  21692

gctcccaccg cgcgcgcatc gagaaggcca ccgccttcga ccgcatgaat caagacatgt  21752

aaaaaaccgg tgtgtgtatg tgaatgcttt attcataata aacagcacat gtttatgcca  21812

ccttctctga ggctctgact ttatttagaa atcgaagggg ttctgccggc tctcggcatg  21872

gcccgcgggc agggatacgt tgcggaactg gtacttgggc agccacttga actcggggat  21932

cagcagcttg ggcacgggga ggtcggggaa cgagtcgctc cacagcttgc gcgtgagttg  21992

cagggcgccc agcaggtcgg gcgcggagat cttgaaatcg cagttgggac ccgcgttctg  22052

cgcgcgagag ttgcggtaca cggggttgca gcactggaac accatcaggg ccgggtgctt  22112

cacgcttgcc agcaccgtcg cgtcggtgat gccctccacg tccagatcct cggcgttggc  22172

catcccgaag ggggtcatct tgcaggtctg ccgccccatg ctgggcacgc agccgggctt  22232

gtggttgcaa tcgcagtgca gggggatcag catcatctgg gcctgctcgg agctcatgcc  22292

cgggtacatg gccttcatga aagcctccag ctggcggaag gcctgctgcg ccttgccgcc  22352
```

ctcggtgaag aagaccccgc aggacttgct agagaactgg ttggtggcgc agccggcgtc 22412

gtgcacgcag cagcgcgcgt cgttgttggc cagctgcacc acgctgcgcc cccagcggtt 22472

ctgggtgatc ttggcccggt tggggttctc cttcagcgcg cgctgcccgt tctcgctcgc 22532

cacatccatc tcgatagtgt gctccttctg gatcatcacg gtcccgtgca ggcaccgcag 22592

cttgccctcg gcttcggtgc agccgtgcag ccacagcgcg cagccggtgc actcccagtt 22652

cttgtgggcg atctgggagt gcgagtgcac gaagccctgc aggaagcggc ccatcatcgc 22712

ggtcagggtc ttgttgctgg tgaaggtcag cgggatgccg cggtgctcct cgttcacata 22772

caggtggcag atgcggcggt acacctcgcc ctgctcgggc atcagctgga aggcggactt 22832

caggtcgctc tccacgcggt accggtccat cagcagcgtc atcacttcca tgcccttctc 22892

ccaggccgaa acgatcggca ggctcagggg gttcttcacc gccattgtca tcttagtcgc 22952

cgccgccgag gtcaggggggt cgttctcgtc cagggtctca aacactcgct tgccgtcctt 23012

ctcgatgatg cgcacggggg gaaagctgaa gcccacggcc gccagctcct cctcggcctg 23072

cctttcgtcc tcgctgtcct ggctgatgtc ttgcaaaggc acatgcttgg tcttgcgggg 23132

tttctttttg ggcggcagag gcggcggcga tgtgctggga gagcgcgagt tctcgttcac 23192

cacgactatt tcttcttctt ggccgtcgtc cgagaccacg cggcggtagg catgcctctt 23252

ctggggcaga ggcggaggcg acgggctctc gcggttcggc gggcggctgg cagagcccct 23312

tccgcgttcg ggggtgcgct cctggcggcg ctgctctgac tgacttcctc cgcggccggc 23372

cattgtgttc tcctagggag caacaacaag catggagact cagccatcgt cgccaacatc 23432

gccatctgcc cccgccgcca ccgccgacga gaaccagcag cagaatgaaa gcttaaccgc 23492

cccgccgccc agccccacct ccgacgccgc ggccccagac atgcaagaga tggaggaatc 23552

catcgagatt gacctgggct acgtgacgcc cgcggagcac gaggaggagc tggcagcgcg 23612

cttttcagcc ccggaagaga accaccaaga gcagccagag caggaagcag agaacgagca 23672

gaaccaggct gggcacgagc atggcgacta cctgagcggg gcagaggacg tgctcatcaa 23732

gcatctggcc cgccaatgca tcatcgtcaa ggacgcgctg ctcgaccgcg ccgaggtgcc 23792

cctcagcgtg gcggagctca gccgcgccta cgagcgcaac ctcttctcgc cgcgcgtgcc 23852

ccccaagcgc cagcccaacg gcacctgtga gcccaacccg cgcctcaact tctacccggt 23912

cttcgcggtg cccgaggccc tggccaccta ccacctcttt ttcaagaacc aaaggatccc 23972

cgtctcctgc cgcgccaacc gcacccgcgc cgacgccctg ctcaacctgg gccccggcgc 24032

ccgcctacct gatatcacct ccttggaaga ggttcccaag atcttcgagg gtctgggcag 24092

cgacgagact cgggccgcga acgctctgca aggaagcgga gaggagcatg agcaccacag 24152

cgccctggtg gagttggaag gcgacaacgc gcgcctggcg gtcctcaagc gcacggtcga 24212

gctgacccac ttcgcctacc cggcgctcaa cctgcccccc aaggtcatga gcgccgtcat 24272

ggaccaggtg ctcatcaagc gcgcctcgcc cctctcggag gaggagatgc aggaccccga 24332

gagttcggac gagggcaagc ccgtggtcag cgacgagcag ctggcgcgct ggctgggagc 24392

gagtagcacc ccccagagcc tggaagagcg gcgcaagctc atgatggccg tggtcctggt 24452

gaccgtggag ctggagtgtc tgcgccgctt ctttgccgac gcggagaccc tgcgcaaggt 24512

cgaggagaac ctgcactacc tcttcaggca cgggttcgtg cgccaggcct gcaagatctc 24572

caacgtggag ctgaccaacc tggtctccta catgggcatc ctgcacgaga accgcctggg 24632

gcaaaacgtg ctgcacacca ccctgcgcgg ggaggcccgc cgcgactaca tccgcgactg 24692

cgtctacctg tacctctgcc acacctggca gacgggcatg ggcgtgtggc agcagtgcct 24752

ggaggagcag aacctgaaag agctctgcaa gctcctgcag aagaacctca aggccctgtg 24812

gaccgggttc gacgagcgta ccaccgcctc ggacctggcc gacctcatct tccccgagcg 24872

cctgcggctg acgctgcgca acgggctgcc cgactttatg agccaaagca tgttgcaaaa 24932

ctttcgctct ttcatcctcg aacgctccgg gatcctgccc gccacctgct ccgcgctgcc 24992

ctcggacttc gtgccgctga ccttccgcga gtgccccccg ccgctctgga gccactgcta 25052

cttgctgcgc ctggccaact acctggccta ccactcggac gtgatcgagg acgtcagcgg 25112

cgagggtctg ctggagtgcc actgccgctg caacctctgc acgccgcacc gctccctggc 25172

ctgcaacccc cagctgctga gcgagaccca gatcatcggc accttcgagt tgcaaggccc 25232

cggcgacggc gagggcaagg ggggtctgaa actcaccccg gggctgtgga cctcggccta 25292

cttgcgcaag ttcgtgcccg aggactacca tcccttcgag atcaggttct acgaggacca 25352

atcccagccg cccaaggccg agctgtcggc ctgcgtcatc acccaggggg ccatcctggc 25412

ccaattgcaa gccatccaga aatcccgcca agaatttctg ctgaaaaagg gccacggggt 25472

ctacttggac ccccagaccg gagaggagct caaccccagc ttcccccagg atgccccgag 25532

gaagcagcaa gaagctgaaa gtggagctgc cgccgccgga ggatttggag gaagactggg 25592

agagcagtca ggcagaggag gaggagatgg aagactggga cagcactcag gcagaggagg 25652

acagcctgca agacagtctg gaggaggaag acgaggtgga ggaggcagag gaagaagcag 25712

ccgccgccag accgtcgtcc tcggcggaga aagcaagcag cacggatacc atctccgctc 25772

cgggtcgggg tcgcggcggc cgggcccaca gtaggtggga cgagaccggg cgcttcccga 25832

accccaccac ccagaccggt aagaaggagc ggcagggata caagtcctgg cgggggcaca 25892

aaaacgccat cgtctcctgc ttgcaagcct gcgggggcaa catctccttc acccggcgct 25952

acctgctctt ccaccgcggg gtgaacttcc cccgcaacat cttgcattac taccgtcacc 26012

tccacagccc ctactactgt ttccaagaag aggcagaaac ccagcagcag cagaaaacca 26072

gcggcagcag cagctagaaa atccacagcg gcggcaggtg gactgaggat cgcggcgaac 26132

gagccggcgc agacccggga gctgaggaac cggatctttc ccaccctcta tgccatcttc 26192

cagcagagtc gggggcagga gcaggaactg aaagtcaaga accgttctct gcgctcgctc 26252

acccgcagtt gtctgtatca caagagcgaa gaccaacttc agcgcactct cgaggacgcc 26312

gaggctctct tcaacaagta ctgcgcgctc actcttaaag agtagcccgc gcccgcccac 26372

acacggaaaa aggcgggaat tacgtcacca cctgcgccct tcgcccgacc atcatgagca 26432

aagagattcc cacgccttac atgtggagct accagcccca gatgggcctg gccgccggcg 26492

ccgcccagga ctactccacc cgcatgaact ggctcagtgc cgggcccgcg atgatctcac 26552

gggtgaatga catccgcgcc caccgaaacc agatactcct agaacagtca gcgatcaccg 26612

ccacgccccg ccatcacctt aatccgcgta attggcccgc cgccctggtg taccaggaaa 26672

ttccccagcc cacgaccgta ctacttccgc gagacgccca ggccgaagtc cagctgacta 26732

actcaggtgt ccagctggcc ggcggcgccg ccctgtgtcg tcaccgcccc gctcagggta 26792

taaagcggct ggtgatccga ggcagaggca cacagctcaa cgacgaggtg gtgagctctt 26852

cgctgggtct gcgacctgac ggagtcttcc aactcgccgg atcggggaga tcttccttca 26912

cgcctcgtca ggccgtcctg actttggaga gttcgtcctc gcagccccgc tcgggcggca 26972

tcggcactct ccagttcgtg gaggagttca ctccctcggt ctacttcaac cccttctccg 27032

gctcccccgg ccactacccg gacgagttca tcccgaactt cgacgccatc agcgagtcgg 27092

tggacggcta cgattgaatg tcccatggtg gcgcagctga cctagctcgg cttcgacacc 27152

tggaccactg ccgccgcttc cgctgcttcg ctcgggatct cgccgagttt gcctactttg 27212

agctgcccga ggagcaccct cagggcccag cccacggagt gcggatcatc gtcgaagggg 27272

gcctcgactc ccacctgctt cggatcttca gccagcgacc gatcctggtc gagcgcgaac 27332

aaggacagac ccttcttact ttgtactgca tctgcaacca ccccggcctg catgaaagtc 27392

tttgttgtct gctgtgtact gagtataata aaagctgaga tcagcgacta ctccggactc 27452

gattgtggtg ttcctgctat caaccggtcc ctgttcttca ccgggaacga gaccgagctc 27512

cagctccagt gtaagcccca caagaagtac ctcacctggc tgttccaggg ctccccgatc 27572

gccgttgtca accactgcga caacgacgga gtcctgctga gcggccctgc caaccttact 27632

ttttccaccc gcagaagcaa gctccagctc ttccaaccct tcctccccgg gacctatcag 27692

tgcgtctcag gaccctgcca tcacaccttc cacctgatcc cgaataccac agcgccgctc 27752

cccgctacta acaaccaaac tacccaccaa cgccaccgtc gcgacctttc ctctgaatct 27812

aataccacta ccggaggtga gctccgaggt cgaccaacct ctgggattta ctacggcccc 27872

tgggaggtgg tggggttaat agcgctaggc ctagttgcgg gtgggctttt ggttctctgc 27932

tacctatacc tcccttgctg ttcgtactta gtggtgctgt gttgctggtt taagaaatgg 27992

ggaagatcac cctagtgagc tgcggtgcgc tggtggcggt gttgctttcg attgtgggac 28052

tgggcggcgc ggctgtagtg aaggagaagg ccgatccctg cttgcatttc aatcccaaca 28112

aatgccagct gagttttcag cccgatggca atcggtgcgc ggtactgatc aagtgcggat 28172

gggaatgcga gaacgtgaga atcgagtaca ataacaagac tcggaacaat actctcgcgt 28232

ccgtgtggca gcccggggac cccgagtggt acaccgtctc tgtccccggt gctgacggct 28292

ccccgcgcac cgtgaataat actttcattt ttgcgcacat gtgcaacacg gtcatgtgga 28352

tgagcaagca gtacgatatg tggcccccca cgaaggagaa catcgtggtc ttctccatcg 28412

cttacagcct gtgcacggcg ctaatcaccg ctatcgtgtg cctgagcatt cacatgctca 28472

tcgctattcg ccccagaaat aatgccgaga aagagaaaca gccataacac gtttttttcac 28532

acaccttgtt tttacagaca atgcgtctgt taaatttttt aaacattgtg ctcagtattg 28592

cttatgcctc tggttatgca aacatacaga aaacccttta tgtaggatct gatggtacac 28652

tagagggtac ccaatcacaa gccaaggttg catggtattt ttatagaacc aacactgatc 28712

cagttaaact ttgtaagggt gaattgccgc gtacacataa aactccactt acatttagtt 28772

gcagcaataa taatcttaca ctttttttcaa ttacaaaaca atatactggt acttattaca 28832

gtacaaactt tcatacagga caagataaat attatactgt taaggtagaa aatcctacca 28892

ctcctagaac taccaccacc accactactg caaagcccac tgtgaaaact acaactagga 28952

ccaccacaac tacagaaacc accaccagca caacacttgc tgcaactaca cacacacaca 29012

ctaagctaac cttacagacc actaatgatt tgatcgccct gctgcaaaag ggggataaca 29072

gcaccacttc caatgaggag atacccaaat ccatgattgg cattattgtt gctgtagtgg 29132

tgtgcatgtt gatcatcgcc ttgtgcatgg tgtactatgc cttctgctac agaaagcaca 29192

gactgaacga caagctggaa cacttactaa gtgttgaatt ttaatttttt agaaccatga 29252

agatcctagg ccttttttagt ttttctatca ttacctctgc tctttgtgaa tcagtggata 29312

gagatgttac tattaccact ggttctaatt atacactgaa agggccaccc tcaggtatgc 29372

tttcgtggta ttgctatttt ggaactgaca ctgatcaaac tgaattatgc aattttcaaa 29432

aaggcaaaac ctcaaactct aaaatctcta attatcaatg caatggcact gatctgatac 29492

tactcaatgt cacgaaagca tatggtggca gttattattg ccctggacaa aacactgaag 29552

aaatgatttt ttacaaagtg gaagtggttg atcccactac accacccacc accacaacta 29612

ttcataccac acacacagaa caaacaccag aggcaacaga agcagagttg gccttccagg 29672

ttcacggaga ttcctttgct gtcaataccc ctacacccga tcagcggtgt ccggggccgc 29732

tagtcagcgg cattgtcggt gtgctttcgg gattagcagt cataatcatc tgcatgttca 29792

tttttgcttg ctgctataga aggctttacc gacaaaaatc agacccactg ctgaacctct 29852

atgtttaatt ttttccagag ccatgaaggc agttagcgct ctagtttttt gttctttgat 29912

tggcattgtt tttaatagta aaattaccag agttagcttt attaaacatg ttaatgtaac 29972

tgaaggagat aacatcacac tagcaggtgt agaaggtgct caaaacacca cctggacaaa 30032

ataccatcta ggatggagag atatttgcac ctggaatgta acttattatt gcataggagt 30092

taatcttacc attgttaacg ctaaccaatc tcagaatggg ttaattaaag gacagagtgt 30152

tagtgtgacc agtgatgggt actataccca gcatagtttt aactacaaca ttactgtcat 30212

accactgcct acgcctagcc cacctagcac taccacacag acaaccacat acagtacatc 30272

aaatcagcct accaccacta cagcagcaga ggttgccagc tcgtctgggg tccgagtggc 30332

atttttgatg ttggccccat ctagcagtcc cactgctagt accaatgagc agactactga 30392

atttttgtcc actgtcgaga gccacaccac agctacctcc agtgccttct ctagcaccgc 30452

caatctctcc tcgctttcct ctacaccaat cagccccgct actactccta gccccgctcc 30512

tcttcccact cccctgaagc aaacagacgg cggcatgcaa tggcagatca ccctgctcat 30572

tgtgatcggg ttggtcatcc tggccgtgtt gctctactac atcttctgcc gccgcattcc 30632

caacgcgcac cgcaagccgg cctacaagcc catcgttatc gggcagccgg agccgcttca 30692

ggtggaaggg ggtctaagga atcttctctt ctcttttaca gtatggtgat tgaactatga 30752

ttcctagaca attcttgatc actattctta tctgcctcct ccaagtctgt gccacctcg 30812

ctctggtggc caacgccagt ccagactgta ttgggccctt cgcctcctac gtgctctttg 30872

ccttcgtcac ctgcatctgc tgctgtagca tagtctgcct gcttatcacc ttcttccagt 30932

```
tcattgactg gatctttgtg cgcatcgcct acctgcgcca ccacccccag taccgcgacc  30992

agcgagtggc gcagctgctc aggctcctct gataagcatg cgggctctgc tacttctcgc  31052

gcttctgctg ttagtgctcc cccgtcccgt cgacccccgg tcccccactc agtccccccga  31112

ggaggttcgc aaatgcaaat tccaagaacc ctggaaattc ctcaaatgct accgccaaaa  31172

atcagacatg catcccagct ggatcatgat cattgggatc gtgaacattc tggcctgcac  31232

cctcatctcc tttgtgattt acccctgctt tgactttggt tggaactcgc cagaggcgct  31292

ctatctcccg cctgaacctg acacaccacc acagcagcaa cctcaggcac acgcactacc  31352

accaccacag cctaggccac aatacatgcc catattagac tatgaggccg agccacagcg  31412

acccatgctc cccgctatta gttacttcaa tctaaccggc ggagatgact gacccactgg  31472

ccaataacaa cgtcaacgac cttctcctgg acatggacgg ccgcgcctcg gagcagcgac  31532

tcgcccaact tcgcattcgt cagcagcagg agagagccgt caaggagctg caggacggca  31592

tagccatcca ccagtgcaag agaggcatct tctgcctggt gaaacaggcc aagatctcct  31652

acgaggtcac ccagaccgac catcgcctct cctacgagct cctgcagcag cgccagaagt  31712

tcacctgcct ggtcggagtc aaccccatcg tcatcaccca gcagtcgggc gataccaagg  31772

ggtgcatcca ctgctcctgc gactcccccg actgcgtcca cactctgatc aagaccctct  31832

gcggcctccg cgacctcctc cccatgaact aatcacccccc ttatccagtg aaataaagat  31892

catattgatg atgatttaaa taaaaaaaat aatcatttga tttgaaataa agatacaatc  31952

atattgatga tttgagttta acaaaaataa agaatcactt acttgaaatc tgataccagg  32012

tctctgtcca tgttttctgc caacaccacc tcactccccc cttcccagct ctggtactgc  32072

aggccccggc gggctgcaaa cttcctccac acgctgaagg ggatgtcaaa ttcctcctgt  32132

ccctcaatct tcattttatc ttctatcag atg tcc aaa aag  cgc gtc cgg gtg   32185
                              Met Ser Lys Lys  Arg Val Arg Val
                                                1475

gat  gat gac ttc gac ccc  gtc tac ccc tac gat  gca gac aac gca     32230
Asp  Asp Asp Phe Asp Pro  Val Tyr Pro Tyr Asp  Ala Asp Asn Ala
1480                 1485                 1490

ccg  acc gtg ccc ttc atc  aac ccc ccc ttc gtc  tct tca gat gga     32275
Pro  Thr Val Pro Phe Ile  Asn Pro Pro Phe Val  Ser Ser Asp Gly
1495                 1500                 1505

ttc  caa gag aag ccc ctg  ggg gtg ttg tcc ctg  cga ctg gct gac     32320
Phe  Gln Glu Lys Pro Leu  Gly Val Leu Ser Leu  Arg Leu Ala Asp
1510                 1515                 1520
```

```
ccc  gtc acc acc aag aac  ggg gaa atc acc ctc  aag ctg gga gag     32365
Pro  Val Thr Thr Lys Asn  Gly Glu Ile Thr Leu  Lys Leu Gly Glu
1525                 1530                 1535

ggg  gtg gac ctc gac tcg  tcg gga aaa ctc atc  tcc aac acg gcc     32410
Gly  Val Asp Leu Asp Ser  Ser Gly Lys Leu Ile  Ser Asn Thr Ala
1540                 1545                 1550

acc  aag gcc gcc gcc cct  ctc agt att tca aac  aac acc att tcc     32455
Thr  Lys Ala Ala Ala Pro  Leu Ser Ile Ser Asn  Asn Thr Ile Ser
1555                 1560                 1565

ctt  aaa act gct gcc cct  ttc tac aac aac aat  gga act tta agc     32500
Leu  Lys Thr Ala Ala Pro  Phe Tyr Asn Asn Asn  Gly Thr Leu Ser
1570                 1575                 1580

ctc  aat gtc tcc aca cca  tta gca gta ttt ccc  aca ttt aac act     32545
Leu  Asn Val Ser Thr Pro  Leu Ala Val Phe Pro  Thr Phe Asn Thr
1585                 1590                 1595

tta  ggc ata agt ctt gga  aac ggt ctt cag act  tca aat aag ttg     32590
Leu  Gly Ile Ser Leu Gly  Asn Gly Leu Gln Thr  Ser Asn Lys Leu
1600                 1605                 1610

ttg  act gta caa cta act  cat cct ctt aca ttc  agc tca aat agc     32635
Leu  Thr Val Gln Leu Thr  His Pro Leu Thr Phe  Ser Ser Asn Ser
1615                 1620                 1625

atc  aca gta aaa aca gac  aaa ggg cta tat att  aac tcc agt gga     32680
Ile  Thr Val Lys Thr Asp  Lys Gly Leu Tyr Ile  Asn Ser Ser Gly
1630                 1635                 1640

aac  aga gga ctt gag gct  aat ata agc cta aaa  aga gga cta gtt     32725
Asn  Arg Gly Leu Glu Ala  Asn Ile Ser Leu Lys  Arg Gly Leu Val
1645                 1650                 1655

ttt  gac ggt aat gct att  gca aca tat att gga  aat ggc tta gac     32770
Phe  Asp Gly Asn Ala Ile  Ala Thr Tyr Ile Gly  Asn Gly Leu Asp
1660                 1665                 1670

tat  gga tct tat gat agt  gat gga aaa aca aga  ccc gta att acc     32815
Tyr  Gly Ser Tyr Asp Ser  Asp Gly Lys Thr Arg  Pro Val Ile Thr
1675                 1680                 1685

aaa  att gga gca gga tta  aat ttt gat gct aac  aaa gca ata gct     32860
Lys  Ile Gly Ala Gly Leu  Asn Phe Asp Ala Asn  Lys Ala Ile Ala
1690                 1695                 1700

gtc  aaa cta ggc aca ggt  tta agt ttt gac tcc  gct ggt gcc ttg     32905
Val  Lys Leu Gly Thr Gly  Leu Ser Phe Asp Ser  Ala Gly Ala Leu
1705                 1710                 1715

aca  gct gga aac aaa cag  gat gac aag cta aca  ctt tgg act acc     32950
Thr  Ala Gly Asn Lys Gln  Asp Asp Lys Leu Thr  Leu Trp Thr Thr
1720                 1725                 1730

cct  gac cca agc cct aat  tgt caa tta ctt tca  gac aga gat gcc     32995
```

```
Pro  Asp Pro Ser Pro Asn  Cys Gln Leu Leu Ser  Asp Arg Asp Ala
1735                 1740                 1745

aaa  ttt act ctc tgt ctt  aca aaa tgc ggt agt  caa ata cta ggc     33040
Lys  Phe Thr Leu Cys Leu  Thr Lys Cys Gly Ser  Gln Ile Leu Gly
1750                 1755                 1760

act  gtg gca gtg gcg gct  gtt act gta gga tca  gca cta aat cca     33085
Thr  Val Ala Val Ala Ala  Val Thr Val Gly Ser  Ala Leu Asn Pro
1765                 1770                 1775

att  aat gac aca gtc aaa  agc gcc ata gtt ttc  ctt aga ttt gat     33130
Ile  Asn Asp Thr Val Lys  Ser Ala Ile Val Phe  Leu Arg Phe Asp
1780                 1785                 1790

tcc  gat ggt gta ctc atg  tca aac tca tca atg  gta ggt gat tac     33175
Ser  Asp Gly Val Leu Met  Ser Asn Ser Ser Met  Val Gly Asp Tyr
1795                 1800                 1805

tgg  aac ttt agg gag gga  cag acc act caa agt  gta gcc tat aca     33220
Trp  Asn Phe Arg Glu Gly  Gln Thr Thr Gln Ser  Val Ala Tyr Thr
1810                 1815                 1820

aat  gct gtg gga ttc atg  cca aat ata ggt gca  tat cca aaa acc     33265
Asn  Ala Val Gly Phe Met  Pro Asn Ile Gly Ala  Tyr Pro Lys Thr
1825                 1830                 1835

caa  agt aaa aca cct aaa  aat agc ata gtc agt  cag gta tat tta     33310
Gln  Ser Lys Thr Pro Lys  Asn Ser Ile Val Ser  Gln Val Tyr Leu
1840                 1845                 1850

act  gga gaa act act atg  cca atg aca cta acc  ata act ttc aat     33355
Thr  Gly Glu Thr Thr Met  Pro Met Thr Leu Thr  Ile Thr Phe Asn
1855                 1860                 1865

ggc  act gat gaa aaa gac  aca acc cca gtt agc  acc tac tct atg     33400
Gly  Thr Asp Glu Lys Asp  Thr Thr Pro Val Ser  Thr Tyr Ser Met
1870                 1875                 1880

act  ttt aca tgg cag tgg  act gga gac tat aag  gac aaa aat att     33445
Thr  Phe Thr Trp Gln Trp  Thr Gly Asp Tyr Lys  Asp Lys Asn Ile
1885                 1890                 1895

acc  ttt gct acc aac tca  ttc tct ttt tcc tac  atc gcc cag gaa     33490
Thr  Phe Ala Thr Asn Ser  Phe Ser Phe Ser Tyr  Ile Ala Gln Glu
1900                 1905                 1910

taa tcccacccag caagccaacc ccttttccca ccacctttgt ctatatggaa        33543

actctgaaac agaaaaataa agttcaagtg ttttattgaa tcaacagttt tacaggactc  33603

gagcagttat ttttcctcca ccctcccagg acatggaata caccaccctc tccccccgca  33663

cagccttgaa catctgaatg ccattggtga tggacatgct tttggtctcc acgttccaca  33723

cagtttcaga gcgagccagt ctcggatcgg tcagggagat gaaaccctcc gggcactccc  33783
```

gcatctgcac ctcacagctc aacagctgag gattgtcctc ggtggtcggg atcacggtta 33843

tctggaagaa gcagaagagc ggcggtggga atcatagtcc gcgaacggga tcggccggtg 33903

gtgtcgcatc aggccccgca gcagtcgctg ccgccgccgc tccgtcaagc tgctgctcag 33963

ggggttcggg tccagggact ccctcagcat gatgcccacg gccctcagca tcagtcgtct 34023

ggtgcggcgg gcgcagcagc gcatgcgaat ctcgctcagg tcactgcagt acgtgcaaca 34083

caggaccacc aggttgttca acagtccata gttcaacacg ctccagccga aactcatcgc 34143

gggaaggatg ctacccacgt ggccgtcgta ccagatcctc aggtaaatca agtggcgctc 34203

cctccagaag acgctgccca tgtacatgat ctccttgggc atgtggcggt tcaccacctc 34263

ccggtaccac atcaccctct ggttgaacat gcagccccgg atgatcctgc ggaaccacag 34323

ggccagcacc gccccgcccg ccatgcagcg aagagacccc ggatcccggc aatgacaatg 34383

gaggacccac cgctcgtacc cgtggatcat ctgggagctg aacaagtcta tgttggcaca 34443

gcacaggcat atgctcatgc atctcttcag cactctcagc tcctcggggg tcaaaaccat 34503

atcccagggc acggggaact cttgcaggac agcgaacccc gcagaacagg gcaatcctcg 34563

cacataactt acattgtgca tggacagggt atcgcaatca ggcagcaccg ggtgatcctc 34623

caccagagaa gcgcgggtct cggtctcctc acagcgtggt aaggggccg gccgatacgg 34683

gtgatggcgg gacgcggctg atcgtgttct cgaccgtgtc atgatgcagt tgctttcgga 34743

cattttcgta cttgctgtag cagaacctgg tccgggcgct gcacaccgat cgccggcggc 34803

ggtctcggcg cttggaacgc tcggtgttaa agttgtaaaa cagccactct ctcagaccgt 34863

gcagcagatc tagggcctca ggagtgatga agatcccatc atgcctgata gctctgatca 34923

catcgaccac cgtggaatgg gccaggccca gccagatgat gcaattttgt tgggtttcgg 34983

tgacggcggg ggagggaaga acaggaagaa ccatgattaa cttttaatcc aaacggtctc 35043

ggagcacttc aaaatgaagg tcacggagat ggcacctctc gcccccgctg tgttggtgga 35103

aaataacagc caggtcaaag gtgatacggt tctcgagatg ttccacggtg gcttccagca 35163

aagcctccac gcgcacatcc agaaacaaga caatagcgaa agcgggaggg ttctctaatt 35223

cctcaaccat catgttacac tcctgcacca tccccagata attttcattt ttccagcctt 35283

gaatgattcg aactagttcc tgaggtaaat ccaagccagc catgataaaa agctcgcgca 35343

gagcaccctc caccggcatt cttaagcaca ccctcataat tccaagatat tctgctcctg 35403

gttcacctgc agcagattga caagcggaat atcaaaatct ctgccgcgat ccctgagctc 35463

ctccctcagc aataactgta agtactcttt catatcgtct ccgaaatttt tagccatagg 35523

acccccagga ataagagaag ggcaagccac attacagata aaccgaagtc ccccccagtg 35583

agcattgcca aatgtaagat tgaaataagc atgctggcta gacccggtga tatcttccag 35643

ataactggac agaaaatcgg gtaagcaatt tttaagaaaa tcaacaaaag aaaaatcttc 35703

caggtgcacg tttagggcct cgggaacaac gatggagtaa gtgcaagggg tgcgttccag 35763

catggttagt tagctgatct gtaaaaaaac aaaaaataaa acattaaacc atgctagcct 35823

ggcgaacagg tgggtaaatc gttctctcca gcaccaggca ggccacgggg tctccggcgc 35883

gaccctcgta aaaattgtcg ctatgattga aaaccatcac agagagacgt tcccggtggc 35943

cggcgtgaat gattcgagaa gaagcataca ccccccggaac attggagtcc gtgagtgaaa 36003

aaaagcggcc gaggaagcaa tgaggcacta caacgctcac tctcaagtcc agcaaagcga 36063

tgccatgcgg atgaagcaca aaattttcag gtgcgtaaaa aatgtaatta ctccccctcct 36123

gcacaggcag cgaagctccc gatccctcca gatacacata caaagcctca gcgtccatag 36183

cttaccgagc ggcagcagca gcggcacaca acaggcgcaa gagtcagaga aaagactgag 36243

ctctaacctg tccgcccgct ctctgctcaa tatatagccc cagatctaca ctgacgtaaa 36303

ggccaaagtc taaaaatacc cgccaaataa tcacacacgc ccagcacacg cccagaaacc 36363

ggtgacacac tcagaaaaat acgcgcactt cctcaaacgg ccaaactgcc gtcatttccg 36423

ggttcccacg ctacgtcatc aaaacacgac tttcaaattc cgtcgaccgt taaaaacatc 36483

acccgccccg cccctaacgg tcgccgctcc cgcagccaat caccttcctc cctccccaaa 36543

ttcaaacagc tcatttgcat attaacgcgc accaaaagtt tgaggtatat tattgatgat 36603

g 36604

<210> 6
<211> 529
<212> PRT
<213> chimpanzee adenovirus serotype Pan6

<400> 6

Met Met Arg Arg Val Tyr Pro Glu Gly Pro Pro Pro Ser Tyr Glu Ser
1 5 10 15

Val Met Gln Gln Ala Val Ala Ala Ala Met Gln Pro Pro Leu Glu Ala
20 25 30

Pro Tyr Val Pro Pro Arg Tyr Leu Ala Pro Thr Glu Gly Arg Asn Ser
35 40 45

```
Ile Arg Tyr Ser Glu Leu Ala Pro Leu Tyr Asp Thr Thr Arg Leu Tyr
    50                  55                  60

Leu Val Asp Asn Lys Ser Ala Asp Ile Ala Ser Leu Asn Tyr Gln Asn
65                  70                  75                  80

Asp His Ser Asn Phe Leu Thr Thr Val Val Gln Asn Asn Asp Phe Thr
                85                  90                  95

Pro Thr Glu Ala Ser Thr Gln Thr Ile Asn Phe Asp Glu Arg Ser Arg
            100                 105                 110

Trp Gly Gly Gln Leu Lys Thr Ile Met His Thr Asn Met Pro Asn Val
        115                 120                 125

Asn Glu Phe Met Tyr Ser Asn Lys Phe Lys Ala Arg Val Met Val Ser
    130                 135                 140

Arg Lys Thr Pro Asn Gly Val Asp Asp Asp Tyr Asp Gly Ser Gln Asp
145                 150                 155                 160

Glu Leu Thr Tyr Glu Trp Val Glu Phe Glu Leu Pro Glu Gly Asn Phe
                165                 170                 175

Ser Val Thr Met Thr Ile Asp Leu Met Asn Asn Ala Ile Ile Asp Asn
            180                 185                 190

Tyr Leu Ala Val Gly Arg Gln Asn Gly Val Leu Glu Ser Asp Ile Gly
        195                 200                 205

Val Lys Phe Asp Thr Arg Asn Phe Arg Leu Gly Trp Asp Pro Val Thr
    210                 215                 220

Glu Leu Val Met Pro Gly Val Tyr Thr Asn Glu Ala Phe His Pro Asp
225                 230                 235                 240

Ile Val Leu Leu Pro Gly Cys Gly Val Asp Phe Thr Glu Ser Arg Leu
                245                 250                 255

Ser Asn Leu Leu Gly Ile Arg Lys Arg Gln Pro Phe Gln Glu Gly Phe
            260                 265                 270
```

```
Gln Ile Leu Tyr Glu Asp Leu Glu Gly Gly Asn Ile Pro Ala Leu Leu
        275                 280                 285

Asp Val Glu Ala Tyr Glu Lys Ser Lys Glu Asp Ser Thr Ala Ala Ala
    290                 295                 300

Thr Ala Ala Val Ala Thr Ala Ser Thr Glu Val Arg Gly Asp Asn Phe
305                 310                 315                 320

Ala Ser Ala Ala Ala Ala Ala Glu Ala Ala Glu Thr Glu Ser Lys Ile
                325                 330                 335

Val Ile Gln Pro Val Glu Lys Asp Ser Lys Asp Arg Ser Tyr Asn Val
            340                 345                 350

Leu Ala Asp Lys Lys Asn Thr Ala Tyr Arg Ser Trp Tyr Leu Ala Tyr
        355                 360                 365

Asn Tyr Gly Asp Pro Glu Lys Gly Val Arg Ser Trp Thr Leu Leu Thr
    370                 375                 380

Thr Ser Asp Val Thr Cys Gly Val Glu Gln Val Tyr Trp Ser Leu Pro
385                 390                 395                 400

Asp Met Met Gln Asp Pro Val Thr Phe Arg Ser Thr Arg Gln Val Ser
                405                 410                 415

Asn Tyr Pro Val Val Gly Ala Glu Leu Leu Pro Val Tyr Ser Lys Ser
            420                 425                 430

Phe Phe Asn Glu Gln Ala Val Tyr Ser Gln Gln Leu Arg Ala Phe Thr
        435                 440                 445

Ser Leu Thr His Val Phe Asn Arg Phe Pro Glu Asn Gln Ile Leu Val
    450                 455                 460

Arg Pro Pro Ala Pro Thr Ile Thr Thr Val Ser Glu Asn Val Pro Ala
465                 470                 475                 480

Leu Thr Asp His Gly Thr Leu Pro Leu Arg Ser Ser Ile Arg Gly Val
                485                 490                 495
```

Gln Arg Val Thr Val Thr Asp Ala Arg Arg Arg Thr Cys Pro Tyr Val
500 505 510

Tyr Lys Ala Leu Gly Val Val Ala Pro Arg Val Leu Ser Ser Arg Thr
515 520 525

Phe

<210> 7
<211> 942
<212> PRT
<213> chimpanzee adenovirus serotype Pan6

<400> 7

Met Ala Thr Pro Ser Met Leu Pro Gln Trp Ala Tyr Met His Ile Ala
1 5 10 15

Gly Gln Asp Ala Ser Glu Tyr Leu Ser Pro Gly Leu Val Gln Phe Ala
20 25 30

Arg Ala Thr Asp Thr Tyr Phe Ser Leu Gly Asn Lys Phe Arg Asn Pro
35 40 45

Thr Val Ala Pro Thr His Asp Val Thr Thr Asp Arg Ser Gln Arg Leu
50 55 60

Thr Leu Arg Phe Val Pro Val Asp Arg Glu Asp Asn Thr Tyr Ser Tyr
65 70 75 80

Lys Val Arg Tyr Thr Leu Ala Val Gly Asp Asn Arg Val Leu Asp Met
85 90 95

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

Phe Lys Pro Tyr Ser Gly Thr Ala Tyr Asn Ser Leu Ala Pro Lys Gly
115 120 125

Ala Pro Asn Ser Ser Gln Trp Glu Gln Ala Lys Thr Gly Asn Gly Gly
130 135 140

Thr Met Glu Thr His Thr Tyr Gly Val Ala Pro Met Gly Gly Glu Asn
145 150 155 160

```
Ile Thr Lys Asp Gly Leu Gln Ile Gly Thr Asp Val Thr Ala Asn Gln
                165                 170                 175

Asn Lys Pro Ile Tyr Ala Asp Lys Thr Phe Gln Pro Glu Pro Gln Val
            180                 185                 190

Gly Glu Glu Asn Trp Gln Glu Thr Glu Asn Phe Tyr Gly Gly Arg Ala
        195                 200                 205

Leu Lys Lys Asp Thr Asn Met Lys Pro Cys Tyr Gly Ser Tyr Ala Arg
    210                 215                 220

Pro Thr Asn Glu Lys Gly Gly Gln Ala Lys Leu Lys Val Gly Asp Asp
225                 230                 235                 240

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

Pro Gly Gly Thr Val Asn Gly Gln Asp Glu Tyr Lys Ala Asp Ile Val
            260                 265                 270

Met Tyr Thr Glu Asn Thr Tyr Leu Glu Thr Pro Asp Thr His Val Val
        275                 280                 285

Tyr Lys Pro Gly Lys Asp Asp Ala Ser Ser Glu Ile Asn Leu Val Gln
    290                 295                 300

Gln Ser Met Pro Asn Arg Pro Asn Tyr Ile Gly Phe Arg Asp Asn Phe
305                 310                 315                 320

Ile Gly Leu Met Tyr Tyr Asn Ser Thr Gly Asn Met Gly Val Leu Ala
                325                 330                 335

Gly Gln Ala Ser Gln Leu Asn Ala Val Val Asp Leu Gln Asp Arg Asn
            340                 345                 350

Thr Glu Leu Ser Tyr Gln Leu Leu Leu Asp Ser Leu Gly Asp Arg Thr
        355                 360                 365

Arg Tyr Phe Ser Met Trp Asn Gln Ala Val Asp Ser Tyr Asp Pro Asp
    370                 375                 380
```

Val Arg Ile Ile Glu Asn His Gly Val Glu Asp Glu Leu Pro Asn Tyr
385 390 395 400

Cys Phe Pro Leu Asp Gly Ser Gly Thr Asn Ala Ala Tyr Gln Gly Val
405 410 415

Lys Val Lys Asp Gly Gln Asp Gly Asp Val Glu Ser Glu Trp Glu Asn
420 425 430

Asp Asp Thr Val Ala Ala Arg Asn Gln Leu Cys Lys Gly Asn Ile Phe
435 440 445

Ala Met Glu Ile Asn Leu Gln Ala Asn Leu Trp Arg Ser Phe Leu Tyr
450 455 460

Ser Asn Val Ala Leu Tyr Leu Pro Asp Ser Tyr Lys Tyr Thr Pro Thr
465 470 475 480

Asn Val Thr Leu Pro Thr Asn Thr Asn Thr Tyr Asp Tyr Met Asn Gly
485 490 495

Arg Val Thr Pro Pro Ser Leu Val Asp Ala Tyr Leu Asn Ile Gly Ala
500 505 510

Arg Trp Ser Leu Asp Pro Met Asp Asn Val Asn Pro Phe Asn His His
515 520 525

Arg Asn Ala Gly Leu Arg Tyr Arg Ser Met Leu Leu Gly Asn Gly Arg
530 535 540

Tyr Val Pro Phe His Ile Gln Val Pro Gln Lys Phe Phe Ala Ile Lys
545 550 555 560

Ser Leu Leu Leu Leu Pro Gly Ser Tyr Thr Tyr Glu Trp Asn Phe Arg
565 570 575

Lys Asp Val Asn Met Ile Leu Gln Ser Ser Leu Gly Asn Asp Leu Arg
580 585 590

Thr Asp Gly Ala Ser Ile Ala Phe Thr Ser Ile Asn Leu Tyr Ala Thr
595 600 605

```
Phe Phe Pro Met Ala His Asn Thr Ala Ser Thr Leu Glu Ala Met Leu
    610                 615                 620

Arg Asn Asp Thr Asn Asp Gln Ser Phe Asn Asp Tyr Leu Ser Ala Ala
625                 630                 635                 640

Asn Met Leu Tyr Pro Ile Pro Ala Asn Ala Thr Asn Val Pro Ile Ser
                645                 650                 655

Ile Pro Ser Arg Asn Trp Ala Ala Phe Arg Gly Trp Ser Phe Thr Arg
            660                 665                 670

Leu Lys Thr Arg Glu Thr Pro Ser Leu Gly Ser Gly Phe Asp Pro Tyr
        675                 680                 685

Phe Val Tyr Ser Gly Ser Ile Pro Tyr Leu Asp Gly Thr Phe Tyr Leu
    690                 695                 700

Asn His Thr Phe Lys Lys Val Ser Ile Thr Phe Asp Ser Ser Val Ser
705                 710                 715                 720

Trp Pro Gly Asn Asp Arg Leu Leu Thr Pro Asn Glu Phe Glu Ile Lys
                725                 730                 735

Arg Thr Val Asp Gly Glu Gly Tyr Asn Val Ala Gln Cys Asn Met Thr
            740                 745                 750

Lys Asp Trp Phe Leu Val Gln Met Leu Ala His Tyr Asn Ile Gly Tyr
        755                 760                 765

Gln Gly Phe Tyr Val Pro Glu Gly Tyr Lys Asp Arg Met Tyr Ser Phe
    770                 775                 780

Phe Arg Asn Phe Gln Pro Met Ser Arg Gln Val Val Asp Glu Val Asn
785                 790                 795                 800

Tyr Lys Asp Tyr Gln Ala Val Thr Leu Ala Tyr Gln His Asn Asn Ser
                805                 810                 815

Gly Phe Val Gly Tyr Leu Ala Pro Thr Met Arg Gln Gly Gln Pro Tyr
            820                 825                 830

Pro Ala Asn Tyr Pro Tyr Pro Leu Ile Gly Lys Ser Ala Val Ala Ser
```

835 840 845

Val Thr Gln Lys Lys Phe Leu Cys Asp Arg Val Met Trp Arg Ile Pro
850 855 860

Phe Ser Ser Asn Phe Met Ser Met Gly Ala Leu Thr Asp Leu Gly Gln
865 870 875 880

Asn Met Leu Tyr Ala Asn Ser Ala His Ala Leu Asp Met Asn Phe Glu
885 890 895

Val Asp Pro Met Asp Glu Ser Thr Leu Leu Tyr Val Val Phe Glu Val
900 905 910

Phe Asp Val Val Arg Val His Gln Pro His Arg Gly Val Ile Glu Ala
915 920 925

Val Tyr Leu Arg Thr Pro Phe Ser Ala Gly Asn Ala Thr Thr
930 935 940

<210> 8
<211> 443
<212> PRT
<213> chimpanzee adenovirus serotype Pan6

<400> 8

Met Ser Lys Lys Arg Val Arg Val Asp Asp Asp Phe Asp Pro Val Tyr
1 5 10 15

Pro Tyr Asp Ala Asp Asn Ala Pro Thr Val Pro Phe Ile Asn Pro Pro
20 25 30

Phe Val Ser Ser Asp Gly Phe Gln Glu Lys Pro Leu Gly Val Leu Ser
35 40 45

Leu Arg Leu Ala Asp Pro Val Thr Thr Lys Asn Gly Glu Ile Thr Leu
50 55 60

Lys Leu Gly Glu Gly Val Asp Leu Asp Ser Ser Gly Lys Leu Ile Ser
65 70 75 80

Asn Thr Ala Thr Lys Ala Ala Ala Pro Leu Ser Ile Ser Asn Asn Thr
85 90 95

```
Ile Ser Leu Lys Thr Ala Ala Pro Phe Tyr Asn Asn Asn Gly Thr Leu
            100                 105                 110

Ser Leu Asn Val Ser Thr Pro Leu Ala Val Phe Pro Thr Phe Asn Thr
        115                 120                 125

Leu Gly Ile Ser Leu Gly Asn Gly Leu Gln Thr Ser Asn Lys Leu Leu
    130                 135                 140

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

Val Lys Thr Asp Lys Gly Leu Tyr Ile Asn Ser Ser Gly Asn Arg Gly
                165                 170                 175

Leu Glu Ala Asn Ile Ser Leu Lys Arg Gly Leu Val Phe Asp Gly Asn
            180                 185                 190

Ala Ile Ala Thr Tyr Ile Gly Asn Gly Leu Asp Tyr Gly Ser Tyr Asp
        195                 200                 205

Ser Asp Gly Lys Thr Arg Pro Val Ile Thr Lys Ile Gly Ala Gly Leu
    210                 215                 220

Asn Phe Asp Ala Asn Lys Ala Ile Ala Val Lys Leu Gly Thr Gly Leu
225                 230                 235                 240

Ser Phe Asp Ser Ala Gly Ala Leu Thr Ala Gly Asn Lys Gln Asp Asp
                245                 250                 255

Lys Leu Thr Leu Trp Thr Thr Pro Asp Pro Ser Pro Asn Cys Gln Leu
            260                 265                 270

Leu Ser Asp Arg Asp Ala Lys Phe Thr Leu Cys Leu Thr Lys Cys Gly
        275                 280                 285

Ser Gln Ile Leu Gly Thr Val Ala Val Ala Ala Val Thr Val Gly Ser
    290                 295                 300

Ala Leu Asn Pro Ile Asn Asp Thr Val Lys Ser Ala Ile Val Phe Leu
305                 310                 315                 320
```

Arg Phe Asp Ser Asp Gly Val Leu Met Ser Asn Ser Ser Met Val Gly
325 330 335

Asp Tyr Trp Asn Phe Arg Glu Gly Gln Thr Thr Gln Ser Val Ala Tyr
340 345 350

Thr Asn Ala Val Gly Phe Met Pro Asn Ile Gly Ala Tyr Pro Lys Thr
355 360 365

Gln Ser Lys Thr Pro Lys Asn Ser Ile Val Ser Gln Val Tyr Leu Thr
370 375 380

Gly Glu Thr Thr Met Pro Met Thr Leu Thr Ile Thr Phe Asn Gly Thr
385 390 395 400

Asp Glu Lys Asp Thr Thr Pro Val Ser Thr Tyr Ser Met Thr Phe Thr
405 410 415

Trp Gln Trp Thr Gly Asp Tyr Lys Asp Lys Asn Ile Thr Phe Ala Thr
420 425 430

Asn Ser Phe Ser Phe Ser Tyr Ile Ala Gln Glu
435 440

<210> 9
<211> 36535
<212> DNA
<213> chimpanzee adenovirus serotype Pan7

<220>
<221> CDS
<222> (13874)..(15469)
<223> L2 Penton

<220>
<221> CDS
<222> (18288)..(21086)
<223> L3 Hexon

<220>
<221> CDS
<222> (32094)..(33425)
<223> L5 Fiber

<400> 9
catcatcaat aatatacctc aaacttttgg tgcgcgttaa tatgcaaatg agctgtttga 60

atttggggag ggaggaaggt gattggccga gagacgggcg accgttaggg gcggggcggg 120

tgacgttttt aatacgtggc cgtgaggcgg agccggtttg caagttctcg tgggaaaagt 180

gacgtcaaac gaggtgtggt ttgaacacgg aaatactcaa ttttcccgcg ctctctgaca 240

ggaaatgagg tgtttctggg cggatgcaag tgaaaacggg ccattttcgc gcgaaaactg 300

aatgaggaag tgaaaatctg agtaatttcg cgtttatggc agggaggagt atttgccgag 360

ggccgagtag actttgaccg attacgtggg ggtttcgatt accgtatttt tcacctaaat 420

ttccgcgtac ggtgtcaaag tccggtgttt ttacgtaggc gtcagctgat cgccagggta 480

tttaaacctg cgctctctag tcaagaggcc actcttgagt gccagcgagt agagttttct 540

cctccgcgcc gcgagtcaga tctacacttt gaaagatgag gcacctgaga gacctgcccg 600

gtaatgtttt cctggctact gggaacgaga ttctggaatt ggtggtggac gccatgatgg 660

gtggcgaccc tcctgagccc cctaccccat ttgaggcgcc ttcgctgtac gatttgtatg 720

atctggaggt ggatgtgccc gagaacgacc ccaacgagga ggcggtgaat gatttgttta 780

gcgatgccgc gctgctggct gccgagcagg ctaatacgga ctctggctca gacagcgatt 840

cctctctcca taccccgaga cccggcagag gtgagaaaaa gatccccgag cttaaagggg 900

aagagctcga cctgcgctgc tatgaggaat gcttgcctcc gagcgatgat gaggaggacg 960

aggaggcgat tcgagctgca tcgaaccagg gagtgaaagc tgcgggcgaa agctttagcc 1020

tggactgtcc tactctgccc ggacacggct gtaagtcttg tgaatttcat cgcatgaata 1080

ctggagataa gaatgtgatg tgtgccctgt gctatatgag agcttacaac cattgtgttt 1140

acagtaagtg tgattaactt tagttgggaa ggcagagggt gactgggtgc tgactggttt 1200

atttatgtat atgtttttt atgtgtaggt cccgtctctg acgtagatga gacccccact 1260

tcagagtgca tttcatcacc cccagaaatt ggcgaggaac cgcccgaaga tattattcat 1320

agaccagttg cagtgagagt caccgggcgg agagcagctg tggagagttt ggatgacttg 1380

ctacagggtg gggatgaacc tttggacttg tgtacccgga aacgccccag gcactaagtg 1440

ccacacatgt gtgtttactt aaggtgatgt cagtatttat agggtgtgga gtgcaataaa 1500

atccgtgttg actttaagtg cgtggtttat gactcagggg tggggactgt gggtatataa 1560

gcaggtgcag acctgtgtgg tcagttcaga gcaggactca tggagatctg gacggtcttg 1620

gaagactttc accagactag acagctgcta gagaactcat cggaggggtt ctcttacctg 1680

tggagattct gcttcggtgg gcctctagct aagctagtct atagggccaa acaggattat 1740

aaggatcaat ttgaggatat tttgagagag tgtcctggta tttttgactc tctcaacttg 1800

ggccatcagt ctcactttaa ccagagtatt ctgagagccc ttgacttttc tactcctggc 1860

agaactaccg ccgcggtagc ctttttttgcc tttatccttg acaaatggag tcaagaaacc 1920

catttcagca gggattaccg tctggactgc ttagcagtag ctttgtggag aacatggagg 1980

tgccagcgcc tgaatgcaat ctccggctac ttgccagtac agccggtaga cacgctgagg 2040

atcctgagtc tccagtcacc ccaggaacac caacgccgcc agcagccgca gcaggagcag 2100

cagcaagagg aggaggagga tcgagaagag aacccgagag ccggtctgga ccctccggtg 2160

gcggaggagg aggagtagct gacttgtttc ccgagctgcg ccgggtgctg actaggtctt 2220

ccagtggacg ggagaggggg attaagcggg agaggcatga ggagactagc cacagaactg 2280

aactgactgt cagtctgatg agccgcaggc gcccagaatc ggtgtggtgg catgaggttc 2340

agtcgcaggg gatagatgag gtctcggtga tgcatgagaa atattccctg gaacaagtca 2400

agacttgttg gttggagcct gaggatgatt gggaggtagc catcaggaat tatgccaagc 2460

tggctctgaa gccagacaag aagtacaaga ttaccaaact gattaatatc agaaattcct 2520

gctacatttc agggaatggg gccgaggtgg agatcagtac ccaggagagg gtggccttca 2580

gatgttgtat gatgaatatg tacccggggg tggtgggcat ggagggagtc acctttatga 2640

acgcgaggtt caggggtgat gggtataatg gggtggtctt tatggccaac accaagctga 2700

cagtgcacgg atgctccttc tttgggttca ataacatgtg catcgaggcc tggggcagtg 2760

tttcagtgag gggatgcagc ttttcagcca actggatggg ggtcgtgggc agaaccaaga 2820

gcaaggtgtc agtgaagaaa tgcctgttcg agaggtgcca cctgggggtg atgagcgagg 2880

gcgaagccaa agtcaaacac tgcgcctcta ctgagacggg ctgctttgtg ctgatcaagg 2940

gcaatgccca agtcaagcat aacatgatct gtggggcctc ggatgagcgc ggctaccaga 3000

tgctgacctg cgccggtggg aacagccata tgctggccac cgtgcatgtg acctcgcacc 3060

cccgcaagac atggcccgag ttcgagcaca acgtcatgac ccgatgcaat gtgcacctgg 3120

ggtcccgccg aggcatgttc atgccctacc agtgcaacat gcaatttgtg aaggtgctgc 3180

tggagcccga tgccatgtcc agagtgagcc tgacgggggt gtttgacatg aatgtggagc 3240

tgtggaaaat tctgagatat gatgaatcca agaccaggtg ccgggcctgc gaatgcggag 3300

gcaagcacgc caggcttcag cccgtgtgtg tggaggtgac ggaggacctg cgacccgatc 3360

atttggtgtt gtcctgcaac gggacggagt tcggctccag cggggaagaa tctgactaga 3420

gtgagtagtg tttgggggag gtggagggct tgtatgaggg gcagaatgac taaaatctgt 3480

gtttttctgt gtgttgcagc agcatgagcg gaagcgcctc ctttgaggga ggggtattca 3540

gcccttatct gacggggcgt ctcccctcct gggcgggagt gcgtcagaat gtgatgggat 3600

ccacggtgga cggccggccc gtgcagcccg cgaactcttc aaccctgacc tacgcgaccc 3660

tgagctcctc gtccgtggac gcagctgccg ccgcagctgc tgcttccgcc gccagcgccg 3720

tgcgcggaat ggccctgggc gccggctact acagctctct ggtggccaac tcgacttcca 3780

ccaataatcc cgccagcctg aacgaggaga agctgctgct gctgatggcc cagctcgagg 3840

ccctgaccca gcgcctgggc gagctgaccc agcaggtggc tcagctgcag gcggagacgc 3900

gggccgcggt tgccacggtg aaaaccaaat aaaaaatgaa tcaataaata aacggagacg 3960

gttgttgatt ttaacacaga gtcttgaatc tttatttgat ttttcgcgcg cggtaggccc 4020

tggaccaccg gtctcgatca ttgagcaccc ggtggatttt ttccaggacc cggtagaggt 4080

gggcttggat gttgaggtac atgggcatga gcccgtcccg ggggtggagg tagctccatt 4140

gcagggcctc gtgctcgggg gtggtgttgt aaatcaccca gtcatagcag gggcgcaggg 4200

cgtggtgctg cacgatgtcc ttgaggagga gactgatggc cacgggcagc cccttggtgt 4260

aggtgttgac gaacctgttg agctgggagg gatgcatgcg gggggagatg agatgcatct 4320

tggcctggat cttgagattg gcgatgttcc cgcccagatc ccgccggggg ttcatgttgt 4380

gcaggaccac cagcacggtg tatccggtgc acttggggaa tttgtcatgc aacttggaag 4440

ggaaggcgtg aaagaatttg gagacgccct tgtgaccgcc caggttttcc atgcactcat 4500

ccatgatgat ggcgatgggc ccgtgggcgg cggcctgggc aaagacgttt cgggggtcgg 4560

acacatcgta gttgtggtcc tgggtgagct cgtcataggc cattttaatg aatttggggc 4620

ggagggtgcc cgactggggg acgaaggtgc cctcgatccc gggggcgtag ttgccctcgc 4680

agatctgcat ctcccaggcc ttgagctcgg agggggggat catgtccacc tgcggggcga 4740

tgaaaaaaac ggtttccggg gcgggggaga tgagctgggc cgaaagcagg ttccggagca 4800

gctgggactt gccgcagccg gtggggccgt agatgacccc gatgaccggc tgcaggtggt 4860

agttgaggga gagacagctg ccgtcctcgc ggaggagggg ggccacctcg ttcatcatct 4920

cgcgcacatg catgttctcg cgcacgagtt ccgccaggag gcgctcgccc cccagcgaga 4980

ggagctcttg cagcgaggcg aagtttttca gcggcttgag yccgtcggcc atgggcattt 5040

tggagagggt ctgttgcaag agttccagac ggtcccagag ctcggtgatg tgctctaggg 5100

catctcgatc cagcagacct cctcgtttcg cgggttgggg cgactgcggg agtagggcac 5160

caggcgatgg gcgtccagcg aggccagggt ccggtccttc cagggtcgca gggtccgcgt 5220

cagcgtggtc tccgtcacgg tgaaggggtg cgcgccgggc tgggcgcttg cgagggtgcg 5280

cttcaggctc atccggctgg tcgagaaccg ctcccggtcg gcgccctgcg cgtcggccag 5340

gtagcaattg agcatgagtt cgtagttgag cgcctcggcc gcgtggccct tggcgcggag 5400

cttacctttg gaagtgtgtc cgcagacggg acagaggagg gacttgaggg cgtagagctt 5460

gggggcgagg aagacggact cgggggcgta ggcgtccgcg ccgcagctgg cgcagacggt 5520

ctcgcactcc acgagccagg tgaggtcggg ccggttgggg tcaaaaacga ggtttcctcc 5580

gtgctttttg atgcgtttct tacctctggt ctccatgagc tcgtgtcccc gctgggtgac 5640

aaagaggctg tccgtgtccc cgtagaccga ctttatgggc cggtcctcga gcggggtgcc 5700

gcggtcctcg tcgtagagga accccgccca ctccgagacg aaggcccggg tccaggccag 5760

cacgaaggag gccacgtggg aggggtagcg gtcgttgtcc accagcgggt ccaccttctc 5820

cagggtatgc aagcacatgt ccccctcgtc cacatccagg aaggtgattg gcttgtaagt 5880

gtaggccacg tgaccggggg tcccggccgg gggggtataa aaggggggcgg gcccctgctc 5940

gtcctcactg tcttccggat cgctgtccag gagcgccagc tgttggggta ggtattcccc 6000

ctcgaaggct ggcataacct cggcactcag gttgtcagtt tctagaaacg aggaggattt 6060

gatattgacg gtgccgttgg agacgccttt catgagcccc tcgtccatct ggtcagaaaa 6120

gacgatcttt ttgttgtcga gcttggtggc gaaggagccg tagagggcgt tggagaggag 6180

cttggcgatg gagcgcatgg tctggttctt ttccttgtcg gcgcgctcct tggcggcgat 6240

gttgagctgc acgtactcgc gcgccacgca cttccattcg gggaagacgg tggtgagctc 6300

gtcgggcacg attctgaccc gccagccgcg gttgtgcagg gtgatgaggt ccacgctggt 6360

ggccacctcg ccgcgcaggg gctcgttggt ccagcagagg cgcccgccct tgcgcgagca 6420

gaaggggggc agcgggtcca gcatgagctc gtcgggggggg tcggcgtcca cggtgaagat 6480

gccgggcaga agctcggggt cgaagtagct gatgcaggtg tccagatcgt ccagcgccgc 6540

ttgccagtcg cgcacggcca gcgcgcgctc gtaggggctg aggggcgtgc cccagggcat 6600

ggggtgcgtg agcgcggagg cgtacatgcc gcagatgtcg tagacgtaga ggggctcctc 6660

gaggacgccg atgtaggtgg ggtagcagcg ccccccgcgg atgctggcgc gcacgtagtc 6720

gtacagctcg tgcgagggcg cgaggagccc cgtgccgagg ttggagcgtt gcggcttttc 6780

ggcgcggtag acgatctggc ggaagatggc gtgggagttg gaggagatgg tgggcctctg 6840

gaagatgttg aagtgggcgt ggggcaggcc gaccgagtcc ctgatgaagt gggcgtagga 6900

gtcctgcagc ttggcgacga gctcggcggt gacgaggacg tccagggcgc agtagtcgag 6960

ggtctcttgg atgatgtcgt acttgagctg gcccttctgc ttccacagct cgcggttgag 7020

aaggaactct tcgcggtcct tccagtactc ttcgaggggg aacccgtcct gatcggcacg 7080

gtaagagccc accatgtaga actggttgac ggccttgtag gcgcagcagc ccttctccac 7140

ggggagggcg taagcttgtg cggccttgcg cagggaggtg tgggtgaggg cgaaggtgtc 7200

gcgcaccatg accttgagga actggtgctt gaagtcgagg tcgtcgcagc cgccctgctc 7260

ccagagctgg aagtccgtgc gcttcttgta ggcggggttg ggcaaagcga aagtaacatc 7320

gttgaagagg atcttgcccg cgcggggcat gaagttgcga gtgatgcgga aaggctgggg 7380

cacctcggcc cggttgttga tgacctgggc ggcgaggacg atctcgtcga agccgttgat 7440

gttgtgcccg acgatgtaga gttccacgaa tcgcgggcgg cccttaacgt ggggcagctt 7500

cttgagctcg tcgtaggtga gctcggcggg gtcgctgagc ccgtgctgct cgagggccca 7560

gtcggcgacg tgggggttgg cgctgaggaa ggaagtccag agatccacgg ccagggcggt 7620

ctgcaagcgg tcccggtact gacggaactg ctggcccacg gccatttttt cgggggtgac 7680

gcagtagaag gtgcgggggt cgccgtgcca gcggtcccac ttgagctgga gggcgaggtc 7740

gtgggcgagc tcgacgagcg gcgggtcccc ggagagtttc atgaccagca tgaaggggac 7800

gagctgcttg ccgaaggacc ccatccaggt gtaggtttcc acatcgtagg tgaggaagag 7860

cctttcggtg cgaggatgcg agccgatggg gaagaactgg atctcctgcc accagttgga 7920

ggaatggctg ttgatgtgat ggaagtagaa atgccgacgg cgcgccgagc actcgtgctt 7980

gtgtttatac aagcgtccgc agtgctcgca acgctgcacg ggatgcacgt gctgcacgag 8040

ctgtacctgg gttcctttga cgaggaattt cagtgggcag tggagcgctg gcggctgcat 8100

ctggtgctgt actacgtcct ggccatcggc gtggccatcg tctgcctcga tggtggtcat 8160

gctgacgagc ccgcgcggga ggcaggtcca gacttcggct cggacgggtc ggagagcgag 8220

gacgagggcg cgcaggccgg agctgtccag ggtcctgaga cgctgcggag tcaggtcagt 8280

gggcagcggc ggcgcgcggt tgacttgcag gagcttttcc agggcgcgcg ggaggtccag 8340

atggtacttg atctccacgg cgccgttggt ggcgacgtcc acggcttgca gggtcccgtg 8400

cccctggggc gccaccaccg tgccccgttt cttcttgggc gctgcttcca tgccggtcag 8460

aagcggcggc gaggacgcgc gccgggcggc aggggcggct cgggacccgg aggcaggggc 8520

ggcaggggca cgtcggcgcc gcgcgcgggc aggttctggt actgcgcccg gagaagactg 8580

gcgtgagcga cgacgcgacg gttgacgtcc tggatctgac gcctctgggt gaaggccacg 8640

ggacccgtga gtttgaacct gaaagagagt tcgacagaat caatctcggt atcgttgacg 8700

gcggcctgcc gcaggatctc ttgcacgtcg cccgagttgt cctggtaggc gatctcggtc 8760

atgaactgct cgatctcctc ctcctgaagg tctccgcggc cggcgcgctc gacggtggcc 8820

gcgaggtcgt tggagatgcg gcccatgagc tgcgagaagg cgttcatgcc ggcctcgttc 8880

cagacgcggc tgtagaccac ggctccgtcg gggtcgcgcg cgcgcatgac cacctgggcg 8940

aggttgagct cgacgtggcg cgtgaagacc gcgtagttgc agaggcgctg gtagaggtag 9000

ttgagcgtgg tggcgatgtg ctcggtgacg aagaagtaca tgatccagcg gcggagcggc 9060

atctcgctga cgtcgcccag ggcttccaag cgctccatgg cctcgtagaa gtccacggcg 9120

aagttgaaaa actgggagtt gcgcgccgag acggtcaact cctcctccag aagacggatg 9180

agctcagcga tggtggcgcg cacctcgcgc tcgaaggccc cgggggggctc ctcttcttcc 9240

atctcttcct cctccactaa catctcttct acttcctcct caggaggcgg cggcggggga 9300

ggggccctgc gtcgccggcg gcgcacgggc agacggtcga tgaagcgctc gatggtctcc 9360

ccgcgccggc gacgcatggt ctcggtgacg gcgcgcccgt cctcgcgggg ccgcagcgtg 9420

aagacgccgc cgcgcatctc caggtggccg ccgggggggt ctccgttggg cagggagagg 9480

gcgctgacga tgcatcttat caattggccc gtagggactc cgcgcaagga cctgagcgtc 9540

tcgagatcca cgggatccga aaaccgctga acgaaggctt cgagccagtc gcagtcgcaa 9600

ggtaggctga gcccggtttc ttgttcttcg gggatttcgg gaggcgggcg ggcgatgctg 9660

ctggtgatga agttgaagta ggcggtcctg agacggcgga tggtggcgag gagcaccagg 9720

tccttgggcc cggcttgctg gatgcgcaga cggtcggcca tgccccaggc gtggtcctga 9780

cacctggcga ggtccttgta gtagtcctgc atgagccgct ccacgggcac ctcctcctcg 9840

cccgcgcggc cgtgcatgcg cgtgagcccg aacccgcgct ggggctggac gagcgccagg 9900

tcggcgacga cgcgctcggc gaggatggcc tgctgtatct gggtgagggt ggtctggaag 9960

tcgtcgaagt cgacgaagcg gtggtaggct ccggtgttga tggtatagga gcagttggcc 10020

atgacggacc agttgacggt ctggtggccg ggtcgcacga gctcgtggta cttgaggcgc 10080

gagtaggcgc gcgtgtcgaa gatgtagtcg ttgcaggtgc gcacgaggta ctggtatccg 10140

acgaggaagt gcggcggcgg ctggcggtag agcggccatc gctcggtggc gggggcgccg 10200

ggcgcgaggt cctcgagcat gaggcggtgg tagccgtaga tgtacctgga catccaggtg 10260

atgccggcgg cggtggtgga ggcgcgcggg aactcgcgga cgcggttcca gatgttgcgc 10320

agcggcagga agtagttcat ggtggccgcg gtctggcccg tgaggcgcgc gcagtcgtgg 10380

atgctctaga catacgggca aaaacgaaag cggtcagcgg ctcgactccg tggcctggag 10440

gctaagcgaa cgggttgggc tgcgcgtgta ccccggttcg aatctcgaat caggctggag 10500

ccgcagctaa cgtggtactg gcactcccgt ctcgacccaa gcctgctaac gaaacctcca 10560

ggatacggag gcgggtcgtt ttttggcctt ggtcgctggt catgaaaaac tagtaagcgc 10620

ggaaagcgac cgcccgcgat ggctcgctgc cgtagtctgg agaaagaatc gccagggttg 10680

cgttgcggtg tgccccggtt cgagcctcag cgctcggcgc cggccggatt ccgcggctaa 10740

cgtgggcgtg gctgccccgt cgtttccaag acccccttagc cagccgactt ctccagttac 10800

ggagcgagcc cctcttttc ttgtgttttt gccagatgca tcccgtactg cggcagatgc 10860

gcccccaccc tccacctcaa ccgcccctac cgccgcagca gcagcaacag ccggcgcttc 10920

tgcccccgcc ccagcagcag ccagccacta ccgcggcggc cgccgtgagc ggagccggcg 10980

ttcagtatga cctggccttg gaagagggcg aggggctggc gcggctgggg gcgtcgtcgc 11040

cggagcggca cccgcgcgtg cagatgaaaa gggacgctcg cgaggcctac gtgcccaagc 11100

agaacctgtt cagagacagg agcggcgagg agcccgagga gatgcgcgcc tcccgcttcc 11160

acgcggggcg ggagctgcgg cgcggcctgg accgaaagcg ggtgctgagg gacgaggatt 11220

tcgaggcgga cgagctgacg gggatcagcc ccgcgcgcgc gcacgtggcc gcggccaacc 11280

tggtcacggc gtacgagcag accgtgaagg aggagagcaa cttccaaaaa tccttcaaca 11340

accacgtgcg cacgctgatc gcgcgcgagg aggtgaccct gggcctgatg cacctgtggg 11400

acctgctgga ggccatcgtg cagaacccca cgagcaagcc gctgacggcg cagctgtttc 11460

tggtggtgca gcacagtcgg gacaacgaga cgttcaggga ggcgctgctg aatatcaccg 11520

agcccgaggg ccgctggctc ctggacctgg tgaacattct gcagagcatc gtggtgcagg 11580

agcgcgggct gccgctgtcc gagaagctgg cggctatcaa cttctcggtg ctgagcctgg 11640

gcaagtacta cgctaggaag atctacaaga ccccgtacgt gcccatagac aaggaggtga 11700

agatcgacgg gttttacatg cgcatgaccc tgaaagtgct gaccctgagc gacgatctgg 11760

gggtgtaccg caacgacagg atgcaccgcg cggtgagcgc cagccgccgg cgcgagctga 11820

gcgaccagga gctgatgcac agcctgcagc gggccctgac cggggccggg accgaggggg 11880

agagctactt tgacatgggc gcggacctgc gctggcagcc cagccgccgg gccttggaag 11940

ctgccggcgg ttccccctac gtggaggagg tggacgatga ggaggaggag ggcgagtacc 12000

tggaagactg atggcgcgac cgtatttttg ctagatgcag caacagccac cgcctcctga 12060

tcccgcgatg cgggcggcgc tgcagagcca gccgtccggc attaactcct cggacgattg 12120

gacccaggcc atgcaacgca tcatggcgct gacgacccgc aatcccgaag cctttagaca 12180

gcagcctcag gccaaccggc tctcggccat cctggaggcc gtggtgccct cgcgctcgaa 12240

ccccacgcac gagaaggtgc tggccatcgt gaacgcgctg gtggagaaca aggccatccg 12300

cggcgacgag gccgggctgg tgtacaacgc gctgctggag cgcgtggccc gctacaacag 12360

caccaacgtg cagacgaacc tggaccgcat ggtgaccgac gtgcgcgagg cggtgtcgca 12420

gcgcgagcgg ttccaccgcg agtcgaacct gggctccatg gtggcgctga acgccttcct 12480

gagcacgcag cccgccaacg tgccccgggg ccaggaggac tacaccaact tcatcagcgc 12540

gctgcggctg atggtggccg aggtgcccca gagcgaggtg taccagtcgg ggccggacta 12600

cttcttccag accagtcgcc agggcttgca gaccgtgaac ctgagccagg ctttcaagaa 12660

cttgcaggga ctgtggggcg tgcaggcccc ggtcggggac cgcgcgacgg tgtcgagcct 12720

gctgacgccg aactcgcgcc tgctgctgct gctggtggcg cccttcacgg acagcggcag 12780

cgtgagccgc gactcgtacc tgggctacct gcttaacctg taccgcgagg ccatcgggca 12840

ggcgcacgtg gacgagcaga cctaccagga gatcacccac gtgagccgcg cgctgggcca 12900

ggaggacccg ggcaacctgg aggccaccct gaacttcctg ctgaccaacc ggtcgcagaa 12960

gatcccgccc cagtacgcgc tgagcaccga ggaggagcgc atcctgcgct acgtgcagca 13020

gagcgtgggg ctgttcctga tgcaggaggg ggccacgccc agcgccgcgc tcgacatgac 13080

cgcgcgcaac atggagccca gcatgtacgc tcgcaaccgc ccgttcatca ataagctgat 13140

ggactacttg catcgggcgg ccgccatgaa ctcggactac tttaccaacg ccatcttgaa 13200

cccgcactgg ctcccgccgc ccgggttcta cacgggcgag tacgacatgc ccgaccccaa 13260

cgacgggttc ctgtgggacg acgtggacag cagcgtgttc tcgccgcgcc ccgccaccac 13320

cgtgtggaag aaagagggcg gggaccggcg gccgtcctcg gcgctgtccg gtcgcgcggg 13380

tgctgccgcg gcggtgcctg aggccgccag ccccttcccg agcctgccct tttcgctgaa 13440

cagcgtgcgc agcagcgagc tgggtcggct gacgcggccg cgcctgctgg gcgaggagga 13500

gtacctgaac gactccttgt tgaggcccga gcgcgagaag aacttcccca ataacgggat 13560

agagagcctg gtggacaaga tgagccgctg gaagacgtac gcgcacgagc acagggacga 13620

gccccgagct agcagcagcg caggcacccg tagacgccag cgacacgaca ggcagcgggg 13680

tctggtgtgg gacgatgagg attccgccga cgacagcagc gtgttggact tgggtgggag 13740

```
tggtggtggt aacccgttcg ctcacttgcg cccccgtatc gggcgcctga tgtaagaatc  13800

tgaaaaaata aaaaacggta ctcaccaagg ccatggcgac cagcgtgcgt tcttctctgt  13860

tgtttgtagt agt atg atg agg cgc gtg tac ccg gag ggt cct cct ccc      13909
              Met Met Arg Arg Val Tyr Pro Glu Gly Pro Pro Pro
              1               5                  10

tcg tac gag agc gtg atg cag cag gcg gtg gcg gcg gcg atg cag ccc    13957
Ser Tyr Glu Ser Val Met Gln Gln Ala Val Ala Ala Ala Met Gln Pro
        15                  20                  25

ccg ctg gag gcg cct tac gtg ccc ccg cgg tac ctg gcg cct acg gag    14005
Pro Leu Glu Ala Pro Tyr Val Pro Pro Arg Tyr Leu Ala Pro Thr Glu
    30                  35                  40

ggg cgg aac agc att cgt tac tcg gag ctg gca ccc ttg tac gat acc    14053
Gly Arg Asn Ser Ile Arg Tyr Ser Glu Leu Ala Pro Leu Tyr Asp Thr
45                  50                  55                  60

acc cgg ttg tac ctg gtg gac aac aag tcg gcg gac atc gcc tcg ctg    14101
Thr Arg Leu Tyr Leu Val Asp Asn Lys Ser Ala Asp Ile Ala Ser Leu
                65                  70                  75

aac tac cag aac gac cac agc aac ttc ctg acc acc gtg gtg cag aac    14149
Asn Tyr Gln Asn Asp His Ser Asn Phe Leu Thr Thr Val Val Gln Asn
            80                  85                  90

aac gat ttc acc ccc acg gag gcc agc acc cag acc atc aac ttt gac    14197
Asn Asp Phe Thr Pro Thr Glu Ala Ser Thr Gln Thr Ile Asn Phe Asp
        95                  100                 105

gag cgc tcg cgg tgg ggc ggc cag ctg aaa acc atc atg cac acc aac    14245
Glu Arg Ser Arg Trp Gly Gly Gln Leu Lys Thr Ile Met His Thr Asn
    110                 115                 120

atg ccc aac gtg aac gag ttc atg tac agc aac aag ttc aag gcg cgg    14293
Met Pro Asn Val Asn Glu Phe Met Tyr Ser Asn Lys Phe Lys Ala Arg
125                 130                 135                 140

gtg atg gtc tcg cgc aag acc ccc aat ggg gtc gcg gtg gat gag aat    14341
Val Met Val Ser Arg Lys Thr Pro Asn Gly Val Ala Val Asp Glu Asn
                145                 150                 155

tat gat ggt agt cag gac gag ctg act tac gag tgg gtg gag ttt gag    14389
Tyr Asp Gly Ser Gln Asp Glu Leu Thr Tyr Glu Trp Val Glu Phe Glu
            160                 165                 170

ctg ccc gag ggc aac ttc tcg gtg acc atg acc atc gat ctg atg aac    14437
Leu Pro Glu Gly Asn Phe Ser Val Thr Met Thr Ile Asp Leu Met Asn
        175                 180                 185

aac gcc atc atc gac aac tac ttg gcg gtg ggg cgt cag aac ggg gtg    14485
Asn Ala Ile Ile Asp Asn Tyr Leu Ala Val Gly Arg Gln Asn Gly Val
    190                 195                 200
```

```
ctg gag agc gac atc ggc gtg aag ttc gac acg cgc aac ttc cgg ctg    14533
Leu Glu Ser Asp Ile Gly Val Lys Phe Asp Thr Arg Asn Phe Arg Leu
205                 210                 215                 220

ggc tgg gac ccc gtg acc gag ctg gtg atg ccg ggc gtg tac acc aac    14581
Gly Trp Asp Pro Val Thr Glu Leu Val Met Pro Gly Val Tyr Thr Asn
                225                 230                 235

gag gcc ttc cac ccc gac atc gtc ctg ctg ccc ggc tgc ggc gtg gac    14629
Glu Ala Phe His Pro Asp Ile Val Leu Leu Pro Gly Cys Gly Val Asp
            240                 245                 250

ttc acc gag agc cgc ctc agc aac ctg ctg ggc atc cgc aag cgg cag    14677
Phe Thr Glu Ser Arg Leu Ser Asn Leu Leu Gly Ile Arg Lys Arg Gln
        255                 260                 265

ccc ttc cag gag ggc ttc cag atc ctg tac gag gac ctg gag ggg ggc    14725
Pro Phe Gln Glu Gly Phe Gln Ile Leu Tyr Glu Asp Leu Glu Gly Gly
    270                 275                 280

aac atc ccc gcg ctc ttg gat gtc gaa gcc tat gag aaa agc aag gag    14773
Asn Ile Pro Ala Leu Leu Asp Val Glu Ala Tyr Glu Lys Ser Lys Glu
285                 290                 295                 300

gag gcc gcc gca gcg gcg acc gca gcc gtg gcc acc gcc tct acc gag    14821
Glu Ala Ala Ala Ala Ala Thr Ala Ala Val Ala Thr Ala Ser Thr Glu
                305                 310                 315

gtg cgg ggc gat aat ttt gct agc gcc gcg gca gtg gcc gag gcg gct    14869
Val Arg Gly Asp Asn Phe Ala Ser Ala Ala Ala Val Ala Glu Ala Ala
            320                 325                 330

gaa acc gaa agt aag ata gtc atc cag ccg gtg gag aag gac agc aag    14917
Glu Thr Glu Ser Lys Ile Val Ile Gln Pro Val Glu Lys Asp Ser Lys
        335                 340                 345

gac agg agc tac aac gtg ctc gcg gac aag aaa aac acc gcc tac cgc    14965
Asp Arg Ser Tyr Asn Val Leu Ala Asp Lys Lys Asn Thr Ala Tyr Arg
    350                 355                 360

agc tgg tac ctg gcc tac aac tac ggc gac ccc gag aag ggc gtg cgc    15013
Ser Trp Tyr Leu Ala Tyr Asn Tyr Gly Asp Pro Glu Lys Gly Val Arg
365                 370                 375                 380

tcc tgg acg ctg ctc acc acc tcg gac gtc acc tgc ggc gtg gag caa    15061
Ser Trp Thr Leu Leu Thr Thr Ser Asp Val Thr Cys Gly Val Glu Gln
                385                 390                 395

gtc tac tgg tcg ctg ccc gac atg atg caa gac ccg gtc acc ttc cgc    15109
Val Tyr Trp Ser Leu Pro Asp Met Met Gln Asp Pro Val Thr Phe Arg
            400                 405                 410

tcc acg cgt caa gtt agc aac tac ccg gtg gtg ggc gcc gag ctc ctg    15157
Ser Thr Arg Gln Val Ser Asn Tyr Pro Val Val Gly Ala Glu Leu Leu
        415                 420                 425

ccc gtc tac tcc aag agc ttc ttc aac gag cag gcc gtc tac tcg cag    15205
```

```
Pro Val Tyr Ser Lys Ser Phe Phe Asn Glu Gln Ala Val Tyr Ser Gln
    430                 435                 440

cag ctg cgc gcc ttc acc tcg ctc acg cac gtc ttc aac cgc ttc ccc   15253
Gln Leu Arg Ala Phe Thr Ser Leu Thr His Val Phe Asn Arg Phe Pro
445                 450                 455                 460

gag aac cag atc ctc gtc cgc ccg ccc gcg ccc acc att acc acc gtc   15301
Glu Asn Gln Ile Leu Val Arg Pro Pro Ala Pro Thr Ile Thr Thr Val
                465                 470                 475

agt gaa aac gtt cct gct ctc aca gat cac ggg acc ctg ccg ctg cgc   15349
Ser Glu Asn Val Pro Ala Leu Thr Asp His Gly Thr Leu Pro Leu Arg
            480                 485                 490

agc agt atc cgg gga gtc cag cgc gtg acc gtc act gac gcc aga cgc   15397
Ser Ser Ile Arg Gly Val Gln Arg Val Thr Val Thr Asp Ala Arg Arg
        495                 500                 505

cgc acc tgc ccc tac gtc tac aag gcc ctg ggc gta gtc gcg ccg cgc   15445
Arg Thr Cys Pro Tyr Val Tyr Lys Ala Leu Gly Val Val Ala Pro Arg
    510                 515                 520

gtc ctc tcg agc cgc acc ttc taa aaaatgtcca ttctcatctc gcccagtaat   15499
Val Leu Ser Ser Arg Thr Phe
525                 530

aacaccggtt ggggcctgcg cgcgcccagc aagatgtacg gaggcgctcg ccaacgctcc   15559

acgcaacacc ccgtgcgcgt gcgcgggcac ttccgcgctc cctggggcgc cctcaagggc   15619

cgcgtgcgct cgcgcaccac cgtcgacgac gtgatcgacc aggtggtggc cgacgcgcgc   15679

aactacacgc ccgccgccgc gcccgcctcc accgtggacg ccgtcatcga cagcgtggtg   15739

gccgatgcgc gccggtacgc ccgcgccaag agccggcggc ggcgcatcgc ccggcggcac   15799

cggagcaccc ccgccatgcg cgcggcgcga gccttgctgc gcagggccag gcgcacggga   15859

cgcagggcca tgctcagggc ggccagacgc gcggcctccg gcagcagcag cgccggcagg   15919

acccgcagac gcgcggccac ggcggcggcg gcggccatcg ccagcatgtc ccgcccgcgg   15979

cgcggcaacg tgtactgggt gcgcgacgcc gccaccggtg tgcgcgtgcc cgtgcgcacc   16039

cgcccccctc gcacttgaag atgctgactt cgcgatgttg atgtgtccca gcggcgagga   16099

ggatgtccaa gcgcaaatac aaggaagaga tgctccaggt catcgcgcct gagatctacg   16159

gccccgcggt gaaggaggaa agaaagcccc gcaaactgaa gcgggtcaaa aaggacaaaa   16219

aggaggagga agatgtggac ggactggtgg agtttgtgcg cgagttcgcc ccccggcggc   16279

gcgtgcagtg gcgcgggcgg aaagtgaaac cggtgctgcg gcccggcacc acggtggtct   16339

tcacgcccgg cgagcgttcc ggctccgcct ccaagcgctc ctacgacgag gtgtacgggg   16399
```

acgaggacat cctcgagcag gcggtcgagc gtctgggcga gtttgcttac ggcaagcgca 16459

gccgccccgc gcccttgaaa gaggaggcgg tgtccatccc gctggaccac ggcaacccca 16519

cgccgagcct gaagccggtg accctgcagc aggtgctgcc gagcgcggcg ccgcgccggg 16579

gcttcaagcg cgagggcggc gaggatctgt acccgaccat gcagctgatg gtgcccaagc 16639

gccagaagct ggaggacgtg ctggagcaca tgaaggtgga ccccgaggtg cagcccgagg 16699

tcaaggtgcg gcccatcaag caggtggccc cgggcctggg cgtgcagacc gtggacatca 16759

agatccccac ggagcccatg gaaacgcaga ccgagcccgt gaagcccagc accagcacca 16819

tggaggtgca gacggatccc tggatgccgg cgccggcttc caccactcgc cgaagacgca 16879

agtacggcgc ggccagcctg ctgatgccca actacgcgct gcatccttcc atcatcccca 16939

cgccgggcta ccgcggcacg cgcttctacc gcggctacac cagcagccgc cgcaagacca 16999

ccacccgccg ccgccgtcgt cgcacccgcc gcagcagcac cgcgacttcc gccgccgccc 17059

tggtgcggag agtgtaccgc agcgggcgcg agcctctgac cctgccgcgc gcgcgctacc 17119

acccgagcat cgccatttaa ctctgccgtc gcctcctact tgcagatatg gccctcacat 17179

gccgcctccg cgtccccatt acgggctacc gaggaagaaa gccgcgccgt agaaggctga 17239

cggggaacgg gctgcgtcgc catcaccacc ggcggcggcg cgccatcagc aagcggttgg 17299

ggggaggctt cctgcccgcg ctgatcccca tcatcgccgc ggcgatcggg gcgatccccg 17359

gcatagcttc cgtggcggtg caggcctctc agcgccactg agacacagct tggaaaattt 17419

gtaataaaaa aatggactga cgctcctggt cctgtgatgt gtgtttttag atggaagaca 17479

tcaatttttc gtccctggca ccgcgacacg gcacgcggcc gtttatgggc acctggagcg 17539

acatcggcaa cagccaactg aacgggggcg ccttcaattg gagcagtctc tggagcgggc 17599

ttaagaattt cgggtccacg ctcaaaacct atggcaacaa ggcgtggaac agcagcacag 17659

ggcaggcgct gagggaaaag ctgaaagagc agaacttcca gcagaaggtg gtcgatggcc 17719

tggcctcggg catcaacggg gtggtggacc tggccaacca ggccgtgcag aaacagatca 17779

acagccgcct ggacgcggtc ccgcccgcgg ggtccgtgga gatgccccag gtggaggagg 17839

agctgcctcc cctggacaag cgcggcgaca agcgaccgcg tcccgacgcg gaggagacgc 17899

tgctgacgca cacggacgag ccgcccccgt acgaggaggc ggtgaaactg ggtctgccca 17959

ccacgcggcc cgtggcgcct ctggccaccg gggtgctgaa acccagcagc agcagccagc 18019

ccgcgaccct ggacttgcct ccgcctgctt cccgcccctc cacagtggct aagcccctgc 18079

cgccggtggc cgtcgcgtcg cgcgcccccc gaggccgccc ccaggcgaac tggcagagca 18139

```
ctctgaacag catcgtgggt ctgggagtgc agagtgtgaa gcgccgccgc tgctattaaa  18199

agacactgta gcgcttaact tgcttgtctg tgtgtatatg tatgtccgcc gaccagaagg  18259

aggaagaggc gcgtcgccga gttgcaag atg gcc acc cca tcg atg ctg ccc      18311
                               Met Ala Thr Pro Ser Met Leu Pro
                                                   535

cag tgg gcg tac atg cac atc gcc gga cag gac gct tcg gag tac ctg    18359
Gln Trp Ala Tyr Met His Ile Ala Gly Gln Asp Ala Ser Glu Tyr Leu
540                 545                 550                 555

agt ccg ggt ctg gtg cag ttc gcc cgc gcc aca gac acc tac ttc agt    18407
Ser Pro Gly Leu Val Gln Phe Ala Arg Ala Thr Asp Thr Tyr Phe Ser
                560                 565                 570

ctg ggg aac aag ttt agg aac ccc acg gtg gcg ccc acg cac gat gtg    18455
Leu Gly Asn Lys Phe Arg Asn Pro Thr Val Ala Pro Thr His Asp Val
            575                 580                 585

acc acc gac cgc agc cag cgg ctg acg ctg cgc ttc gtg ccc gtg gac    18503
Thr Thr Asp Arg Ser Gln Arg Leu Thr Leu Arg Phe Val Pro Val Asp
        590                 595                 600

cgc gag gac aac acc tac tcg tac aaa gtg cgc tac acg ctg gcc gtg    18551
Arg Glu Asp Asn Thr Tyr Ser Tyr Lys Val Arg Tyr Thr Leu Ala Val
    605                 610                 615

ggc gac aac cgc gtg ctg gac atg gcc agc acc tac ttt gac atc cgc    18599
Gly Asp Asn Arg Val Leu Asp Met Ala Ser Thr Tyr Phe Asp Ile Arg
620                 625                 630                 635

ggc gtg ctg gat cgg ggg ccc agc ttc aaa ccc tac tcc ggc acc gcc    18647
Gly Val Leu Asp Arg Gly Pro Ser Phe Lys Pro Tyr Ser Gly Thr Ala
                640                 645                 650

tac aac agc ctg gct ccc aag gga gcg ccc aac act tgc cag tgg aca    18695
Tyr Asn Ser Leu Ala Pro Lys Gly Ala Pro Asn Thr Cys Gln Trp Thr
            655                 660                 665

tat aaa gct ggt gat act gat aca gaa aaa acc tat aca tat gga aat    18743
Tyr Lys Ala Gly Asp Thr Asp Thr Glu Lys Thr Tyr Thr Tyr Gly Asn
        670                 675                 680

gca cct gtg caa ggc att agc att aca aag gat ggt att caa ctt gga    18791
Ala Pro Val Gln Gly Ile Ser Ile Thr Lys Asp Gly Ile Gln Leu Gly
    685                 690                 695

act gac agc gat ggt cag gca atc tat gca gac gaa act tat caa cca    18839
Thr Asp Ser Asp Gly Gln Ala Ile Tyr Ala Asp Glu Thr Tyr Gln Pro
700                 705                 710                 715

gag cct caa gtg ggt gat gct gaa tgg cat gac atc act ggt act gat    18887
Glu Pro Gln Val Gly Asp Ala Glu Trp His Asp Ile Thr Gly Thr Asp
                720                 725                 730
```

```
gaa aaa tat gga ggc aga gct ctt aag cct gac acc aaa atg aag cct    18935
Glu Lys Tyr Gly Gly Arg Ala Leu Lys Pro Asp Thr Lys Met Lys Pro
            735                 740                 745

tgc tat ggt tct ttt gcc aag cct acc aat aaa gaa gga ggc cag gca    18983
Cys Tyr Gly Ser Phe Ala Lys Pro Thr Asn Lys Glu Gly Gly Gln Ala
        750                 755                 760

aat gtg aaa acc gaa aca ggc ggt acc aaa gaa tat gac att gac atg    19031
Asn Val Lys Thr Glu Thr Gly Gly Thr Lys Glu Tyr Asp Ile Asp Met
    765                 770                 775

gca ttc ttc gat aat cga agt gca gct gcc gcc ggc cta gcc cca gaa    19079
Ala Phe Phe Asp Asn Arg Ser Ala Ala Ala Ala Gly Leu Ala Pro Glu
780                 785                 790                 795

att gtt ttg tat act gag aat gtg gat ctg gaa act cca gat acc cat    19127
Ile Val Leu Tyr Thr Glu Asn Val Asp Leu Glu Thr Pro Asp Thr His
                800                 805                 810

att gta tac aag gca ggt aca gat gac agt agc tct tct atc aat ttg    19175
Ile Val Tyr Lys Ala Gly Thr Asp Asp Ser Ser Ser Ser Ile Asn Leu
            815                 820                 825

ggt cag cag tcc atg ccc aac aga ccc aac tac att ggc ttc aga gac    19223
Gly Gln Gln Ser Met Pro Asn Arg Pro Asn Tyr Ile Gly Phe Arg Asp
        830                 835                 840

aac ttt atc ggt ctg atg tac tac aac agc act ggc aat atg ggt gta    19271
Asn Phe Ile Gly Leu Met Tyr Tyr Asn Ser Thr Gly Asn Met Gly Val
    845                 850                 855

ctg gct gga cag gcc tcc cag ctg aat gct gtg gtg gac ttg cag gac    19319
Leu Ala Gly Gln Ala Ser Gln Leu Asn Ala Val Val Asp Leu Gln Asp
860                 865                 870                 875

aga aac acc gaa ctg tcc tac cag ctc ttg ctt gac tct ctg ggt gac    19367
Arg Asn Thr Glu Leu Ser Tyr Gln Leu Leu Leu Asp Ser Leu Gly Asp
                880                 885                 890

aga acc agg tat ttc agt atg tgg aat cag gcg gtg gac agt tat gac    19415
Arg Thr Arg Tyr Phe Ser Met Trp Asn Gln Ala Val Asp Ser Tyr Asp
            895                 900                 905

ccc gat gtg cgc att att gaa aat cac ggt gtg gag gat gaa ctt cct    19463
Pro Asp Val Arg Ile Ile Glu Asn His Gly Val Glu Asp Glu Leu Pro
        910                 915                 920

aac tat tgc ttc ccc ctg gat gct gtg ggt aga act gat act tac cag    19511
Asn Tyr Cys Phe Pro Leu Asp Ala Val Gly Arg Thr Asp Thr Tyr Gln
    925                 930                 935

gga att aag gcc aat ggt gat aat caa acc acc tgg acc aaa gat gat    19559
Gly Ile Lys Ala Asn Gly Asp Asn Gln Thr Thr Trp Thr Lys Asp Asp
940                 945                 950                 955

act gtt aat gat gct aat gaa ttg ggc aag ggc aat cct ttc gcc atg    19607
```

```
Thr Val Asn Asp Ala Asn Glu Leu Gly Lys Gly Asn Pro Phe Ala Met
                960                 965                 970

gag atc aac atc cag gcc aac ctg tgg cgg aac ttc ctc tac gcg aac    19655
Glu Ile Asn Ile Gln Ala Asn Leu Trp Arg Asn Phe Leu Tyr Ala Asn
            975                 980                 985

gtg gcg ctg tac ctg ccc gac tcc tac aag tac acg ccg  gcc aac atc   19703
Val Ala Leu Tyr Leu Pro Asp Ser Tyr Lys Tyr Thr Pro  Ala Asn Ile
        990                 995                 1000

acg ctg  ccc acc aac acc aac  acc tac gat tac atg  aac ggc cgc     19748
Thr Leu  Pro Thr Asn Thr Asn  Thr Tyr Asp Tyr Met  Asn Gly Arg
    1005                1010                1015

gtg gtg  gcg ccc tcg ctg gtg  gac gcc tac atc aac  atc ggg gcg     19793
Val Val  Ala Pro Ser Leu Val  Asp Ala Tyr Ile Asn  Ile Gly Ala
    1020                1025                1030

cgc tgg  tcg ctg gac ccc atg  gac aac gtc aac ccc  ttc aac cac     19838
Arg Trp  Ser Leu Asp Pro Met  Asp Asn Val Asn Pro  Phe Asn His
    1035                1040                1045

cac cgc  aac gcg ggc ctg cga  tac cgc tcc atg ctc  ctg ggc aac     19883
His Arg  Asn Ala Gly Leu Arg  Tyr Arg Ser Met Leu  Leu Gly Asn
    1050                1055                1060

ggg cgc  tac gtg ccc ttc cac  atc cag gtg ccc caa  aag ttt ttc     19928
Gly Arg  Tyr Val Pro Phe His  Ile Gln Val Pro Gln  Lys Phe Phe
    1065                1070                1075

gcc atc  aag agc ctc ctg ctc  ctg ccc ggg tcc tac  acc tac gag     19973
Ala Ile  Lys Ser Leu Leu Leu  Leu Pro Gly Ser Tyr  Thr Tyr Glu
    1080                1085                1090

tgg aac  ttc cgc aag gac gtc  aac atg atc ctg cag  agc tcc ctc     20018
Trp Asn  Phe Arg Lys Asp Val  Asn Met Ile Leu Gln  Ser Ser Leu
    1095                1100                1105

ggc aac  gac ctg cgc acg gac  ggg gcc tcc atc gcc  ttc acc agc     20063
Gly Asn  Asp Leu Arg Thr Asp  Gly Ala Ser Ile Ala  Phe Thr Ser
    1110                1115                1120

atc aac  ctc tac gcc acc ttc  ttc ccc atg gcg cac  aac acc gcc     20108
Ile Asn  Leu Tyr Ala Thr Phe  Phe Pro Met Ala His  Asn Thr Ala
    1125                1130                1135

tcc acg  ctc gag gcc atg ctg  cgc aac gac acc aac  gac cag tcc     20153
Ser Thr  Leu Glu Ala Met Leu  Arg Asn Asp Thr Asn  Asp Gln Ser
    1140                1145                1150

ttc aac  gac tac ctc tcg gcg  gcc aac atg ctc tac  ccc atc ccg     20198
Phe Asn  Asp Tyr Leu Ser Ala  Ala Asn Met Leu Tyr  Pro Ile Pro
    1155                1160                1165

gcc aac  gcc acc aac gtg ccc  atc tcc atc ccc tcg  cgc aac tgg     20243
Ala Asn  Ala Thr Asn Val Pro  Ile Ser Ile Pro Ser  Arg Asn Trp
```

```
    1170                1175                1180

gcc gcc  ttc cgc ggc tgg tcc  ttc acg cgc ctc aag  acc cgc gag     20288
Ala Ala  Phe Arg Gly Trp Ser  Phe Thr Arg Leu Lys  Thr Arg Glu
    1185                1190                1195

acg ccc  tcg ctc ggc tcc ggg  ttc gac ccc tac ttc  gtc tac tcg     20333
Thr Pro  Ser Leu Gly Ser Gly  Phe Asp Pro Tyr Phe  Val Tyr Ser
    1200                1205                1210

ggc tcc  atc ccc tac ctc gac  ggc acc ttc tac ctc  aac cac acc     20378
Gly Ser  Ile Pro Tyr Leu Asp  Gly Thr Phe Tyr Leu  Asn His Thr
    1215                1220                1225

ttc aag  aag gtc tcc atc acc  ttc gac tcc tcc gtc  agc tgg ccc     20423
Phe Lys  Lys Val Ser Ile Thr  Phe Asp Ser Ser Val  Ser Trp Pro
    1230                1235                1240

ggc aac  gac cgc ctc ctg acg  ccc aac gag ttc gaa  atc aag cgc     20468
Gly Asn  Asp Arg Leu Leu Thr  Pro Asn Glu Phe Glu  Ile Lys Arg
    1245                1250                1255

acc gtc  gac gga gag ggg tac  aac gtg gcc cag tgc  aac atg acc     20513
Thr Val  Asp Gly Glu Gly Tyr  Asn Val Ala Gln Cys  Asn Met Thr
    1260                1265                1270

aag gac  tgg ttc ctg gtc cag  atg ctg gcc cac tac  aac atc ggc     20558
Lys Asp  Trp Phe Leu Val Gln  Met Leu Ala His Tyr  Asn Ile Gly
    1275                1280                1285

tac cag  ggc ttc tac gtg ccc  gag ggc tac aag gac  cgc atg tac     20603
Tyr Gln  Gly Phe Tyr Val Pro  Glu Gly Tyr Lys Asp  Arg Met Tyr
    1290                1295                1300

tcc ttc  ttc cgc aac ttc cag  ccc atg agc cgc cag  gtc gtg gac     20648
Ser Phe  Phe Arg Asn Phe Gln  Pro Met Ser Arg Gln  Val Val Asp
    1305                1310                1315

gag gtc  aac tac aag gac tac  cag gcc gtc acc ctg  gcc tac cag     20693
Glu Val  Asn Tyr Lys Asp Tyr  Gln Ala Val Thr Leu  Ala Tyr Gln
    1320                1325                1330

cac aac  aac tcg ggc ttc gtc  ggc tac ctc gcg ccc  acc atg cgc     20738
His Asn  Asn Ser Gly Phe Val  Gly Tyr Leu Ala Pro  Thr Met Arg
    1335                1340                1345

cag ggc  cag ccc tac ccc gcc  aac tac ccc tac ccg  ctc atc ggc     20783
Gln Gly  Gln Pro Tyr Pro Ala  Asn Tyr Pro Tyr Pro  Leu Ile Gly
    1350                1355                1360

aag agc  gcc gtc gcc agc gtc  acc cag aaa aag ttc  ctc tgc gac     20828
Lys Ser  Ala Val Ala Ser Val  Thr Gln Lys Lys Phe  Leu Cys Asp
    1365                1370                1375

cgg gtc  atg tgg cgc atc ccc  ttc tcc agc aac ttc  atg tcc atg     20873
Arg Val  Met Trp Arg Ile Pro  Phe Ser Ser Asn Phe  Met Ser Met
    1380                1385                1390
```

```
ggc gcg  ctc acc gac ctc ggc  cag aac atg ctc tac  gcc aac tcc     20918
Gly Ala  Leu Thr Asp Leu Gly  Gln Asn Met Leu Tyr  Ala Asn Ser
    1395                 1400                 1405

gcc cac  gcg cta gac atg aat  ttc gaa gtc gac ccc  atg gat gag     20963
Ala His  Ala Leu Asp Met Asn  Phe Glu Val Asp Pro  Met Asp Glu
    1410                 1415                 1420

tcc acc  ctt ctc tat gtt gtc  ttc gaa gtc ttc gac  gtc gtc cga     21008
Ser Thr  Leu Leu Tyr Val Val  Phe Glu Val Phe Asp  Val Val Arg
    1425                 1430                 1435

gtg cac  cag ccc cac cgc ggc  gtc atc gag gcc gtc  tac ctg cgc     21053
Val His  Gln Pro His Arg Gly  Val Ile Glu Ala Val  Tyr Leu Arg
    1440                 1445                 1450

acg ccc  ttc tcg gcc ggc aac  gcc acc acc taa gcctcttgct          21096
Thr Pro  Phe Ser Ala Gly Asn  Ala Thr Thr
    1455                 1460

tcttgcaaga tgacggcctg cgcgggctcc ggcgagcagg agctcagggc catcctccgc  21156

gacctgggct gcgggccctg cttcctgggc accttcgaca agcgcttccc gggattcatg  21216

gccccgcaca agctggcctg cgccatcgtc aacacggccg gccgcgagac cgggggcgag  21276

cactggctgg ccttcgcctg gaacccgcgc tcccacacct gctacctctt cgaccccttc  21336

gggttctcgg acgagcgcct caagcagatc taccagttcg agtacgaggg cctgctgcgt  21396

cgcagcgccc tggccaccga ggaccgctgc gtcaccctgg aaaagtccac ccagaccgtg  21456

cagggtccgc gctcggccgc ctgcgggctc ttctgctgca tgttcctgca cgccttcgtg  21516

cactggcccg accgccccat ggacaagaac cccaccatga acttgctgac gggggtgccc  21576

aacggcatgc tccagtcgcc ccaggtggaa cccaccctgc gccgcaacca ggaggcgctc  21636

taccgcttcc tcaacgccca ctccgcctac tttcgctccc accgcgcgcg catcgagaag  21696

gccaccgcct tcgaccgcat gaatcaagac atgtaatccg gtgtgtgtat gtgaatgctt  21756

tattcatcat aataaacagc acatgtttat gccaccttct ctgaggctct gactttattt  21816

agaaatcgaa ggggttctgc cggctctcgg catggcccgc gggcagggat acgttgcgga  21876

actggtactt gggcagccac ttgaactcgg ggatcagcag cttcggcacg gggaggtcgg  21936

ggaacgagtc gctccacagc ttgcgcgtga gttgcagggc gcccagcagg tcgggcgcgg  21996

agatcttgaa atcgcagttg ggacccgcgt tctgcgcgcg agagttacgg tacacggggt  22056

tgcagcactg gaacaccatc agggccgggt gcttcacgct cgccagcacc gtcgcgtcgg  22116

tgatgccctc cacgtccaga tcctcggcgt tggccatccc gaaggggggtc atcttgcagg  22176
```

tctgccgccc catgctgggc acgcagccgg gcttgtggtt gcaatcgcag tgcaggggga 22236

tcagcatcat ctgggcctgc tcggagctca tgcccgggta catggccttc atgaaagcct 22296

ccagctggcg gaaggcctgc tgcgccttgc cgccctcggt gaagaagacc ccgcaggact 22356

tgctagagaa ctggttggtg gcgcagccag cgtcgtgcac gcagcagcgc gcgtcgttgt 22416

tggccagctg caccacgctg cgcccccagc ggttctgggt gatcttggcc cggtcggggt 22476

tctccttcag cgcgcgctgc ccgttctcgc tcgccacatc catctcgatc gtgtgctcct 22536

tctggatcat cacggtcccg tgcaggcacc gcagcttgcc ctcggcctcg gtgcacccgt 22596

gcagccacag cgcgcagccg gtgctctccc agttcttgtg ggcgatctgg gagtgcgagt 22656

gcacgaagcc ctgcaggaag cggcccatca tcgtggtcag ggtcttgttg ctggtgaagg 22716

tcagcggaat gccgcggtgc tcctcgttca catacaggtg gcagatacgg cggtacacct 22776

cgccctgctc gggcatcagc tggaaggcgg acttcaggtc gctctccacg cggtaccggt 22836

ccatcagcag cgtcatcact tccatgccct tctcccaggc cgaaacgatc ggcaggctca 22896

gggggttctt caccgttgtc atcttagtcg ccgccgccga agtcaggggg tcgttctcgt 22956

ccagggtctc aaacactcgc ttgccgtcct tctcggtgat gcgcacgggg ggaaagctga 23016

agcccacggc cgccagctcc tcctcggcct gcctttcgtc ctcgctgtcc tggctgatgt 23076

cttgcaaagg cacatgcttg gtcttgcggg gtttcttttt gggcggcaga ggcggcggcg 23136

gagacgtgct gggcgagcgc gagttctcgc tcaccacgac tatttcttct ccttggccgt 23196

cgtccgagac cacgcggcgg taggcatgcc tcttctgggg cagaggcgga ggcgacgggc 23256

tctcgcggtt cggcgggcgg ctggcagagc cccttccgcg ttcgggggtg cgctcctggc 23316

ggcgctgctc tgactgactt cctccgcggc cggccattgt gttctcctag ggagcaagca 23376

tggagactca gccatcgtcg ccaacatcgc catctgcccc cgccgccgcc gacgagaacc 23436

agcagcagca gaatgaaagc ttaaccgccc cgccgcccag ccccacctcc gacgccgcag 23496

ccccagacat gcaagagatg gaggaatcca tcgagattga cctgggctac gtgacgcccg 23556

cggagcacga ggaggagctg gcagcgcgct tttcagcccc ggaagagaac caccaagagc 23616

agccagagca ggaagcagag agcgagcaga accaggctgg gctcgagcat ggcgactacc 23676

tgagcggggc agaggacgtg ctcatcaagc atctggcccg ccaatgcatc atcgtcaagg 23736

acgcgctgct cgaccgcgcc gaggtgcccc tcagcgtggc ggagctcagc cgcgcctacg 23796

agcgcaacct cttctcgccg cgcgtgcccc ccaagcgcca gcccaacggc acctgcgagc 23856

ccaacccgcg cctcaacttc tacccggtct tcgcggtgcc cgaggccctg gccacctacc 23916

acctcttttt caagaaccaa aggatccccg tctcctgccg cgccaaccgc acccgcgccg 23976

acgccctgct caacctgggc cccggcgccc gcctacctga tatcgcctcc ttggaagagg 24036

ttcccaagat cttcgagggt ctgggcagcg acgagactcg ggccgcgaac gctctgcaag 24096

gaagcggaga ggagcatgag caccacagcg ccctggtgga gttggaaggc gacaacgcgc 24156

gcctggcggt cctcaagcgc acggtcgagc tgacccactt cgcctacccg gcgctcaacc 24216

tgcccccaa ggtcatgagc gccgtcatgg accaggtgct catcaagcgc gcctcgcccc 24276

tctcggagga ggagatgcag gaccccgaga gctcggacga gggcaagccc gtggtcagcg 24336

acgagcagct ggcgcgctgg ctgggagcga gtagcacccc ccagagcctg gaagagcggc 24396

gcaagctcat gatggccgtg gtcctggtga ccgtggagct ggagtgtctg cgccgcttct 24456

tcgccgacgc ggagaccctg cgcaaggtcg aggagaacct gcactacctc ttcagacacg 24516

ggttcgtgcg ccaggcctgc aagatctcca acgtggagct gaccaacctg gtctcctaca 24576

tgggcatcct gcacgagaac cgcctggggc agaacgtgct gcacaccacc ctgcgcgggg 24636

aggcccgccg cgactacatc cgcgactgcg tctacctgta cctctgccac acctggcaga 24696

cgggcatggg cgtgtggcag cagtgcctgg aggagcagaa cctgaaagag ctctgcaagc 24756

tcctgcagaa gaacctcaag gccctgtgga ccgggttcga cgagcgcacc accgccgcgg 24816

acctggccga cctcatcttc cccgagcgcc tgcggctgac gctgcgcaac gggctgcccg 24876

actttatgag ccaaagcatg ttgcaaaact ttcgctcttt catcctcgaa cgctccggga 24936

tcctgcccgc cacctgctcc gcgctgccct cggacttcgt gccgctgacc ttccgcgagt 24996

gccccccgcc gctctggagc cactgctacc tgctgcgcct ggccaactac ctggcctacc 25056

actcggacgt gatcgaggac gtcagcggcg agggcctgct cgagtgccac tgccgctgca 25116

acctctgcac gccgcaccgc tccctggcct gcaaccccca gctgctgagc gagacccaga 25176

tcatcggcac cttcgagttg caaggccccg gcgagggcaa ggggggtctg aaactcaccc 25236

cggggctgtg gacctcggcc tacttgcgca agttcgtgcc cgaggactac catcccttcg 25296

agatcaggtt ctacgaggac caatcccagc cgcccaaggc cgagctgtcg gcctgcgtca 25356

tcacccaggg ggccatcctg gcccaattgc aagccatcca gaaatcccgc caagaatttc 25416

tgctgaaaaa gggccacggg gtctacttgg acccccagac cggagaggag ctcaacccca 25476

gcttcccccа ggatgccccg aggaagcagc aagaagctga aagtggagct gccgccgccg 25536

ccggaggatt tggaggaaga ctgggagagc agtcaggcag aggaggagga gatggaagac 25596

tgggacagca ctcaggcaga ggaggacagc ctgcaagaca gtctggagga ggaagacgag 25656

gtggaggagg cagaggaaga agcagccgcc gccagaccgt cgtcctcggc ggaggaggag 25716

aaagcaagca gcacggatac catctccgct ccgggtcggg gtcgcggcgg ccgggcccac 25776

agtagatggg acgagaccgg gcgcttcccg aaccccacca cccagaccgg taagaaggag 25836

cggcagggat acaagtcctg gcgggggcac aaaaacgcca tcgtctcctg cttgcaagcc 25896

tgcgggggca acatctcctt cacccggcgc tacctgctct tccaccgcgg ggtgaacttc 25956

ccccgcaaca tcttgcatta ctaccgtcac ctccacagcc cctactactg tttccaagaa 26016

gaggcagaaa cccagcagca gcagcagcag cagaaaacca gcggcagcag ctagaaaatc 26076

cacagcggcg gcaggtggac tgaggatcgc ggcgaacgag ccggcgcaga cccgggagct 26136

gaggaaccgg atctttccca ccctctatgc catcttccag cagagtcggg ggcaagagca 26196

ggaactgaaa gtcaagaacc gttctctgcg ctcgctcacc cgcagttgtc tgtatcacaa 26256

gagcgaagac caacttcagc gcactctcga ggacgccgag gctctcttca acaagtactg 26316

cgcgctcact cttaaagagt agcccgcgcc cgcccacaca cggaaaaagg cgggaattac 26376

gtcaccacct gcgcccttcg cccgaccatc atcatgagca aagagattcc cacgccttac 26436

atgtggagct accagcccca gatgggcctg gccgccggcg ccgcccagga ctactccacc 26496

cgcatgaact ggctcagtgc cgggcccgcg atgatctcac gggtgaatga catccgcgcc 26556

caccgaaacc agatactcct agaacagtca gcgatcaccg ccacgccccg ccatcacctt 26616

aatccgcgta attggcccgc cgccctggtg taccaggaaa ttccccagcc cacgaccgta 26676

ctacttccgc gagacgccca ggccgaagtc cagctgacta actcaggtgt ccagctggcc 26736

ggcggcgccg ccctgtgtcg tcaccgcccc gctcagggta taaagcggct ggtgatccga 26796

ggcagaggca cacagctcaa cgacgaggtg gtgagctctt cgctgggtct gcgacctgac 26856

ggagtcttcc aactcgccgg atcggggaga tcttccttca cgcctcgtca ggccgtcctg 26916

actttggaga gttcgtcctc gcagccccgc tcgggtggca tcggcactct ccagttcgtg 26976

gaggagttca ctccctcggt ctacttcaac cccttctccg gctcccccgg ccactacccg 27036

gacgagttca tcccgaactt cgacgccatc agcgagtcgg tggacggcta cgattgaatg 27096

tcccatggtg gcgcggctga cctagctcgg cttcgacacc tggaccactg ccgccgcttc 27156

cgctgcttcg ctcgggatct cgccgagttt gcctactttg agctgcccga ggagcaccct 27216

cagggcccgg cccacggagt gcggatcgtc gtcgaagggg gtctcgactc ccacctgctt 27276

cggatcttca gccagcgtcc gatcctggcc gagcgcgagc aaggacagac ccttctgacc 27336

ctgtactgca tctgcaacca ccccggcctg catgaaagtc tttgttgtct gctgtgtact 27396

gagtataata aaagctgaga tcagcgacta ctccggactt ccgtgtgttc ctgctatcaa 27456

ccagtccctg ttcttcaccg ggaacgagac cgagctccag ctccagtgta agccccacaa 27516

gaagtacctc acctggctgt tccagggctc tccgatcgcc gttgtcaacc actgcgacaa 27576

cgacggagtc ctgctgagcg gccctgccaa ccttactttt tccacccgca gaagcaagct 27636

ccagctcttc caacccttcc tccccgggac ctatcagtgc gtctcgggac cctgccatca 27696

caccttccac ctgatcccga ataccacagc gtcgctcccc gctactaaca accaaactac 27756

ccaccaacgc caccgtcgcg acctttcctc tgggtctaat accactaccg gaggtgagct 27816

ccgaggtcga ccaacctctg ggatttacta cggcccctgg gaggtggtag ggttaatagc 27876

gctaggccta gttgcgggtg ggcttttggc tctctgctac ctatacctcc cttgctgttc 27936

gtacttagtg gtgctgtgtt gctggtttaa gaaatgggga agatcaccct agtgagctgc 27996

ggtgtgctgg tggcggtggt gctttcgatt gtgggactgg gcggcgcggc tgtagtgaag 28056

gagaaggccg atccctgctt gcatttcaat cccgacaaat gccagctgag ttttcagccc 28116

gatggcaatc ggtgcgcggt gctgatcaag tgcggatggg aatgcgagaa cgtgagaatc 28176

gagtacaata acaagactcg gaacaatact ctcgcgtccg tgtggcagcc cggggacccc 28236

gagtggtaca ccgtctctgt ccccggtgct gacggctccc cgcgcaccgt gaataatact 28296

ttcatttttg cgcacatgtg cgacacggtc atgtggatga gcaagcagta cgatatgtgg 28356

ccccccacga aggagaacat cgtggtcttc tccatcgctt acagcgtgtg cacggcgcta 28416

atcaccgcta tcgtgtgcct gagcattcac atgctcatcg ctattcgccc cagaaataat 28476

gccgaaaaag aaaaacagcc ataacacgtt ttttcacaca ccttttcag accatggcct 28536

ctgttaaatt tttgctttta tttgccagtc tcattgccgt cattcatgga atgagtaatg 28596

agaaaattac tatttacact ggcactaatc acacattgaa aggtccagaa aaagccacag 28656

aagtttcatg gtattgttat tttaatgaat cagatgtatc tactgaactc tgtggaaaca 28716

ataacaaaaa aaatgagagc attactctca tcaagtttca atgtggatct gacttaaccc 28776

taattaacat cactagagac tatgtaggta tgtattatgg aactacagca ggcatttcgg 28836

acatggaatt ttatcaagtt tctgtgtctg aacccaccac gcctagaatg accacaacca 28896

caaaaactac acctgttacc actatacagc tcactaccaa tggctttctt gccatgcttc 28956

aagtggctga aaatagcacc agcattcaac ccaccccacc cagtgaggaa attcccagat 29016

ccatgattgg cattattgtt gctgtagtgg tgtgcatgtt gatcatcgcc ttgtgcatgg 29076

tgtactatgc cttctgctac agaaagcaca gactgaacga caagctggaa cacttactaa 29136

gtgttgaatt ttaatttttt agaaccatga agatcctagg ccttttagtt ttttctatca 29196

ttacctctgc tctatgcaat tctgacaatg aggacgttac tgtcgttgtc ggatcaaatt 29256

atacactaaa aggtccagca aaaggtatgc tttcgtggta ttgttggttc ggaactgacg 29316

agcaacagac agaactttgc aatgctcaaa aaggcaaaac ctcaaattct aaaatctcta 29376

attatcaatg caatggcact gacttagtat tgctcaatgt cacgaaagca tatgctggca 29436

gttacacctg ccctggagat gatgccgaca atatgatttt ttacaaagtg gaagtggttg 29496

atcccactac tccaccgccc accaccacaa ctactcatac cacacacaca gaacaaacac 29556

cagaggcagc agaagcagag ttggccttcc aggttcacgg agattccttt gctgtcaata 29616

ccctacacc cgatcagcgg tgtccggggc tgctcgtcag cggcattgtc ggtgtgcttt 29676

cgggattagc agtcataatc atctgcatgt tcatttttgc ttgctgctat agaaggcttt 29736

accgacaaaa atcagaccca ctgctgaacc tctatgttta atttttttcca gagccatgaa 29796

ggcagttagc gctctagttt tttgttcttt gattggcatt gtttttagtg ctgggttttt 29856

gaaaaatctt accatttatg aaggtgagaa tgccactcta gtgggcatca gtggtcaaaa 29916

tgtcagctgg ctaaaatacc atctagatgg gtggaaagac atttgcgatt ggaatgtcac 29976

tgtgtataca tgtaatggag ttaacctcac cattactaat gccacccaag atcagaatgg 30036

taggtttaag ggccagagtt tcactagaaa taatgggtat gaatcccata acatgtttat 30096

ctatgacgtc actgtcatca gaaatgagac tgccaccacc acacagatgc ccactacaca 30156

cagttctacc actactacca tgcaaaccac acagacaacc actacatcaa ctcagcatat 30216

gaccaccact acagcagcaa agccaagtag tgcagcgcct cagccccagg ctttggcttt 30276

gaaagctgca caacctagta caactactag gaccaatgag cagactactg aatttttgtc 30336

cactgtcgag agccacacca cagctacctc cagtgccttc tctagcaccg ccaatctctc 30396

ctcgctttcc tctacaccaa tcagtcccgc tactactccc accccagctc ttctccccac 30456

tcccctgaag caaactgagg acagcggcat gcaatggcag atcaccctgc tcattgtgat 30516

cgggttggtc atcctggccg tgttgctcta ctacatcttc tgccgccgca ttcccaacgc 30576

gcaccgcaaa ccggcctaca agcccatcgt tatcgggcag ccggagccgc ttcaggtgga 30636

aggggggtcta aggaatcttc tcttctcttt tacagtatgg tgattgaact atgattccta 30696

gacaattctt gatcactatt cttatctgcc tcctccaagt ctgtgccacc ctcgctctgg 30756

```
tggccaacgc cagtccagac tgtattgggc ccttcgcctc ctacgtgctc tttgccttca  30816

tcacctgcat ctgctgctgt agcatagtct gcctgcttat caccttcttc cagttcattg  30876

actggatctt tgtgcgcatc gcctacctgc gccaccaccc ccagtaccgc gaccagcgag  30936

tggcgcggct gctcaggctc ctctgataag catgcgggct ctgctacttc tcgcgcttct  30996

gctgttagtg ctcccccgcc ccgtcgaccc ccggtccccc actcagtccc ccgaagaggt  31056

ccgcaaatgc aaattccaag aaccctggaa attcctcaaa tgctaccgcc aaaaatcaga  31116

catgcttccc agctggatca tgatcattgg gatcgtgaac attctggcct gcaccctcat  31176

ctcctttgtg atttaccccc gctttgactt tggttggaac tcgccagagg cgctctatct  31236

cccgcctgaa cctgacacac caccacagca acctcaggca cacgcactac caccaccaca  31296

gcctaggcca caatacatgc ccatattaga ctatgaggcc gagccacagc gacccatgct  31356

ccccgctatt agttacttca atctaaccgg cggagatgac tgacccactg gccaacaaca  31416

acgtcaacga ccttctcctg gacatggacg gccgcgcctc ggagcagcga ctcgcccaac  31476

ttcgcattcg ccagcagcag gagagagccg tcaaggagct gcaggacggc atagccatcc  31536

accagtgcaa gaaaggcatc ttctgcctgg tgaaacaggc caagatctcc tacgaggtca  31596

ccccgaccga ccatcgcctc tcctacgagc tcctgcagca gcgccagaag ttcacctgcc  31656

tggtcggagt caaccccatc gtcatcaccc agcagtcggg cgataccaag gggtgcatcc  31716

actgctcctg cgactccccc gactgcgtcc acactctgat caagaccctc tgcggcctcc  31776

gcgacctcct ccccatgaac taatcacccc cttatccagt gaaataaata tcatattgat  31836

gatgatttaa ataaaaaata atcatttgat ttgaaataaa gatacaatca tattgatgat  31896

ttgagtttta aaaaataaag aatcacttac ttgaaatctg ataccaggtc tctgtccatg  31956

ttttctgcca acaccacctc actcccctct tcccagctct ggtactgcag accccggcgg  32016

gctgcaaact tcctccacac gctgaagggg atgtcaaatt cctcctgtcc ctcaatcttc  32076

attttatctt ctatcag atg tcc  aaa aag cgc gtc cgg  gtg gat gat gac   32126
                  Met Ser  Lys Lys Arg Val Arg  Val Asp Asp Asp
                      1465                1470

ttc  gac ccc gtc tac ccc  tac gat gca gac aac  gca ccg acc gtg     32171
Phe  Asp Pro Val Tyr Pro  Tyr Asp Ala Asp Asn  Ala Pro Thr Val
1475                 1480                1485

ccc  ttc atc aac ccc ccc  ttc gtc tct tca gat  gga ttc caa gag     32216
Pro  Phe Ile Asn Pro Pro  Phe Val Ser Ser Asp  Gly Phe Gln Glu
1490                 1495                1500
```

```
aag  ccc ctg ggg gtg ctg  tcc ctg cga ctg gct  gac ccc gtc acc     32261
Lys  Pro Leu Gly Val Leu  Ser Leu Arg Leu Ala  Asp Pro Val Thr
1505                 1510                 1515

acc  aag aac ggg gaa atc  acc ctc aag ctg gga  gag ggg gtg gac     32306
Thr  Lys Asn Gly Glu Ile  Thr Leu Lys Leu Gly  Glu Gly Val Asp
1520                 1525                 1530

ctc  gac tcc tcg gga aaa  ctc atc tcc aac acg  gcc acc aag gcc     32351
Leu  Asp Ser Ser Gly Lys  Leu Ile Ser Asn Thr  Ala Thr Lys Ala
1535                 1540                 1545

gcc  gcc cct ctc agt ttt  tcc aac aac acc att  tcc ctt aac atg     32396
Ala  Ala Pro Leu Ser Phe  Ser Asn Asn Thr Ile  Ser Leu Asn Met
1550                 1555                 1560

gat  acc cct ctt tat acc  aaa gat gga aaa tta  tcc tta caa gtt     32441
Asp  Thr Pro Leu Tyr Thr  Lys Asp Gly Lys Leu  Ser Leu Gln Val
1565                 1570                 1575

tct  cca ccg tta aac ata  tta aaa tca acc att  ctg aac aca tta     32486
Ser  Pro Pro Leu Asn Ile  Leu Lys Ser Thr Ile  Leu Asn Thr Leu
1580                 1585                 1590

gct  gta gct tat gga tca  ggt tta gga ctg agt  ggt ggc act gct     32531
Ala  Val Ala Tyr Gly Ser  Gly Leu Gly Leu Ser  Gly Gly Thr Ala
1595                 1600                 1605

ctt  gca gta cag ttg gcc  tct cca ctc act ttt  gat gaa aaa gga     32576
Leu  Ala Val Gln Leu Ala  Ser Pro Leu Thr Phe  Asp Glu Lys Gly
1610                 1615                 1620

aat  att aaa att aac cta  gcc agt ggt cca tta  aca gtt gat gca     32621
Asn  Ile Lys Ile Asn Leu  Ala Ser Gly Pro Leu  Thr Val Asp Ala
1625                 1630                 1635

agt  cga ctt agt atc aac  tgc aaa aga ggg gtc  act gtc act acc     32666
Ser  Arg Leu Ser Ile Asn  Cys Lys Arg Gly Val  Thr Val Thr Thr
1640                 1645                 1650

tca  gga gat gca att gaa  agc aac ata agc tgg  cct aaa ggt ata     32711
Ser  Gly Asp Ala Ile Glu  Ser Asn Ile Ser Trp  Pro Lys Gly Ile
1655                 1660                 1665

aga  ttt gaa ggt aat ggc  ata gct gca aac att  ggc aga gga ttg     32756
Arg  Phe Glu Gly Asn Gly  Ile Ala Ala Asn Ile  Gly Arg Gly Leu
1670                 1675                 1680

gaa  ttt gga acc act agt  aca gag act gat gtc  aca gat gca tac     32801
Glu  Phe Gly Thr Thr Ser  Thr Glu Thr Asp Val  Thr Asp Ala Tyr
1685                 1690                 1695

cca  att caa gtt aaa ttg  ggt act ggc ctt acc  ttt gac agt aca     32846
Pro  Ile Gln Val Lys Leu  Gly Thr Gly Leu Thr  Phe Asp Ser Thr
1700                 1705                 1710

ggc  gcc att gtt gct tgg  aac aaa gag gat gat  aaa ctt aca tta     32891
```

```
Gly  Ala Ile Val Ala Trp  Asn Lys Glu Asp Asp  Lys Leu Thr Leu
1715                 1720                 1725

tgg  acc aca gcc gac ccc  tcg cca aat tgc aaa  ata tac tct gaa     32936
Trp  Thr Thr Ala Asp Pro  Ser Pro Asn Cys Lys  Ile Tyr Ser Glu
1730                 1735                 1740

aaa  gat gcc aaa ctc aca  ctt tgc ttg aca aag  tgt gga agt caa     32981
Lys  Asp Ala Lys Leu Thr  Leu Cys Leu Thr Lys  Cys Gly Ser Gln
1745                 1750                 1755

att  ctg ggt act gtg act  gta ttg gca gtg aat  aat gga agt ctc     33026
Ile  Leu Gly Thr Val Thr  Val Leu Ala Val Asn  Asn Gly Ser Leu
1760                 1765                 1770

aac  cca atc aca aac aca  gta agc act gca ctc  gtc tcc ctc aag     33071
Asn  Pro Ile Thr Asn Thr  Val Ser Thr Ala Leu  Val Ser Leu Lys
1775                 1780                 1785

ttt  gat gca agt gga gtt  ttg cta agc agc tcc  aca tta gac aaa     33116
Phe  Asp Ala Ser Gly Val  Leu Leu Ser Ser Ser  Thr Leu Asp Lys
1790                 1795                 1800

gaa  tat tgg aac ttc aga  aag gga gat gtt aca  cct gct gag ccc     33161
Glu  Tyr Trp Asn Phe Arg  Lys Gly Asp Val Thr  Pro Ala Glu Pro
1805                 1810                 1815

tat  act aat gct ata ggt  ttt atg cct aac ata  aag gcc tat cct     33206
Tyr  Thr Asn Ala Ile Gly  Phe Met Pro Asn Ile  Lys Ala Tyr Pro
1820                 1825                 1830

aaa  aac aca tct gca gct  tca aaa agc cat att  gtc agt caa gtt     33251
Lys  Asn Thr Ser Ala Ala  Ser Lys Ser His Ile  Val Ser Gln Val
1835                 1840                 1845

tat  ctc aat ggg gat gag  gcc aaa cca ctg atg  ctg att att act     33296
Tyr  Leu Asn Gly Asp Glu  Ala Lys Pro Leu Met  Leu Ile Ile Thr
1850                 1855                 1860

ttt  aat gaa act gag gat  gca act tgc acc tac  agt atc act ttt     33341
Phe  Asn Glu Thr Glu Asp  Ala Thr Cys Thr Tyr  Ser Ile Thr Phe
1865                 1870                 1875

caa  tgg aaa tgg gat agt  act aag tac aca ggt  gaa aca ctt gct     33386
Gln  Trp Lys Trp Asp Ser  Thr Lys Tyr Thr Gly  Glu Thr Leu Ala
1880                 1885                 1890

acc  agc tcc ttc acc ttc  tcc tac atc gcc caa  gaa tga acactgtatc  33435
Thr  Ser Ser Phe Thr Phe  Ser Tyr Ile Ala Gln  Glu
1895                 1900                 1905

ccaccctgca tgccaaccct tcccacccca ctctgtctat ggaaaaaact ctgaagcaca  33495

aaataaaata aagttcaagt gttttattga ttcaacagtt ttacaggatt cgagcagtta  33555

ttttttcctcc accctcccag gacatggaat acaccaccct ctccccccgc acagccttga  33615
```

acatctgaat gccattggtg atggacatgc ttttggtctc cacgttccac acagtttcag 33675

agcgagccag tctcgggtcg gtcagggaga tgaaaccctc cgggcactcc cgcatctgca 33735

cctcacagct caacagctga ggattgtcct cggtggtcgg gatcacggtt atctggaaga 33795

agcagaagag cggcggtggg aatcatagtc cgcgaacggg atcggccggt ggtgtcgcat 33855

caggccccgc agcagtcgct gccgccgccg ctccgtcaag ctgctgctca gggggtccgg 33915

gtccagggac tccctcagca tgatgcccac ggccctcagc atcagtcgtc tggtgcggcg 33975

ggcgcagcag cgcatgcgga tctcgctcag gtcgctgcag tacgtgcaac acaggaccac 34035

caggttgttc aacagtccat agttcaacac gctccagccg aaactcatcg cgggaaggat 34095

gctacccacg tggccgtcgt accagatcct caggtaaatc aagtggcgct ccctccagaa 34155

cacgctgccc acgtacatga tctccttggg catgtggcgg ttcaccacct cccggtacca 34215

catcaccctc tggttgaaca tgcagccccg gatgatcctg cggaaccaca gggccagcac 34275

cgccccgccc gccatgcagc gaagagaccc cgggtcccgg caatggcaat ggaggaccca 34335

ccgctcgtac ccgtggatca tctgggagct gaacaagtct atgttggcac agcacaggca 34395

tatgctcatg catctcttca gcactctcag ctcctcgggg gtcaaaacca tatcccaggg 34455

cacggggaac tcttgcagga cagcgaaccc cgcagaacag ggcaatcctc gcacataact 34515

tacattgtgc atggacaggg tatcgcaatc aggcagcacc gggtgatcct ccaccagaga 34575

agcgcgggtc tcggtctcct cacagcgtgg taaggggccc ggccgatacg ggtgatggcg 34635

ggacgcggct gatcgtgttc gcgaccgtgt catgatgcag ttgctttcgg acattttcgt 34695

acttgctgta gcagaacctg gtccgggcgc tgcacaccga tcgccggcgg cggtcccggc 34755

gcttggaacg ctcggtgttg aaattgtaaa acagccactc tctcagaccg tgcagcagat 34815

ctagggcctc aggagtgatg aagatcccat catgcctgat agctctgatc acatcgacca 34875

ccgtggaatg ggccagaccc agccagatga tgcaattttg ttgggtttcg gtgacggcgg 34935

gggagggaag aacaggaaga accatgatta acttttaatc caaacggtct cggagcactt 34995

caaaatgaag gtcgcggaga tggcacctct cgccccccgct gtgttggtgg aaaataacag 35055

ccaggtcaaa ggtgatacgg ttctcgagat gttccacggt ggcttccagc aaagcctcca 35115

cgcgcacatc cagaaacaag acaatagcga aagcgggagg gttctctaat tcctcaatca 35175

tcatgttaca ctcctgcacc atccccagat aattttcatt tttccagcct tgaatgattc 35235

gaactagttc ctgaggtaaa tccaagccag ccatgataaa gagctcgcgc agagcgccct 35295

ccaccggcat tcttaagcac accctcataa ttccaagata ttctgctcct ggttcacctg 35355

cagcagattg acaagcggaa tatcaaaatc tctgccgcga tccctaagct cctccctcag 35415

caataactgt aagtactctt tcatatcctc tccgaaattt ttagccatag gaccaccagg 35475

aataagatta gggcaagcca cagtacagat aaaccgaagt cctccccagt gagcattgcc 35535

aaatgcaaga ctgctataag catgctggct agacccggtg atatcttcca gataactgga 35595

cagaaaatca cccaggcaat ttttaagaaa atcaacaaaa gaaaaatcct ccaggtgcac 35655

gtttagagcc tcgggaacaa cgatgaagta aatgcaagcg gtgcgttcca gcatggttag 35715

ttagctgatc tgtaaaaaac aaaaaataaa acattaaacc atgctagcct ggcgaacagg 35775

tgggtaaatc gttctctcca gcaccaggca ggccacgggg tctccggcgc gaccctcgta 35835

aaaattgtcg ctatgattga aaaccatcac agagagacgt tcccggtggc cggcgtgaat 35895

gattcgacaa gatgaataca ccccccggaac attggcgtcc gcgagtgaaa aaaagcgccc 35955

gaggaagcaa taaggcacta caatgctcag tctcaagtcc agcaaagcga tgccatgcgg 36015

atgaagcaca aaatcctcag gtgcgtacaa aatgtaatta ctccccctcct gcacaggcag 36075

cgaagccccc gatccctcca gatacacata caaagcctca gcgtccatag cttaccgagc 36135

agcagcacac aacaggcgca agagtcagag aaaggctgag ctctaacctg tccacccgct 36195

ctctgctcaa tatatagccc agatctacac tgacgtaaag gccaaagtct aaaaataccc 36255

gccaaataat cacacacgcc cagcacacgc ccagaaaccg gtgacacact caaaaaaata 36315

cgcgcacttc ctcaaacgcc caaactgccg tcatttccgg gttcccacgc tacgtcatcg 36375

gaattcgact ttcaaattcc gtcgaccgtt aaaaacgtca cccgccccgc ccctaacggt 36435

cgcccgtctc tcggccaatc accttcctcc ctccccaaat tcaaacagct catttgcata 36495

ttaacgcgca ccaaaagttt gaggtatatt attgatgatg 36535

<210> 10
<211> 531
<212> PRT
<213> chimpanzee adenovirus serotype Pan7

<400> 10

Met Met Arg Arg Val Tyr Pro Glu Gly Pro Pro Pro Ser Tyr Glu Ser
1 5 10 15

Val Met Gln Gln Ala Val Ala Ala Ala Met Gln Pro Pro Leu Glu Ala
20 25 30

```
Pro Tyr Val Pro Pro Arg Tyr Leu Ala Pro Thr Glu Gly Arg Asn Ser
        35                  40                  45

Ile Arg Tyr Ser Glu Leu Ala Pro Leu Tyr Asp Thr Thr Arg Leu Tyr
    50                  55                  60

Leu Val Asp Asn Lys Ser Ala Asp Ile Ala Ser Leu Asn Tyr Gln Asn
65                  70                  75                  80

Asp His Ser Asn Phe Leu Thr Thr Val Val Gln Asn Asn Asp Phe Thr
                85                  90                  95

Pro Thr Glu Ala Ser Thr Gln Thr Ile Asn Phe Asp Glu Arg Ser Arg
            100                 105                 110

Trp Gly Gly Gln Leu Lys Thr Ile Met His Thr Asn Met Pro Asn Val
        115                 120                 125

Asn Glu Phe Met Tyr Ser Asn Lys Phe Lys Ala Arg Val Met Val Ser
    130                 135                 140

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

Gln Asp Glu Leu Thr Tyr Glu Trp Val Glu Phe Glu Leu Pro Glu Gly
                165                 170                 175

Asn Phe Ser Val Thr Met Thr Ile Asp Leu Met Asn Asn Ala Ile Ile
            180                 185                 190

Asp Asn Tyr Leu Ala Val Gly Arg Gln Asn Gly Val Leu Glu Ser Asp
        195                 200                 205

Ile Gly Val Lys Phe Asp Thr Arg Asn Phe Arg Leu Gly Trp Asp Pro
    210                 215                 220

Val Thr Glu Leu Val Met Pro Gly Val Tyr Thr Asn Glu Ala Phe His
225                 230                 235                 240

Pro Asp Ile Val Leu Leu Pro Gly Cys Gly Val Asp Phe Thr Glu Ser
                245                 250                 255

Arg Leu Ser Asn Leu Leu Gly Ile Arg Lys Arg Gln Pro Phe Gln Glu
```

```
            260                 265                 270

Gly Phe Gln Ile Leu Tyr Glu Asp Leu Glu Gly Gly Asn Ile Pro Ala
        275                 280                 285

Leu Leu Asp Val Glu Ala Tyr Glu Lys Ser Lys Glu Glu Ala Ala Ala
    290                 295                 300

Ala Ala Thr Ala Ala Val Ala Thr Ala Ser Thr Glu Val Arg Gly Asp
305                 310                 315                 320

Asn Phe Ala Ser Ala Ala Ala Val Ala Glu Ala Ala Glu Thr Glu Ser
                325                 330                 335

Lys Ile Val Ile Gln Pro Val Glu Lys Asp Ser Lys Asp Arg Ser Tyr
            340                 345                 350

Asn Val Leu Ala Asp Lys Lys Asn Thr Ala Tyr Arg Ser Trp Tyr Leu
        355                 360                 365

Ala Tyr Asn Tyr Gly Asp Pro Glu Lys Gly Val Arg Ser Trp Thr Leu
    370                 375                 380

Leu Thr Thr Ser Asp Val Thr Cys Gly Val Glu Gln Val Tyr Trp Ser
385                 390                 395                 400

Leu Pro Asp Met Met Gln Asp Pro Val Thr Phe Arg Ser Thr Arg Gln
                405                 410                 415

Val Ser Asn Tyr Pro Val Val Gly Ala Glu Leu Leu Pro Val Tyr Ser
            420                 425                 430

Lys Ser Phe Phe Asn Glu Gln Ala Val Tyr Ser Gln Gln Leu Arg Ala
        435                 440                 445

Phe Thr Ser Leu Thr His Val Phe Asn Arg Phe Pro Glu Asn Gln Ile
    450                 455                 460

Leu Val Arg Pro Pro Ala Pro Thr Ile Thr Thr Val Ser Glu Asn Val
465                 470                 475                 480

Pro Ala Leu Thr Asp His Gly Thr Leu Pro Leu Arg Ser Ser Ile Arg
                485                 490                 495
```

Gly Val Gln Arg Val Thr Val Thr Asp Ala Arg Arg Arg Thr Cys Pro
500 505 510

Tyr Val Tyr Lys Ala Leu Gly Val Val Ala Pro Arg Val Leu Ser Ser
515 520 525

Arg Thr Phe
530

<210> 11
<211> 932
<212> PRT
<213> chimpanzee adenovirus serotype Pan7

<400> 11

Met Ala Thr Pro Ser Met Leu Pro Gln Trp Ala Tyr Met His Ile Ala
1 5 10 15

Gly Gln Asp Ala Ser Glu Tyr Leu Ser Pro Gly Leu Val Gln Phe Ala
20 25 30

Arg Ala Thr Asp Thr Tyr Phe Ser Leu Gly Asn Lys Phe Arg Asn Pro
35 40 45

Thr Val Ala Pro Thr His Asp Val Thr Thr Asp Arg Ser Gln Arg Leu
50 55 60

Thr Leu Arg Phe Val Pro Val Asp Arg Glu Asp Asn Thr Tyr Ser Tyr
65 70 75 80

Lys Val Arg Tyr Thr Leu Ala Val Gly Asp Asn Arg Val Leu Asp Met
85 90 95

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

Phe Lys Pro Tyr Ser Gly Thr Ala Tyr Asn Ser Leu Ala Pro Lys Gly
115 120 125

Ala Pro Asn Thr Cys Gln Trp Thr Tyr Lys Ala Gly Asp Thr Asp Thr
130 135 140

```
Glu Lys Thr Tyr Thr Tyr Gly Asn Ala Pro Val Gln Gly Ile Ser Ile
145                 150                 155                 160

Thr Lys Asp Gly Ile Gln Leu Gly Thr Asp Ser Asp Gly Gln Ala Ile
                165                 170                 175

Tyr Ala Asp Glu Thr Tyr Gln Pro Glu Pro Gln Val Gly Asp Ala Glu
            180                 185                 190

Trp His Asp Ile Thr Gly Thr Asp Glu Lys Tyr Gly Gly Arg Ala Leu
        195                 200                 205

Lys Pro Asp Thr Lys Met Lys Pro Cys Tyr Gly Ser Phe Ala Lys Pro
    210                 215                 220

Thr Asn Lys Glu Gly Gly Gln Ala Asn Val Lys Thr Glu Thr Gly Gly
225                 230                 235                 240

Thr Lys Glu Tyr Asp Ile Asp Met Ala Phe Phe Asp Asn Arg Ser Ala
                245                 250                 255

Ala Ala Ala Gly Leu Ala Pro Glu Ile Val Leu Tyr Thr Glu Asn Val
            260                 265                 270

Asp Leu Glu Thr Pro Asp Thr His Ile Val Tyr Lys Ala Gly Thr Asp
        275                 280                 285

Asp Ser Ser Ser Ser Ile Asn Leu Gly Gln Gln Ser Met Pro Asn Arg
    290                 295                 300

Pro Asn Tyr Ile Gly Phe Arg Asp Asn Phe Ile Gly Leu Met Tyr Tyr
305                 310                 315                 320

Asn Ser Thr Gly Asn Met Gly Val Leu Ala Gly Gln Ala Ser Gln Leu
                325                 330                 335

Asn Ala Val Val Asp Leu Gln Asp Arg Asn Thr Glu Leu Ser Tyr Gln
            340                 345                 350

Leu Leu Leu Asp Ser Leu Gly Asp Arg Thr Arg Tyr Phe Ser Met Trp
        355                 360                 365

Asn Gln Ala Val Asp Ser Tyr Asp Pro Asp Val Arg Ile Ile Glu Asn
```

370 375 380

His Gly Val Glu Asp Glu Leu Pro Asn Tyr Cys Phe Pro Leu Asp Ala
385 390 395 400

Val Gly Arg Thr Asp Thr Tyr Gln Gly Ile Lys Ala Asn Gly Asp Asn
405 410 415

Gln Thr Thr Trp Thr Lys Asp Asp Thr Val Asn Asp Ala Asn Glu Leu
420 425 430

Gly Lys Gly Asn Pro Phe Ala Met Glu Ile Asn Ile Gln Ala Asn Leu
435 440 445

Trp Arg Asn Phe Leu Tyr Ala Asn Val Ala Leu Tyr Leu Pro Asp Ser
450 455 460

Tyr Lys Tyr Thr Pro Ala Asn Ile Thr Leu Pro Thr Asn Thr Asn Thr
465 470 475 480

Tyr Asp Tyr Met Asn Gly Arg Val Val Ala Pro Ser Leu Val Asp Ala
485 490 495

Tyr Ile Asn Ile Gly Ala Arg Trp Ser Leu Asp Pro Met Asp Asn Val
500 505 510

Asn Pro Phe Asn His His Arg Asn Ala Gly Leu Arg Tyr Arg Ser Met
515 520 525

Leu Leu Gly Asn Gly Arg Tyr Val Pro Phe His Ile Gln Val Pro Gln
530 535 540

Lys Phe Phe Ala Ile Lys Ser Leu Leu Leu Leu Pro Gly Ser Tyr Thr
545 550 555 560

Tyr Glu Trp Asn Phe Arg Lys Asp Val Asn Met Ile Leu Gln Ser Ser
565 570 575

Leu Gly Asn Asp Leu Arg Thr Asp Gly Ala Ser Ile Ala Phe Thr Ser
580 585 590

Ile Asn Leu Tyr Ala Thr Phe Phe Pro Met Ala His Asn Thr Ala Ser
595 600 605

```
Thr Leu Glu Ala Met Leu Arg Asn Asp Thr Asn Asp Gln Ser Phe Asn
    610                 615                 620

Asp Tyr Leu Ser Ala Ala Asn Met Leu Tyr Pro Ile Pro Ala Asn Ala
625                 630                 635                 640

Thr Asn Val Pro Ile Ser Ile Pro Ser Arg Asn Trp Ala Ala Phe Arg
                645                 650                 655

Gly Trp Ser Phe Thr Arg Leu Lys Thr Arg Glu Thr Pro Ser Leu Gly
            660                 665                 670

Ser Gly Phe Asp Pro Tyr Phe Val Tyr Ser Gly Ser Ile Pro Tyr Leu
        675                 680                 685

Asp Gly Thr Phe Tyr Leu Asn His Thr Phe Lys Lys Val Ser Ile Thr
    690                 695                 700

Phe Asp Ser Ser Val Ser Trp Pro Gly Asn Asp Arg Leu Leu Thr Pro
705                 710                 715                 720

Asn Glu Phe Glu Ile Lys Arg Thr Val Asp Gly Glu Gly Tyr Asn Val
                725                 730                 735

Ala Gln Cys Asn Met Thr Lys Asp Trp Phe Leu Val Gln Met Leu Ala
            740                 745                 750

His Tyr Asn Ile Gly Tyr Gln Gly Phe Tyr Val Pro Glu Gly Tyr Lys
        755                 760                 765

Asp Arg Met Tyr Ser Phe Phe Arg Asn Phe Gln Pro Met Ser Arg Gln
    770                 775                 780

Val Val Asp Glu Val Asn Tyr Lys Asp Tyr Gln Ala Val Thr Leu Ala
785                 790                 795                 800

Tyr Gln His Asn Asn Ser Gly Phe Val Gly Tyr Leu Ala Pro Thr Met
                805                 810                 815

Arg Gln Gly Gln Pro Tyr Pro Ala Asn Tyr Pro Tyr Pro Leu Ile Gly
            820                 825                 830
```

```
Lys Ser Ala Val Ala Ser Val Thr Gln Lys Lys Phe Leu Cys Asp Arg
        835                 840                 845

Val Met Trp Arg Ile Pro Phe Ser Ser Asn Phe Met Ser Met Gly Ala
    850                 855                 860

Leu Thr Asp Leu Gly Gln Asn Met Leu Tyr Ala Asn Ser Ala His Ala
865                 870                 875                 880

Leu Asp Met Asn Phe Glu Val Asp Pro Met Asp Glu Ser Thr Leu Leu
                885                 890                 895

Tyr Val Val Phe Glu Val Phe Asp Val Val Arg Val His Gln Pro His
            900                 905                 910

Arg Gly Val Ile Glu Ala Val Tyr Leu Arg Thr Pro Phe Ser Ala Gly
        915                 920                 925

Asn Ala Thr Thr
    930
```

<210> 12
<211> 443
<212> PRT
<213> chimpanzee adenovirus serotype Pan7

<400> 12

```
Met Ser Lys Lys Arg Val Arg Val Asp Asp Asp Phe Asp Pro Val Tyr
1               5                   10                  15

Pro Tyr Asp Ala Asp Asn Ala Pro Thr Val Pro Phe Ile Asn Pro Pro
            20                  25                  30

Phe Val Ser Ser Asp Gly Phe Gln Glu Lys Pro Leu Gly Val Leu Ser
        35                  40                  45

Leu Arg Leu Ala Asp Pro Val Thr Thr Lys Asn Gly Glu Ile Thr Leu
    50                  55                  60

Lys Leu Gly Glu Gly Val Asp Leu Asp Ser Ser Gly Lys Leu Ile Ser
65                  70                  75                  80

Asn Thr Ala Thr Lys Ala Ala Ala Pro Leu Ser Phe Ser Asn Asn Thr
```

85 90 95

Ile Ser Leu Asn Met Asp Thr Pro Leu Tyr Thr Lys Asp Gly Lys Leu
100 105 110

Ser Leu Gln Val Ser Pro Pro Leu Asn Ile Leu Lys Ser Thr Ile Leu
115 120 125

Asn Thr Leu Ala Val Ala Tyr Gly Ser Gly Leu Gly Leu Ser Gly Gly
130 135 140

Thr Ala Leu Ala Val Gln Leu Ala Ser Pro Leu Thr Phe Asp Glu Lys
145 150 155 160

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

Ser Arg Leu Ser Ile Asn Cys Lys Arg Gly Val Thr Val Thr Thr Ser
180 185 190

Gly Asp Ala Ile Glu Ser Asn Ile Ser Trp Pro Lys Gly Ile Arg Phe
195 200 205

Glu Gly Asn Gly Ile Ala Ala Asn Ile Gly Arg Gly Leu Glu Phe Gly
210 215 220

Thr Thr Ser Thr Glu Thr Asp Val Thr Asp Ala Tyr Pro Ile Gln Val
225 230 235 240

Lys Leu Gly Thr Gly Leu Thr Phe Asp Ser Thr Gly Ala Ile Val Ala
245 250 255

Trp Asn Lys Glu Asp Asp Lys Leu Thr Leu Trp Thr Thr Ala Asp Pro
260 265 270

Ser Pro Asn Cys Lys Ile Tyr Ser Glu Lys Asp Ala Lys Leu Thr Leu
275 280 285

Cys Leu Thr Lys Cys Gly Ser Gln Ile Leu Gly Thr Val Thr Val Leu
290 295 300

Ala Val Asn Asn Gly Ser Leu Asn Pro Ile Thr Asn Thr Val Ser Thr
305 310 315 320

Ala Leu Val Ser Leu Lys Phe Asp Ala Ser Gly Val Leu Leu Ser Ser
325 330 335

Ser Thr Leu Asp Lys Glu Tyr Trp Asn Phe Arg Lys Gly Asp Val Thr
340 345 350

Pro Ala Glu Pro Tyr Thr Asn Ala Ile Gly Phe Met Pro Asn Ile Lys
355 360 365

Ala Tyr Pro Lys Asn Thr Ser Ala Ala Ser Lys Ser His Ile Val Ser
370 375 380

Gln Val Tyr Leu Asn Gly Asp Glu Ala Lys Pro Leu Met Leu Ile Ile
385 390 395 400

Thr Phe Asn Glu Thr Glu Asp Ala Thr Cys Thr Tyr Ser Ile Thr Phe
405 410 415

Gln Trp Lys Trp Asp Ser Thr Lys Tyr Thr Gly Glu Thr Leu Ala Thr
420 425 430

Ser Ser Phe Thr Phe Ser Tyr Ile Ala Gln Glu
435 440

<210> 13
<211> 338
<212> PRT
<213> simian serotype C1

<400> 13

Ala Pro Lys Gly Ala Pro Asn Thr Ser Gln Trp Leu Asp Lys Gly Val
1 5 10 15

Thr Thr Thr Asp Asn Asn Thr Glu Asn Gly Asp Glu Glu Asp Glu Val
20 25 30

Ala Glu Glu Gly Glu Glu Glu Lys Gln Ala Thr Tyr Thr Phe Gly Asn
35 40 45

Ala Pro Val Lys Ala Glu Ala Glu Ile Thr Lys Glu Gly Leu Pro Ile
50 55 60

```
Gly Leu Glu Val Pro Ser Glu Gly Asp Pro Lys Pro Ile Tyr Ala Asp
65                  70                  75                  80

Lys Leu Tyr Gln Pro Glu Pro Gln Val Gly Glu Glu Ser Trp Thr Asp
                85                  90                  95

Thr Asp Gly Thr Asp Glu Lys Tyr Gly Gly Arg Ala Leu Lys Pro Glu
            100                 105                 110

Thr Lys Met Lys Pro Cys Tyr Gly Ser Phe Ala Lys Pro Thr Asn Val
        115                 120                 125

Lys Gly Gly Gln Ala Lys Val Lys Lys Val Glu Glu Gly Lys Val Glu
    130                 135                 140

Tyr Asp Ile Asp Met Asn Phe Phe Asp Leu Arg Ser Gln Lys Thr Gly
145                 150                 155                 160

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

Pro Asp Thr His Val Val Tyr Lys Pro Gly Ala Ser Asp Ala Ser Ser
            180                 185                 190

His Ala Asn Leu Gly Gln Gln Ser Met Pro Asn Arg Pro Asn Tyr Ile
        195                 200                 205

Gly Phe Arg Asp Asn Phe Ile Gly Leu Met Tyr Tyr Asn Ser Thr Gly
    210                 215                 220

Asn Met Gly Val Leu Ala Gly Gln Ala Ser Gln Leu Asn Ala Val Val
225                 230                 235                 240

Asp Leu Gln Asp Arg Asn Thr Glu Leu Ser Tyr Gln Leu Leu Leu Asp
                245                 250                 255

Ser Leu Gly Asp Arg Thr Arg Tyr Phe Ser Met Trp Asn Gln Ala Val
            260                 265                 270

Asp Ser Tyr Asp Pro Asp Val Arg Val Ile Glu Asn His Gly Val Glu
        275                 280                 285

Asp Glu Leu Pro Asn Tyr Cys Phe Pro Leu Asp Gly Val Gly Pro Arg
```

290 295 300

Thr Asp Ser Tyr Lys Gly Ile Glu Thr Asn Gly Asp Glu Asn Thr Thr
305 310 315 320

Trp Lys Asp Leu Asp Pro Asn Gly Ile Ser Glu Leu Ala Lys Gly Asn
325 330 335

Pro Phe

<210> 14
<211> 315
<212> PRT
<213> chimpanzee adenovirus Pan-9

<400> 14

Ala Pro Lys Gly Ala Pro Asn Thr Cys Gln Trp Thr Tyr Lys Ala Asp
1 5 10 15

Gly Glu Thr Ala Thr Glu Lys Thr Tyr Thr Tyr Gly Asn Ala Pro Val
20 25 30

Gln Gly Ile Asn Ile Thr Lys Asp Gly Ile Gln Leu Gly Thr Asp Thr
35 40 45

Asp Asp Gln Pro Ile Tyr Ala Asp Lys Thr Tyr Gln Pro Glu Pro Gln
50 55 60

Val Gly Asp Ala Glu Trp His Asp Ile Thr Gly Thr Asp Glu Lys Tyr
65 70 75 80

Gly Gly Arg Ala Leu Lys Pro Asp Thr Lys Met Lys Pro Cys Tyr Gly
85 90 95

Ser Phe Ala Lys Pro Thr Asn Lys Glu Gly Gly Gln Ala Asn Val Lys
100 105 110

Thr Gly Thr Gly Thr Thr Lys Glu Tyr Asp Ile Asp Met Ala Phe Phe
115 120 125

Asp Asn Arg Ser Ala Ala Ala Ala Gly Leu Ala Pro Glu Ile Val Leu
130 135 140

```
Tyr Thr Glu Asn Val Asp Leu Glu Thr Pro Asp Thr His Ile Val Tyr
145                 150                 155                 160

Lys Ala Gly Thr Asp Asp Ser Ser Ser Ser Ile Asn Leu Gly Gln Gln
                165                 170                 175

Ala Met Pro Asn Arg Pro Asn Tyr Ile Gly Phe Arg Asp Asn Phe Ile
            180                 185                 190

Gly Leu Met Tyr Tyr Asn Ser Thr Gly Asn Met Gly Val Leu Ala Gly
        195                 200                 205

Gln Ala Ser Gln Leu Asn Ala Val Val Asp Leu Gln Asp Arg Asn Thr
    210                 215                 220

Glu Leu Ser Tyr Gln Leu Leu Leu Asp Ser Leu Gly Asp Arg Thr Arg
225                 230                 235                 240

Tyr Phe Ser Met Trp Asn Gln Ala Val Asp Ser Tyr Asp Pro Asp Val
                245                 250                 255

Arg Ile Ile Glu Asn His Gly Val Glu Asp Glu Leu Pro Asn Tyr Cys
            260                 265                 270

Phe Pro Leu Asp Ala Val Gly Arg Thr Asp Thr Tyr Gln Gly Ile Lys
        275                 280                 285

Ala Asn Gly Thr Asp Gln Thr Thr Trp Thr Lys Asp Asp Ser Val Asn
    290                 295                 300

Asp Ala Asn Glu Ile Gly Lys Gly Asn Pro Phe
305                 310                 315
```

<210> 15
<211> 315
<212> PRT
<213> chimpanzee adenovirus Pan-5

<400> 15

```
Ala Pro Lys Gly Ala Pro Asn Thr Cys Gln Trp Thr Tyr Lys Ala Asp
1               5                   10                  15

Gly Asp Thr Gly Thr Glu Lys Thr Tyr Thr Tyr Gly Asn Ala Pro Val
```

```
            20                  25                  30

Gln Gly Ile Ser Ile Thr Lys Asp Gly Ile Gln Leu Gly Thr Asp Thr
        35                  40                  45

Asp Asp Gln Pro Ile Tyr Ala Asp Lys Thr Tyr Gln Pro Glu Pro Gln
    50                  55                  60

Val Gly Asp Ala Glu Trp His Asp Ile Thr Gly Thr Asp Glu Lys Tyr
65                  70                  75                  80

Gly Gly Arg Ala Leu Lys Pro Asp Thr Lys Met Lys Pro Cys Tyr Gly
                85                  90                  95

Ser Phe Ala Lys Pro Thr Asn Lys Glu Gly Gly Gln Ala Asn Val Lys
            100                 105                 110

Thr Glu Thr Gly Gly Thr Lys Glu Tyr Asp Ile Asp Met Ala Phe Phe
        115                 120                 125

Asp Asn Arg Ser Ala Ala Ala Ala Gly Leu Ala Pro Glu Ile Val Leu
    130                 135                 140

Tyr Thr Glu Asn Val Asp Leu Glu Thr Pro Asp Thr His Ile Val Tyr
145                 150                 155                 160

Lys Ala Gly Thr Asp Asp Ser Ser Ser Ser Ile Asn Leu Gly Gln Gln
                165                 170                 175

Ser Met Pro Asn Arg Pro Asn Tyr Ile Gly Phe Arg Asp Asn Phe Ile
            180                 185                 190

Gly Leu Met Tyr Tyr Asn Ser Thr Gly Asn Met Gly Val Leu Ala Gly
        195                 200                 205

Gln Ala Ser Gln Leu Asn Ala Val Val Asp Leu Gln Asp Arg Asn Thr
    210                 215                 220

Glu Leu Ser Tyr Gln Leu Leu Leu Asp Ser Leu Gly Asp Arg Thr Arg
225                 230                 235                 240

Tyr Phe Ser Met Trp Asn Gln Ala Val Asp Ser Tyr Asp Pro Asp Val
                245                 250                 255
```

Arg Ile Ile Glu Asn His Gly Val Glu Asp Glu Leu Pro Asn Tyr Cys
260 265 270

Phe Pro Leu Asp Ala Val Gly Arg Thr Asp Thr Tyr Gln Gly Ile Lys
275 280 285

Ala Asn Gly Ala Asp Gln Thr Thr Trp Thr Lys Asp Asp Thr Val Asn
290 295 300

Asp Ala Asn Glu Leu Gly Lys Gly Asn Pro Phe
305 310 315

<210> 16
<211> 324
<212> PRT
<213> chimpanzee adenovirus Pan-6

<400> 16

Ala Pro Lys Gly Ala Pro Asn Ser Ser Gln Trp Glu Gln Ala Lys Thr
1 5 10 15

Gly Asn Gly Gly Thr Met Glu Thr His Thr Tyr Gly Val Ala Pro Met
20 25 30

Gly Gly Glu Asn Ile Thr Lys Asp Gly Leu Gln Ile Gly Thr Asp Val
35 40 45

Thr Ala Asn Gln Asn Lys Pro Ile Tyr Ala Asp Lys Thr Phe Gln Pro
50 55 60

Glu Pro Gln Val Gly Glu Glu Asn Trp Gln Glu Thr Glu Asn Phe Tyr
65 70 75 80

Gly Gly Arg Ala Leu Lys Lys Asp Thr Lys Met Lys Pro Cys Tyr Gly
85 90 95

Ser Tyr Ala Arg Pro Thr Asn Glu Lys Gly Gly Gln Ala Lys Leu Lys
100 105 110

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

```
Phe Phe Asp Thr Pro Gly Gly Thr Val Asn Gly Gln Asp Glu Tyr Lys
    130                 135                 140

Ala Asp Ile Val Met Tyr Thr Glu Asn Thr Tyr Leu Glu Thr Pro Asp
145                 150                 155                 160

Thr His Val Val Tyr Lys Pro Gly Lys Asp Asp Ala Ser Ser Glu Ile
                165                 170                 175

Asn Leu Val Gln Gln Ser Met Pro Asn Arg Pro Asn Tyr Ile Gly Phe
            180                 185                 190

Arg Asp Asn Phe Ile Gly Leu Met Tyr Tyr Asn Ser Thr Gly Asn Met
        195                 200                 205

Gly Val Leu Ala Gly Gln Ala Ser Gln Leu Asn Ala Val Val Asp Leu
    210                 215                 220

Gln Asp Arg Asn Thr Glu Leu Ser Tyr Gln Leu Leu Leu Asp Ser Leu
225                 230                 235                 240

Gly Asp Arg Thr Arg Tyr Phe Ser Met Trp Asn Gln Ala Val Asp Ser
                245                 250                 255

Tyr Asp Pro Asp Val Arg Ile Ile Glu Asn His Gly Val Glu Asp Glu
            260                 265                 270

Leu Pro Asn Tyr Cys Phe Pro Leu Asp Gly Ser Gly Thr Asn Ala Ala
        275                 280                 285

Tyr Gln Gly Val Lys Val Lys Asp Gly Gln Asp Gly Asp Val Glu Ser
    290                 295                 300

Glu Trp Glu Asn Asp Asp Thr Val Ala Ala Arg Asn Gln Leu Cys Lys
305                 310                 315                 320

Gly Asn Ile Phe
```

<210> 17
<211> 314
<212> PRT
<213> chimpanzee adenovirus Pan-7

<400> 17

```
Ala Pro Lys Gly Ala Pro Asn Thr Cys Gln Trp Thr Tyr Lys Ala Gly
1               5                  10                  15

Asp Thr Asp Thr Glu Lys Thr Tyr Thr Tyr Gly Asn Ala Pro Val Gln
            20                  25                  30

Gly Ile Ser Ile Thr Lys Asp Gly Ile Gln Leu Gly Thr Asp Ser Asp
        35                  40                  45

Gly Gln Ala Ile Tyr Ala Asp Glu Thr Tyr Gln Pro Glu Pro Gln Val
    50                  55                  60

Gly Asp Ala Glu Trp His Asp Ile Thr Gly Thr Asp Glu Lys Tyr Gly
65                  70                  75                  80

Gly Arg Ala Leu Lys Pro Asp Thr Lys Met Lys Pro Cys Tyr Gly Ser
                85                  90                  95

Phe Ala Lys Pro Thr Asn Lys Glu Gly Gly Gln Ala Asn Val Lys Thr
            100                 105                 110

Glu Thr Gly Gly Thr Lys Glu Tyr Asp Ile Asp Met Ala Phe Phe Asp
        115                 120                 125

Asn Arg Ser Ala Ala Ala Ala Gly Leu Ala Pro Glu Ile Val Leu Tyr
    130                 135                 140

Thr Glu Asn Val Asp Leu Glu Thr Pro Asp Thr His Ile Val Tyr Lys
145                 150                 155                 160

Ala Gly Thr Asp Asp Ser Ser Ser Ser Ile Asn Leu Gly Gln Gln Ser
                165                 170                 175

Met Pro Asn Arg Pro Asn Tyr Ile Gly Phe Arg Asp Asn Phe Ile Gly
            180                 185                 190

Leu Met Tyr Tyr Asn Ser Thr Gly Asn Met Gly Val Leu Ala Gly Gln
        195                 200                 205

Ala Ser Gln Leu Asn Ala Val Val Asp Leu Gln Asp Arg Asn Thr Glu
    210                 215                 220
```

Leu Ser Tyr Gln Leu Leu Leu Asp Ser Leu Gly Asp Arg Thr Arg Tyr
225 230 235 240

Phe Ser Met Trp Asn Gln Ala Val Asp Ser Tyr Asp Pro Asp Val Arg
245 250 255

Ile Ile Glu Asn His Gly Val Glu Asp Glu Leu Pro Asn Tyr Cys Phe
260 265 270

Pro Leu Asp Ala Val Gly Arg Thr Asp Thr Tyr Gln Gly Ile Lys Ala
275 280 285

Asn Gly Asp Asn Gln Thr Thr Trp Thr Lys Asp Asp Thr Val Asn Asp
290 295 300

Ala Asn Glu Leu Gly Lys Gly Asn Pro Phe
305 310

<210> 18
<211> 179
<212> PRT
<213> chimpanzee adenovirus Pan9

<400> 18

Thr Leu Trp Thr Thr Pro Asp Pro Ser Pro Asn Cys Gln Ile Leu Ala
1 5 10 15

Glu Asn Asp Ala Lys Leu Thr Leu Cys Leu Thr Lys Cys Gly Ser Gln
20 25 30

Ile Leu Ala Thr Val Ser Val Leu Val Val Gly Ser Gly Asn Leu Asn
35 40 45

Pro Ile Thr Gly Thr Val Ser Ser Ala Gln Val Phe Leu Arg Phe Asp
50 55 60

Ala Asn Gly Val Leu Leu Thr Glu His Ser Thr Leu Lys Lys Tyr Trp
65 70 75 80

Gly Tyr Arg Gln Gly Asp Ser Ile Asp Gly Thr Pro Tyr Thr Asn Ala
85 90 95

Val Gly Phe Met Pro Asn Leu Lys Ala Tyr Pro Lys Ser Gln Ser Ser

100 105 110

Thr Thr Lys Asn Asn Ile Val Gly Gln Val Tyr Met Asn Gly Asp Val
115 120 125

Ser Lys Pro Met Leu Leu Thr Ile Thr Leu Asn Gly Thr Asp Asp Ser
130 135 140

Asn Ser Thr Tyr Ser Met Ser Phe Ser Tyr Thr Trp Thr Asn Gly Ser
145 150 155 160

Tyr Val Gly Ala Thr Phe Gly Ala Asn Ser Tyr Thr Phe Ser Tyr Ile
165 170 175

Ala Gln Glu

<210> 19
<211> 185
<212> PRT
<213> chimpanzee adenovirus Pan6

<400> 19

Thr Leu Trp Thr Thr Pro Asp Pro Ser Pro Asn Cys Gln Leu Leu Ser
1 5 10 15

Asp Arg Asp Ala Lys Phe Thr Leu Cys Leu Thr Lys Cys Gly Ser Gln
20 25 30

Ile Leu Gly Thr Val Ala Val Ala Ala Val Thr Val Gly Ser Ala Leu
35 40 45

Asn Pro Ile Asn Asp Thr Val Lys Ser Ala Ile Val Phe Leu Arg Phe
50 55 60

Asp Ser Asp Gly Val Leu Met Ser Asn Ser Ser Met Val Gly Asp Tyr
65 70 75 80

Trp Asn Phe Arg Glu Gly Gln Thr Thr Gln Ser Val Ala Tyr Thr Asn
85 90 95

Ala Val Gly Phe Met Pro Asn Ile Gly Ala Tyr Pro Lys Thr Gln Ser
100 105 110

Lys Thr Pro Lys Asn Ser Ile Val Ser Gln Val Tyr Leu Thr Gly Glu
115 120 125

Thr Thr Met Pro Met Thr Leu Thr Ile Thr Phe Asn Gly Thr Asp Glu
130 135 140

Lys Asp Thr Thr Pro Val Ser Thr Tyr Ser Met Thr Phe Thr Trp Gln
145 150 155 160

Trp Thr Gly Asp Tyr Lys Asp Lys Asn Ile Thr Phe Ala Thr Asn Ser
165 170 175

Phe Ser Phe Ser Tyr Ile Ala Gln Glu
180 185

<210> 20
<211> 179
<212> PRT
<213> chimpanzee adenovirus Pan7

<400> 20

Thr Leu Trp Thr Thr Ala Asp Pro Ser Pro Asn Cys Lys Ile Tyr Ser
1 5 10 15

Glu Lys Asp Ala Lys Leu Thr Leu Cys Leu Thr Lys Cys Gly Ser Gln
20 25 30

Ile Leu Gly Thr Val Thr Val Leu Ala Val Asn Asn Gly Ser Leu Asn
35 40 45

Pro Ile Thr Asn Thr Val Ser Thr Ala Leu Val Ser Leu Lys Phe Asp
50 55 60

Ala Ser Gly Val Leu Leu Ser Ser Ser Thr Leu Asp Lys Glu Tyr Trp
65 70 75 80

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

Ile Gly Phe Met Pro Asn Ile Lys Ala Tyr Pro Lys Asn Thr Ser Ala
100 105 110

Ala Ser Lys Ser His Ile Val Ser Gln Val Tyr Leu Asn Gly Asp Glu

115 120 125

Ala Lys Pro Leu Met Leu Ile Ile Thr Phe Asn Glu Thr Glu Asp Ala
130 135 140

Thr Cys Thr Tyr Ser Ile Thr Phe Gln Trp Lys Trp Asp Ser Thr Lys
145 150 155 160

Tyr Thr Gly Glu Thr Leu Ala Thr Ser Ser Phe Thr Phe Ser Tyr Ile
165 170 175

Ala Gln Glu

<210> 21
<211> 179
<212> PRT
<213> chimpanzee adenovirus Pan5

<400> 21

Thr Leu Trp Thr Thr Ala Asp Pro Ser Pro Asn Cys His Ile Tyr Ser
1 5 10 15

Glu Lys Asp Ala Lys Leu Thr Leu Cys Leu Thr Lys Cys Gly Ser Gln
20 25 30

Ile Leu Gly Thr Val Ser Leu Ile Ala Val Asp Thr Gly Ser Leu Asn
35 40 45

Pro Ile Thr Gly Thr Val Thr Thr Ala Leu Val Ser Leu Lys Phe Asp
50 55 60

Ala Asn Gly Val Leu Gln Ser Ser Ser Thr Leu Asp Ser Asp Tyr Trp
65 70 75 80

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

Ile Gly Phe Met Pro Asn Leu Lys Ala Tyr Pro Lys Asn Thr Ser Gly
100 105 110

Ala Ala Lys Ser His Ile Val Gly Lys Val Tyr Leu His Gly Asp Thr
115 120 125

Gly Lys Pro Leu Asp Leu Ile Ile Thr Phe Asn Glu Thr Ser Asp Glu
130 135 140

Ser Cys Thr Tyr Cys Ile Asn Phe Gln Trp Gln Trp Gly Ala Asp Gln
145 150 155 160

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

Ala Lys Glu

<210> 22
<211> 183
<212> PRT
<213> human adenovirus Ad 2

<400> 22

Thr Leu Trp Thr Thr Pro Asp Pro Ser Pro Asn Cys Arg Ile His Ser
1 5 10 15

Asp Asn Asp Cys Lys Phe Thr Leu Val Leu Thr Lys Cys Gly Ser Gln
20 25 30

Val Leu Ala Thr Val Ala Ala Leu Ala Val Ser Gly Asp Leu Ser Ser
35 40 45

Met Thr Gly Thr Val Ala Ser Val Ser Ile Phe Leu Arg Phe Asp Gln
50 55 60

Asn Gly Val Leu Met Glu Asn Ser Ser Leu Lys Lys His Tyr Trp Asn
65 70 75 80

Phe Arg Asn Gly Asn Ser Thr Asn Ala Asn Pro Tyr Thr Asn Ala Val
85 90 95

Gly Phe Met Pro Asn Leu Leu Ala Tyr Pro Lys Thr Gln Ser Gln Thr
100 105 110

Ala Lys Asn Asn Ile Val Ser Gln Val Tyr Leu His Gly Asp Lys Thr
115 120 125

Lys Pro Met Ile Leu Thr Ile Thr Leu Asn Gly Thr Ser Glu Ser Thr

130 135 140

Glu Thr Ser Glu Val Ser Thr Tyr Ser Met Ser Phe Thr Trp Ser Trp
145 150 155 160

Glu Ser Gly Lys Tyr Thr Thr Glu Thr Phe Ala Thr Asn Ser Tyr Thr
165 170 175

Phe Ser Tyr Ile Ala Gln Glu
180

<210> 23
<211> 182
<212> PRT
<213> human adenovirus Ad 5

<400> 23

Thr Leu Trp Thr Thr Pro Ala Pro Ser Pro Asn Cys Arg Leu Asn Ala
1 5 10 15

Glu Lys Asp Ala Lys Leu Thr Leu Val Leu Thr Lys Cys Gly Ser Gln
20 25 30

Ile Leu Ala Thr Val Ser Val Leu Ala Val Lys Gly Ser Leu Ala Pro
35 40 45

Ile Ser Gly Thr Val Gln Ser Ala His Leu Ile Ile Arg Phe Asp Glu
50 55 60

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

Phe Arg Asn Gly Asp Leu Thr Glu Gly Thr Ala Tyr Thr Asn Ala Val
85 90 95

Gly Phe Met Pro Asn Leu Ser Ala Tyr Pro Lys Ser His Gly Lys Thr
100 105 110

Ala Lys Ser Asn Ile Val Ser Gln Val Tyr Leu Asn Gly Asp Lys Thr
115 120 125

Lys Pro Val Thr Leu Thr Ile Thr Leu Asn Gly Thr Gln Glu Thr Gly
130 135 140

Asp Thr Thr Pro Ser Ala Tyr Ser Met Ser Phe Ser Trp Asp Trp Ser
145 150 155 160

Gly His Asn Tyr Ile Asn Glu Ile Phe Ala Thr Ser Ser Tyr Thr Glu
165 170 175

Ser Tyr Ile Ala Gln Glu
180

<210> 24
<211> 34264
<212> DNA
<213> simian adenovirus SV-1

<220>
<221> CDS
<222> (12454)..(13965)
<223> L2 Penton

<220>
<221> CDS
<222> (16841)..(19636)
<223> L3 Hexon

<220>
<221> CDS
<222> (28059)..(29150)
<223> L5 Fiber #2

<220>
<221> CDS
<222> (29183)..(30865)
<223> L5 Fiber #1

<400> 24
tccttattct ggaaacgtgc caatatgata atgagcgggg aggagcgagg cggggccggg 60

gtgacgtgcg gtgacgtggg gtgacgcggg gtggcgcgag ggcggggcgg gagtggggag 120

gcgcttagtt tttacgtatg cggaaggagg ttttataccg gaagttgggt aatttgggcg 180

tatacttgta agttttgtgt aatttggcgc gaaaaccggg taatgaggaa gttgaggtta 240

atatgtactt tttatgactg ggcggaattt ctgctgatca gcagtgaact ttgggcgctg 300

acggggaggt ttcgctacgt ggcagtacca cgagaaggct caaaggtccc atttattgta 360

ctcctcagcg ttttcgctgg gtatttaaac gctgtcagat catcaagagg ccactcttga 420

gtgccggcga gtagagtttt ctcctccgcg ctgccgcgat gaggctggtt cccgagatgt 480

acggtgtttt ctgcagcgag acggcccgga actcagatga gctgcttaat acagatctgc 540

tggatgttcc caactcgcct gtggcttcgc ctccgtcgct tcatgatctt ttcgatgtgg 600

aagtggatcc accgcaagat cccaacgagg acgcggtaaa cagtatgttc cctgaatgtc 660

tgtttgaggc ggctgaggag ggttctcaca gcagtgaaga gagcagacgg ggagaggaac 720

tggacttgaa atgctacgag gaatgtctgc cttctagcga ttctgaaacg gaacagacag 780

ggggagacgg ctgtgagtcg gcaatgaaaa atgaacttgt attagactgt ccagaacatc 840

ctggtcatgg ctgccgtgcc tgtgcttttc atagaaatgc cagcggaaat cctgagactc 900

tatgtgctct gtgttatctg cgccttacca gcgattttgt atacagtaag taaagtgttt 960

tcattggcgt acggtagggg attcgttgaa gtgctttgtg acttattatg tgtcattatt 1020

tctaggtgac gtgtccgacg tggaagggga aggagataga tcaggggctg ctaattctcc 1080

ttgcactttg ggggctgtgg ttccagttgg cattttaaa ccgagtggtg gaggagaacg 1140

agccggagga gaccgagaat ctgagagccg gcctggaccc tccagtggaa gactaggtgc 1200

tgaggatgat cctgaagagg ggactagtgg gggtgctagg aaaaagcaaa aaactgagcc 1260

tgaacctaga aacttttga atgagttgac tgtaagccta atgaatcggc agcgtcctga 1320

gacggtgttt tggactgagt tggaggatga gttcaagaag ggggaattaa acctcttgta 1380

caagtatggg tttgagcagt tgaaaactca ctggttggag ccgtgggagg atatggaaat 1440

ggctctagac acctttgcta aagtggctct gcggccggat aaagtttaca ctattcgccg 1500

cactgttaat ataaaaaaga gtgtttatgt tatcggccat ggagctctgg tgcaggtgca 1560

gaccccagac cgggtggctt tcaattgcgg catgcagagt ttgggccccg gggtgatagg 1620

tttgaatgga gttacatttc aaaatgtcag gtttactggt gatgatttta atggctctgt 1680

gtttgtgact agcacccagc taaccctcca cggtgtttac ttttttaact ttaacaatac 1740

atgtgtggag tcatggggta gggtgtctct gaggggctgc agttttcatg gttgctggaa 1800

ggcggtggtg ggaagaatta aaagtgtcat gtctgtgaag aaatgcatat ttgaacgctg 1860

tgtgatagct ctagcagtag aggggtacgg acggatcagg aataacgccg catctgagaa 1920

tggatgtttt cttttgctga aaggtacggc cagcgttaag cataatatga tttgcggcag 1980

cggcctgtgc ccctcgcagc tcttaacttg cgcagatgga aactgtcaca ccttgcgcac 2040

cgtgcacata gtgtcccact cgcgccgcac ctggccaaca tttgagcaca atatgctcat 2100

gcgttgcgcc gttcacctag gtgctagacg cggcgtgttt atgccttatc aatgtaactt 2160

tagtcatact aagattttgc tggaaactga ttccttccct cgagtatgtt tcaatggggt 2220

gtttgacatg tcaatggaac tttttaaagt gataagatat gatgaaacca agtctcgttg 2280

tcgctcatgt gaatgcggag ctaatcattt gaggttgtat cctgtaaccc tgaacgttac 2340

cgaggagctg aggacggacc accacatgct gtcttgcctg cgtaccgact atgaatccag 2400

cgatgaggag tgaggtgagg ggcggagcca caaagggtat aaaggggcat gaggggtggg 2460

cgcggtgttt caaaatgagc gggacgacgg acggcaatgc gtttgagggg ggagtgttca 2520

gcccatatct gacatctcgt cttccttcct gggcaggagt tcgtcagaat gtagtgggct 2580

ccaccgtgga cggacggccg gtcgcccctg caaattccgc caccctcacc tatgccaccg 2640

tgggatcatc gttggacact gccgcggcag ctgccgcttc tgctgccgct tctactgctc 2700

gcggcatggc ggctgatttt ggactatata accaactggc cactgcagct gtggcgtctc 2760

ggtctctggt tcaagaagat gccctgaatg tgatcttgac tcgcctggag atcatgtcac 2820

gtcgcctgga cgaactggct gcgcagatat cccaagctaa ccccgatacc gcttcagaat 2880

cttaaaataa agacaaacaa atttgttgaa aagtaaaatg gctttatttg ttttttttgg 2940

ctcggtaggc tcgggtccac ctgtctcggt cgttaaggac tttgtgtatg ttttccaaaa 3000

cacggtacag atgggcttgg atgttcaagt acatgggcat gaggccatct ttggggtgga 3060

gataggacca ctgaagagcg tcatgttccg gggtggtatt gtaaatcacc cagtcgtagc 3120

agggtttttg agcgtggaac tggaatatgt ccttcaggag caggctaatg gccaagggta 3180

gacccttagt gtaggtgttt acaaagcggt tgagctggga gggatgcatg cgggggggaga 3240

tgatatgcat cttggcttgg attttgaggt tagctatgtt accacccagg tctctgcggg 3300

ggttcatgtt atgaaggacc accagcacgg tatagccagt gcatttgggg aacttgtcat 3360

gcagtttgga ggggaaggcg tggaagaatt tagatacccc cttgtgcccc cctaggtttt 3420

ccatgcactc atccataata atggcaatgg gacccctggc ggccgcttta gcaaacacgt 3480

tttgggggtt ggaaacatca tagttttgct ctagagtgag ctcatcatag gccatcttta 3540

caaagcgggg taggagggtg cccgactggg ggatgatagt tccatctggg cctggagcgt 3600

agttgccctc acagatctgc atctcccagg ccttaatttc cgagggggggg atcatgtcca 3660

cctgggggggc gataaaaaac acggtttctg gcgggggggtt aatgagctgg gtggaaagca 3720

agttacgcaa cagctgggat ttgccgcaac cggtgggacc gtagatgacc ccgatgacgg 3780

gttgcagctg gtagttcaga gaggaacagc tgccgtcggg gcgcaggagg ggagctacct 3840

cattcatcat gcttctgaca tgtttatttt cactcactaa gttttgcaag agcctctccc 3900

cacccaggga taagagttct tccaggctgt tgaagtgttt cagcggtttc aggccgtcgg 3960

ccatgggcat cttttcaagc gactgacgaa gcaagtacag tcggtcccag agctcggtga 4020

cgtgctctat ggaatctcga tccagcagac ttcttggttt cgggggttgg gccgactttc 4080

gctgtagggc accagccggt gggcgtccag ggccgcgagg gttctgtcct tccagggtct 4140

cagcgttcgg gtgagggtgg tctcggtgac ggtgaaggga tgagccccgg gctgggcgct 4200

tgcgagggtg cgcttcaggc tcatcctgct ggtgctgaag cgggcgtcgt ctccctgtga 4260

gtcggccaga tagcaacgaa gcatgaggtc gtagctgagg gactcggccg cgtgtccctt 4320

ggcgcgcagc tttcccttgg aaacgtgctg acatttggtg cagtgcagac acttgagggc 4380

gtagagtttt ggggccagga agaccgactc gggcgagtag gcgtcggctc cgcactgagc 4440

gcagacggtc tcgcactcca ccagccacgt gagctcgggt ttagcgggat caaaaaccaa 4500

gttgcctcca tttttttga tgcgtttctt accttgcgtc tccatgagtc tgtgtcccgc 4560

ttccgtgaca aaaaggctgt cggtatcccc gtagaccgac ttgagggggc gatcttccaa 4620

aggtgttccg aggtcttccg cgtacaggaa ctgggaccac tccgagacaa aggctcgggt 4680

ccaggctaac acgaaggagg cgatctgcga ggggtatctg tcgttttcaa tgaggggtc 4740

caccttttcc agggtgtgca gacacaggtc gtcctcctcc gcgtccacga aggtgattgg 4800

cttgtaagtg taggtcacgt gacccgcacc cccccaaggg gtataaaagg gggcgtgccc 4860

actctccccg tcactttctt ccgcatcgct gtggaccaga gccagctgtt cgggtgagta 4920

ggccctctca aaagccggca tgatttcggc gctcaagttg tcagtttcta caaacgaggt 4980

ggatttgata ttcacgtgcc ccgcggcgat gcttttgatg gtggaggggt ccatctgatc 5040

agaaaacacg atctttttat tgtcaagttt ggtggcgaaa gacccgtaga gggcgttgga 5100

aagcaacttg gcgatggagc gcagggtctg atttttctcc cgatcggccc tctccttggc 5160

ggcgatgttg agttgcacgt actcgcgggc cacgcaccgc cactcgggga acacggcggt 5220

gcgctcgtcg ggcaggatgc gcacgcgcca gccgcggttg tgcagggtga tgaggtccac 5280

gctggtggcc acctccccgc ggaggggctc gttggtccaa cacaatcgcc ccccttttct 5340

ggagcagaac ggaggcaggg gatctagcaa gttggcgggc ggggggtcgg cgtcgatggt 5400

aaatatgccg ggtagcagaa ttttattaaa ataatcgatt tcggtgtccg tgtcttgcaa 5460

cgcgtcttcc cacttcttca ccgccagggc cctttcgtag ggattcaggg gcggtcccca 5520

gggcatgggg tgggtcaggg ccgaggcgta catgccgcag atgtcgtaca cgtacagggg 5580

ctccctcaac accccgatgt aagtggggta acagcgcccc ccgcggatgc tggctcgcac 5640

gtagtcgtac atctcgtgag agggagccat gagcccgtct cccaagtggg tcttgtgggg 5700

tttttcggcc cggtagagga tctgcctgaa gatggcgtgg gagttggaag agatagtggg 5760

gcgttggaag acgttaaagt tggctccggg cagtcccacg gagtcttgga tgaactgggc 5820

gtaggattcc cggagcttgt ccaccagggc tgcggttacc agcacgtcga gagcgcagta 5880

gtccaacgtc tcgcggacca ggttgtaggc cgtctcttgt tttttctccc acagttcgcg 5940

attgaggagg tattcctcgc ggtctttcca gtactcttcg gcgggaaatc ctttttcgtc 6000

cgctcggtaa gaacctaaca tgtaaaattc gttcacggct ttgtatggac aacagccttt 6060

ttctaccggc agggcgtacg cttgagcggc ctttctgaga gaggtgtggg tgagggcgaa 6120

ggtgtcccgc accatcactt tcaggtactg atgtttgaag tccgtgtcgt cgcaggcgcc 6180

ctgttcccac agcgtgaagt cggtgcgctt tttctgcctg ggattgggga gggcgaatgt 6240

gacgtcgtta aagaggattt tcccggcgcg gggcatgaag ttgcgagaga tcctgaaggg 6300

tccgggcacg tccgagcggt tgttgatgac ttgcgccgcc aggacgatct cgtcgaagcc 6360

gttgatgttg tggcccacga tgtaaagttc gataaagcgc ggctgtccct tgagggccgg 6420

cgctttttte aactcctcgt aggtgagaca gtccggcgag gagagaccca gctccgcccg 6480

ggcccagtcg gagagctgag ggttagccgc gaggaaagag ctccacaggt caagggctag 6540

cagagtttgc aagcggtcgc ggaactcgcg aaacttttte cccacggcca ttttctccgg 6600

cgtcaccacg tagaaagtgc aggggcggtc gttccagacg tcccatcgga gctctagggc 6660

cagctcgcag gcttgacgaa cgagggtctc ctcgcccgag acgtgcatga ccagcatgaa 6720

gggtaccaac tgtttcccga acgagcccat ccatgtgtag gtttctacgt cgtaggtgac 6780

aaagagccgc tgggtgcgcg cgtgggagcc gatcgggaag aagctgatct cctgccacca 6840

gttggaggaa tgggtgttga tgtggtgaaa gtagaagtcc cgccggcgca cagagcattc 6900

gtgctgatgt ttgtaaaagc gaccgcagta gtcgcagcgc tgcacgctct gtatctcctg 6960

aatgagatgc gcttttcgcc cgcgcaccag aaaccggagg gggaagttga gacgggggct 7020

tggtggggcg gcatcccctt cgccttggcg gtgggagtct gcgtctgcgc cctccttctc 7080

tgggtggacg acggtgggga cgacgacgcc ccgggtgccg caagtccaga tctccgccac 7140

ggaggggcgc aggcgttgca ggaggggacg cagctgcccg ctgtccaggg agtcgagggc 7200

ggccgcgctg aggtcggcgg gaagcgtttg caagttcact ttcagaagac cggtaagagc 7260

gtgagccagg tgcacatggt acttgatttc caggggggtg ttggaagagg cgtccacggc 7320

gtagaggagg ccgtgtccgc gcggggccac caccgtgccc cgaggaggtt ttatctcact 7380

cgtcgagggc gagcgccggg gggtagaggc ggctctgcgc cgggggcag cggaggcagt 7440

ggcacgtttt cgtgaggatt cggcagcggt tgatgacgag cccggagact gctggcgtgg 7500

gcgacgacgc ggcggttgag gtcctggatg tgccgtctct gcgtgaagac caccggcccc 7560

cgggtcctga acctgaaaga gagttccaca gaatcaatgt ctgcatcgtt aacggcggcc 7620

tgcctgagga tctcctgtac gtcgcccgag ttgtcttgat aggcgatctc ggccatgaac 7680

tgctccactt cttcctcgcg gaggtcgccg tggcccgctc gctccacggt ggcggccagg 7740

tcgttggaga tgcgacgcat gagttgagag aaggcgttga ggccgttctc gttccacacg 7800

cggctgtaca ccacgtttcc gaaggagtcg cgcgctcgca tgaccacctg ggccacgttg 7860

agttccacgt ggcgggcgaa gacggcgtag tttctgaggc gctggaagag gtagttgagc 7920

gtggtggcga tgtgctcgca gacgaagaag tacatgatcc agcgccgcag ggtcatctcg 7980

ttgatgtctc cgatggcttc gagacgctcc atggcctcgt agaagtcgac ggcgaagttg 8040

aaaaattggg agttgcgggc ggccaccgtg agttcttctt gcaggaggcg gatgagatcg 8100

gcgaccgtgt cgcgcacctc ctgctcgaaa gcgccccgag gcgcctctgc ttcttcctcc 8160

ggctcctcct cttccagggg cacgggttcc tccggcagct ctgcgacggg gacggggcgg 8220

cgacgtcgtc gtctgaccgg caggcggtcc acgaagcgct cgatcatttc gccgcgccgg 8280

cgacgcatgg tctcggtgac ggcgcgtccg ttttcgcgag gtcgcagttc gaagacgccg 8340

ccgcgcagag cgcccccgtg cagggagggt aagtggttag ggccgtcggg cagggacacg 8400

gcgctgacga tgcattttat caattgctgc gtaggcactc cgtgcaggga tctgagaacg 8460

tcgaggtcga cgggatccga gaacttctct aggaaagcgt ctatccaatc gcagtcgcaa 8520

ggtaagctga ggacggtggg ccgctggggg gcgtccgcgg gcagttggga ggtgatgctg 8580

ctgatgatgt aattaaagta ggcggtcttc aggcggcgga tggtggcgag gaggaccacg 8640

tctttgggcc cggcctgttg aatgcgcagg cgctcggcca tgccccaggc ctcgctctga 8700

cagcgacgca ggtctttgta gtagtcttgc atcagtctct ccaccggaac ctctgcttct 8760

cccctgtctg ccatgcgagt cgagccgaac ccccgcaggg gctgcagcaa cgctaggtcg 8820

gccacgaccc tctcggccag cacggcctgt tggatctgcg tgagggtggt ctggaagtcg 8880

tccaggtcca cgaagcggtg ataggccccc gtgttgatgg tgtaggtgca gttggccatg 8940

acggaccagt tgacgacttg catgccgggt tgggtgatct ccgtgtactt gaggcgcgag 9000

taggcgcggg actcgaacac gtagtcgttg catgtgcgta ccagatactg gtagccaacc 9060

aggaagtggg gaggcggttc tcggtacagg ggccagccga ctgtggcggg ggcgccgggg 9120

gacaggtcgt ccagcatgag gcgatggtag tggtagatgt agcgggagag ccaggtgatg 9180

ccggccgagg tggtcgcggc cctggtgaat tcgcggacgc ggttccagat gttgcgcagg 9240

gggcgaaagc gctccatggt gggcacgctc tgccccgtga ggcgggcgca atcttgtacg 9300

ctctagatgg aaaaaagaca gggcggtcat cgactccctt ccgtagctcg gggggtaaag 9360

tcgcaagggt gcggcggcgg ggaaccccgg ttcgagaccg gccggatccg ccgctcccga 9420

tgcgcctggc cccgcatcca cgacgtccgc gtcgagaccc agccgcgacg ctccgcccca 9480

atacggaggg gagtcttttg gtgtttttc gtagatgcat ccggtgctgc ggcagatgcg 9540

acctcagacg cccaccacca ccgccgcggc ggcagtaaac ctgagcggag gcggtgacag 9600

ggaggaggag gagctggctt tagacctgga agagggagag gggctggccc ggctgggagc 9660

gccgtcccca gagagacacc ctagggttca gctcgtgagg gacgccaggc aggcttttgt 9720

gccgaagcag aacctgttta gggaccgcag cggtcaggag gcggaggaga tgcgcgattg 9780

caggtttcgg gcgggtagag agctgagggc gggcttcgat cgggagcggc tcctgagggc 9840

ggaggatttc gagcccgacg agcgttctgg ggtgagcccg gcccgcgctc acgtctcggc 9900

ggccaacctg gtgagcgcgt acgagcagac ggtgaacgag gagcgcaact tccaaaagag 9960

ctttaacaat cacgtgagga ccctgatcgc gagggaggag gtgaccatcg ggctgatgca 10020

tctgtgggac ttcgtggagg cctacgtgca gaacccggcc agcaaacctc tgacggccca 10080

gctgttcctg atcgtgcagc acagccgcga caacgagacg ttccgcgacg ccatgttgaa 10140

catcgcggag cccgagggtc gctggctctt ggatctgatt aacatcctgc agagcatcgt 10200

ggtgcaggag aggggcctca gcttagcgga caaggtggcg gccattaact attcgatgca 10260

gagcctgggg aagttctacg ctcgcaagat ctacaagagc ccttacgtgc ccatagacaa 10320

ggaggtgaag atagacagct tttacatgcg catggcgctg aaggtgctga cgctgagcga 10380

cgacctcggc gtgtaccgta acgacaagat ccacaaggcg gtgagcgcca gccgccggcg 10440

ggagctgagc gacagggagc tgatgcacag cctgcagagg gcgctggcgg gcgccgggga 10500

cgaggagcgc gaggcttact tcgacatggg agccgatctg cagtggcgtc ccagcgcgcg 10560

cgccttggag gcggcgggct accccgacga ggaggatcgg gacgatttgg aggaggcagg 10620

cgagtacgag gacgaagcct gaccgggcag gtgttgtttt agatgcagcg gccggcggac 10680

ggggccaccg cggatcccgc acttttggca tccatgcaga gtcaaccttc gggcgtgacc 10740

gcctccgatg actgggcggc ggccatggac cgcattatgg cgctgactac ccgcaacccc 10800

gaggctttta gacagcaacc ccaggccaac cgtttttcgg ccatcttgga agcggtggtg 10860

ccctcccgca ccaaccccac acacgagaaa gtcctgacta tcgtgaacgc cctggtagac 10920

agcaaggcca tccgccgcga cgaggcgggc ttgatttaca acgctctgct ggaacgggtg 10980

gcgcgctaca acagcactaa cgttcagacc aatctggatc gcctcaccac cgacgtgaag 11040

gaggcgctgg ctcagaagga gcggtttctg agggacagca atctgggctc tctggtggca 11100

ctcaacgcct tcctgagcac gcagccggcc aacgtgcccc gcgggcagga ggactacgtg 11160

agcttcatca gcgctctgag gctgctggtg tccgaggtgc cccagagcga ggtgtatcag 11220

tctgggccgg attacttctt ccagacgtcc cgacagggct tgcaaacggt gaacctgact 11280

caggccttta aaaacttgca aggcatgtgg ggcgttaagg ccccggtggg cgatcgagcc 11340

accatctcca gtctgctgac ccccaacact cgcctgctgc tgctcttgat cgcgccgttc 11400

accaacagta gcactatcag ccgtgactcg tacctgggtc atctcatcac tttgtaccgc 11460

gaggccatcg gtcaggctca gatcgacgag cacacatatc aggagatcac taacgtgagc 11520

cgggccctgg gtcaggaaga taccggcagc ctggaagcca cgttgaactt tttgctaacc 11580

aaccggaggc aaaaaatacc ctcccagttt acgttaagcg ccgaggagga gaggattctg 11640

cgatacgtgc agcagtccgt gagtctgtac ttgatgcggg agggcgccac cgcttccacg 11700

gctttagaca tgacggctcg gaacatggaa ccgtcctttt actccgccca ccggccgttc 11760

attaaccgtc tgatggacta cttccatcgc gcggccgcca tgaacgggga gtacttcacc 11820

aatgccatcc tgaatccgca ttggatgccc ccgtccggct tctacaccgg cgagtttgac 11880

ctgcccgaag ccgacgacgg ctttctttgg gacgacgtgt ccgacagcat tttcacgccg 11940

ggcaatcgcc gattccagaa gaaggagggc ggagacgagc tccccctctc cagcgtggag 12000

gcggcctcta ggggagagag tcccctttccc agtctgtctt ccgccagcag tggtcgggta 12060

acgcgcccgc ggttgccggg ggagagcgac tacctgaacg acccccttgct gcggccggct 12120

aggaagaaaa atttccccaa caacggggtg gaaagcttgg tggataaaat gaatcgttgg 12180

aagacctacg cccaggagca gcggtagtgg gaggacagtc agccgcgacc gctggttccg 12240

ccgcactggc gtcgtcagag agaagacccg gacgactccg cagacgatag tagcgtgttg 12300

gacctgggag ggagcggagc caacccccttt gctcacttgc aacccaaggg gcgttccagt 12360

cgcctctact aataaaaaag acgcggaaac ttaccagagc catggccaca gcgtgtgtcc 12420

```
tttcttcctc tctttcttcc tcggcgcggc aga atg aga aga gcg gtg aga gtc   12474
                                      Met Arg Arg Ala Val Arg Val
                                      1               5

acg ccg gcg gcg tat gag ggt ccg ccc cct tct tac gaa agc gtg atg    12522
Thr Pro Ala Ala Tyr Glu Gly Pro Pro Ser Tyr Glu Ser Val Met
        10                  15                  20

gga tca gcg aac gtg ccg gcc acg ctg gag gcg cct tac gtt cct ccc    12570
Gly Ser Ala Asn Val Pro Ala Thr Leu Glu Ala Pro Tyr Val Pro Pro
    25                  30                  35

aga tac ctg gga cct acg gag ggc aga aac agc atc cgt tac tcc gag    12618
Arg Tyr Leu Gly Pro Thr Glu Gly Arg Asn Ser Ile Arg Tyr Ser Glu
40                  45                  50                  55

ctg gca ccc ctg tac gat acc acc aag gtg tac ctg gtg gac aac aag    12666
Leu Ala Pro Leu Tyr Asp Thr Thr Lys Val Tyr Leu Val Asp Asn Lys
                60                  65                  70

tcg gcg gac atc gcc tcc ctg aat tat caa aac gat cac agc aat ttt    12714
Ser Ala Asp Ile Ala Ser Leu Asn Tyr Gln Asn Asp His Ser Asn Phe
            75                  80                  85

ctg act acc gtg gtg cag aac aat gac ttc acc ccg acg gag gcg ggc    12762
Leu Thr Thr Val Val Gln Asn Asn Asp Phe Thr Pro Thr Glu Ala Gly
        90                  95                  100

acg cag acc att aac ttt gac gag cgt tcc cgc tgg ggc ggt cag ctg    12810
Thr Gln Thr Ile Asn Phe Asp Glu Arg Ser Arg Trp Gly Gly Gln Leu
    105                 110                 115

aaa acc atc ctg cac acc aac atg ccc aac atc aac gag ttc atg tcc    12858
Lys Thr Ile Leu His Thr Asn Met Pro Asn Ile Asn Glu Phe Met Ser
120                 125                 130                 135

acc aac aag ttc agg gcc agg ctg atg gtt aaa aag gct gaa aac cag    12906
Thr Asn Lys Phe Arg Ala Arg Leu Met Val Lys Lys Ala Glu Asn Gln
                140                 145                 150

cct ccc gag tac gaa tgg ttt gag ttc acc att ccc gag ggc aac tat    12954
Pro Pro Glu Tyr Glu Trp Phe Glu Phe Thr Ile Pro Glu Gly Asn Tyr
            155                 160                 165

tcc gag acc atg act atc gat ctg atg aac aat gcg atc gtg gac aat    13002
Ser Glu Thr Met Thr Ile Asp Leu Met Asn Asn Ala Ile Val Asp Asn
        170                 175                 180

tac ctg caa gtg ggg agg cag aac ggg gta ttg gaa agc gat atc ggc    13050
Tyr Leu Gln Val Gly Arg Gln Asn Gly Val Leu Glu Ser Asp Ile Gly
    185                 190                 195

gta aaa ttt gat acc aga aac ttc cga ctg ggg tgg gat ccc gtg acc    13098
Val Lys Phe Asp Thr Arg Asn Phe Arg Leu Gly Trp Asp Pro Val Thr
200                 205                 210                 215
```

```
aag ctg gtg atg cca ggc gtg tac acc aac gag gct ttt cac ccc gac    13146
Lys Leu Val Met Pro Gly Val Tyr Thr Asn Glu Ala Phe His Pro Asp
                220                 225                 230

atc gtg ctg ctg ccg ggg tgc ggt gtg gac ttc act cag agc cgt ttg    13194
Ile Val Leu Leu Pro Gly Cys Gly Val Asp Phe Thr Gln Ser Arg Leu
            235                 240                 245

agt aac ctg tta ggg atc aga aag cgc cgc ccc ttc caa gag ggc ttt    13242
Ser Asn Leu Leu Gly Ile Arg Lys Arg Arg Pro Phe Gln Glu Gly Phe
        250                 255                 260

cag atc atg tat gag gac ctg gaa gga ggt aac att cca ggt ttg cta    13290
Gln Ile Met Tyr Glu Asp Leu Glu Gly Gly Asn Ile Pro Gly Leu Leu
    265                 270                 275

gac gtg ccg gcg tat gaa gag agt gtt aaa cag gcg gag gcg cag gga    13338
Asp Val Pro Ala Tyr Glu Glu Ser Val Lys Gln Ala Glu Ala Gln Gly
280                 285                 290                 295

cga gag att cga ggc gac acc ttt gcc acg gaa cct cac gaa ctg gta    13386
Arg Glu Ile Arg Gly Asp Thr Phe Ala Thr Glu Pro His Glu Leu Val
                300                 305                 310

ata aaa cct ctg gaa caa gac agt aaa aaa cgg agt tac aac att ata    13434
Ile Lys Pro Leu Glu Gln Asp Ser Lys Lys Arg Ser Tyr Asn Ile Ile
            315                 320                 325

tcc ggc act atg aat acc ttg tac cgg agc tgg ttt ctg gct tac aac    13482
Ser Gly Thr Met Asn Thr Leu Tyr Arg Ser Trp Phe Leu Ala Tyr Asn
        330                 335                 340

tac ggg gat ccc gaa aag gga gtg aga tca tgg acc ata ctc acc acc    13530
Tyr Gly Asp Pro Glu Lys Gly Val Arg Ser Trp Thr Ile Leu Thr Thr
    345                 350                 355

acg gac gtg acc tgc ggc tcg cag caa gtg tac tgg tcc ctg ccg gat    13578
Thr Asp Val Thr Cys Gly Ser Gln Gln Val Tyr Trp Ser Leu Pro Asp
360                 365                 370                 375

atg atg caa gac ccg gtc acc ttc cgc ccc tcc acc caa gtc agc aac    13626
Met Met Gln Asp Pro Val Thr Phe Arg Pro Ser Thr Gln Val Ser Asn
                380                 385                 390

ttc ccg gtg gtg ggc acc gag ctg ctg ccc gtc cat gcc aag agc ttc    13674
Phe Pro Val Val Gly Thr Glu Leu Leu Pro Val His Ala Lys Ser Phe
            395                 400                 405

tac aac gaa cag gcc gtc tac tcg caa ctc att cgc cag tcc acc gcg    13722
Tyr Asn Glu Gln Ala Val Tyr Ser Gln Leu Ile Arg Gln Ser Thr Ala
        410                 415                 420

ctt acc cac gtg ttc aat cgc ttt ccc gag aac cag att ctg gtg cgc    13770
Leu Thr His Val Phe Asn Arg Phe Pro Glu Asn Gln Ile Leu Val Arg
    425                 430                 435

cct ccc gct cct acc att acc acc gtc agt gaa aac gtt ccc gcc ctc    13818
```

```
Pro Pro Ala Pro Thr Ile Thr Thr Val Ser Glu Asn Val Pro Ala Leu
440                 445                 450                 455

aca gat cac gga acc ctg ccg ctg cgc agc agt atc agt gga gtt cag    13866
Thr Asp His Gly Thr Leu Pro Leu Arg Ser Ser Ile Ser Gly Val Gln
                460                 465                 470

cgc gtg acc atc acc gac gcc aga cgt cga acc tgt ccc tac gtt tac    13914
Arg Val Thr Ile Thr Asp Ala Arg Arg Arg Thr Cys Pro Tyr Val Tyr
            475                 480                 485

aaa gct ctt ggc gta gtg gct cct aaa gtg ctc tct agt cgc acc ttc    13962
Lys Ala Leu Gly Val Val Ala Pro Lys Val Leu Ser Ser Arg Thr Phe
        490                 495                 500

taa acatgtccat cctcatctct cccgataaca acaccggctg gggactgggc         14015

tccggcaaga tgtacggcgg agccaaaagg cgctccagtc agcacccagt tcgagttcgg  14075

ggccacttcc gtgctccctg gggagcttac aagcgaggac tctcgggccg aacggcggta  14135

gacgatacca tagatgccgt gattgccgac gcccgccggt acaaccccgg accggtcgct  14195

agcgccgcct ccaccgtgga ttccgtgatc gacagcgtgg tagctggcgc tcgggcctat  14255

gctcgccgca agaggcggct gcatcggaga cgtcgcccca ccgccgccat gctggcagcc  14315

agggccgtgc tgaggcgggc ccggagggta ggcagaaggg ctatgcgccg cgctgccgcc  14375

aacgccgccg ccgggagggc ccgccgacag gctgcccgcc aggctgctgc cgccatcgct  14435

agcatggcca gacccaggag agggaacgtg tactgggtgc gcgattctgt gacgggagtc  14495

cgagtgccgg tgcgcagccg acctccccga agttagaaga tccaagctgc gaagacggcg  14555

gtactgagtc tccctgttgt tatcagccca acatgagcaa gcgcaagttt aaagaagaac  14615

tgctgcagac gctggtgcct gagatctatg gccctccgga cgtgaagcct gacattaagc  14675

cccgcgatat caagcgtgtt aaaaagcggg aaaagaaaga ggaactcgcg gtggtagacg  14735

atggcggagt ggaatttatt aggagtttcg ccccgcgacg cagggttcaa tggaaagggc  14795

ggcgggtaca acgcgttttg aggccgggca ccgcggtagt ttttaccccg ggagagcggt  14855

cggccgttag gggtttcaaa aggcagtacg acgaggtgta cggcgacgag gacatattgg  14915

aacaggcggc tcaacagatc ggagaatttg cctacggaaa gcgttcgcgt cgcgaagacc  14975

tggccatcgc tttagacagc ggcaacccca cgcccagcct caaacctgtg acgctgcagc  15035

aggtgctccc cgtgagcgcc agcacggaca gcaagagggg aataaaaaga gaaatggaag  15095

atctgcagcc caccatccag ctcatggtcc ctaaacggca gaggctggaa gaggtcctgg  15155

agaaaatgaa agtggaccca agcatagagc cggacgtcaa agtcaggccg atcaaagaag  15215
```

```
tggcccctgg tctcggggtg cagacggtgg atatccagat ccccgtcacg tcagcttcga  15275

ccgccgtgga agccatggaa acgcaaacgg aaacccctgc cgcgatcggt accagggaag  15335

tggcgttgca aaccgacccc tggtacgaat acgccgcccc tcggcgtcag aggcgacccg  15395

ctcgttacgg ccccgccaac gccatcatgc cagaatatgc gctgcatccg tctatcctgc  15455

ccacccccgg ctaccgggga gtgacgtatc gcccgtcagg aacccgccgc cgaacccgtc  15515

gccgccgccg ctcccgtcgt gctctggccc ccgtgtcggt gcgccgcgta acacgccggg  15575

gaaagacagt taccattccc aacccgcgct accaccctag catcctttaa tgactctgcc  15635

gttttgcaga tggctctgac ttgccgcgtg cgccttcccg ttccgcacta tcgaggaaga  15695

tctcgtcgta ggagaggcat ggcgggtagt ggtcgccggc gggctttgcg caggcgcatg  15755

aaaggcggaa ttttacccgc tctgataccc ataatcgccg ccgccatcgg tgccataccc  15815

ggcgtcgctt cagtggcctt gcaagcagct cgtaataaat aaacgaaggc ttttgcactt  15875

atgtcctggt cctgactatt ttatgcagaa agagcatgga agacatcaat tttacgtcgc  15935

tggctccgcg gcacggctcg cggccgctca tgggcacctg gaacgacatc ggcaccagtc  15995

agctcaacgg gggcgctttc aattgggggа gcctttggag cggcattaaa aactttggct  16055

ccacgattaa atcctacggc agcaaagcct ggaacagtag tgctggtcag atgctccgag  16115

ataaactgaa ggacaccaac ttccaagaaa aagtggtcaa tggggtggtg accggcatcc  16175

acggcgcggt agatctcgcc aaccaagcgg tgcagaaaga gattgacagg cgtttggaaa  16235

gctcgcgggt gccgccgcag agaggggatg aggtggaggt cgaggaagta gaagtagagg  16295

aaaagctgcc cccgctggag aaagttcccg gtgcgcctcc gagaccgcag aagcgaccca  16355

ggccagaact agaagaaact ctggtgacgg agagcaagga gcctccctcg tacgagcaag  16415

ccttgaaaga gggcgcctct ccaccctacc caatgacaaa accgatcgcg cctatggctc  16475

ggccggtgta cgggaaggac tacaagcctg tcacgctaga gctccccccg ccgccaccgc  16535

cgcccccccac gcgcccgacc gttcccccccc ccctgccggc tccgtcggcg ggacccgtgt  16595

ccgcacccgt cgccgtgcct ctgccagccg cccgcccagt ggccgtggcc actgccagaa  16655

accccagagg ccagagagga gccaactggc aaagcacgct gaacagcatc gtgggcctgg  16715

gagtgaaaag cctgaaacgc cgccgttgct attattaaaa gtgtagctaa aaaatttccc  16775

gttgtatacg cctcctatgt taccgccaga gacgcgtgac tgtcgccgcg agcgccgctt  16835

tcaag atg gcc acc cca tcg atg atg ccg cag tgg tct tac atg cac atc  16885
     Met Ala Thr Pro Ser Met Met Pro Gln Trp Ser Tyr Met His Ile
         505                 510                 515
```

```
gcc ggg cag gac gcc tcg gag tac ctg agc ccc ggt ctc gtg cag ttc    16933
Ala Gly Gln Asp Ala Ser Glu Tyr Leu Ser Pro Gly Leu Val Gln Phe
    520                 525                 530

gcc cgc gcc acc gac acc tac ttc agc ttg gga aac aag ttt aga aac    16981
Ala Arg Ala Thr Asp Thr Tyr Phe Ser Leu Gly Asn Lys Phe Arg Asn
535                 540                 545                 550

ccc acc gtg gcc ccc acc cac gat gta acc acg gac cgc tcg caa agg    17029
Pro Thr Val Ala Pro Thr His Asp Val Thr Thr Asp Arg Ser Gln Arg
                555                 560                 565

ctg acc ctg cgt ttt gtg ccc gta gac cgg gag gac acc gcg tac tct    17077
Leu Thr Leu Arg Phe Val Pro Val Asp Arg Glu Asp Thr Ala Tyr Ser
            570                 575                 580

tac aaa gtg cgc tac acg ctg gcc gta ggg gac aac cga gtg ctg gac    17125
Tyr Lys Val Arg Tyr Thr Leu Ala Val Gly Asp Asn Arg Val Leu Asp
        585                 590                 595

atg gcc agc acc tac ttt gac atc cgg gga gtg ctg gat cgc ggt ccc    17173
Met Ala Ser Thr Tyr Phe Asp Ile Arg Gly Val Leu Asp Arg Gly Pro
    600                 605                 610

agt ttt aag ccc tac tcg ggt acc gcg tac aat tcc ctg gct ccc aag    17221
Ser Phe Lys Pro Tyr Ser Gly Thr Ala Tyr Asn Ser Leu Ala Pro Lys
615                 620                 625                 630

ggc gct ccc aac cct gca gaa tgg acg aat tca gac agc aaa gtt aaa    17269
Gly Ala Pro Asn Pro Ala Glu Trp Thr Asn Ser Asp Ser Lys Val Lys
                635                 640                 645

gtg agg gca cag gcg cct ttt gtt agc tcg tat ggt gct aca gcg att    17317
Val Arg Ala Gln Ala Pro Phe Val Ser Ser Tyr Gly Ala Thr Ala Ile
            650                 655                 660

aca aaa gag ggt att cag gtg gga gta acc tta aca gac tcc gga tca    17365
Thr Lys Glu Gly Ile Gln Val Gly Val Thr Leu Thr Asp Ser Gly Ser
        665                 670                 675

aca cca cag tat gca gat aaa acg tat cag cct gag ccg caa att gga    17413
Thr Pro Gln Tyr Ala Asp Lys Thr Tyr Gln Pro Glu Pro Gln Ile Gly
    680                 685                 690

gaa cta cag tgg aac agc gat gtt gga acc gat gac aaa ata gca gga    17461
Glu Leu Gln Trp Asn Ser Asp Val Gly Thr Asp Asp Lys Ile Ala Gly
695                 700                 705                 710

aga gtg cta aag aaa aca acg ccc atg ttc cct tgt tac ggc tca tat    17509
Arg Val Leu Lys Lys Thr Thr Pro Met Phe Pro Cys Tyr Gly Ser Tyr
                715                 720                 725

gcc agg ccc act aat gaa aaa gga gga cag gca aca ccg tcc gct agt    17557
Ala Arg Pro Thr Asn Glu Lys Gly Gly Gln Ala Thr Pro Ser Ala Ser
            730                 735                 740
```

```
caa gac gtg caa aat ccc gaa tta caa ttt ttt gcc tct act aat gtc    17605
Gln Asp Val Gln Asn Pro Glu Leu Gln Phe Phe Ala Ser Thr Asn Val
        745                 750                 755

gcc aat aca cca aaa gca gtt cta tat gcg gag gac gtg tca att gaa    17653
Ala Asn Thr Pro Lys Ala Val Leu Tyr Ala Glu Asp Val Ser Ile Glu
    760                 765                 770

gcg cca gac act cac ttg gtg ttc aaa cca aca gtc act gaa ggc att    17701
Ala Pro Asp Thr His Leu Val Phe Lys Pro Thr Val Thr Glu Gly Ile
775                 780                 785                 790

aca agt tca gag gct cta ctg acc caa caa gct gct ccc aac cgt cca    17749
Thr Ser Ser Glu Ala Leu Leu Thr Gln Gln Ala Ala Pro Asn Arg Pro
                795                 800                 805

aac tac ata gcc ttt aga gat aat ttt att ggt ctc atg tac tac aat    17797
Asn Tyr Ile Ala Phe Arg Asp Asn Phe Ile Gly Leu Met Tyr Tyr Asn
            810                 815                 820

agc aca ggt aac atg gga gta ctg gca ggc cag gct tct cag cta aat    17845
Ser Thr Gly Asn Met Gly Val Leu Ala Gly Gln Ala Ser Gln Leu Asn
        825                 830                 835

gca gtt gtt gac ctg caa gac aga aat act gag ctg tcc tac caa ctc    17893
Ala Val Val Asp Leu Gln Asp Arg Asn Thr Glu Leu Ser Tyr Gln Leu
    840                 845                 850

atg ttg gac gcc ctc gga gac cgc agt cgg tac ttt tct atg tgg aac    17941
Met Leu Asp Ala Leu Gly Asp Arg Ser Arg Tyr Phe Ser Met Trp Asn
855                 860                 865                 870

caa gct gtg gat agt tac gat cct gat gta aga atc ata gaa aac cat    17989
Gln Ala Val Asp Ser Tyr Asp Pro Asp Val Arg Ile Ile Glu Asn His
                875                 880                 885

ggc gta gaa gat gaa ttg cct aat tat tgc ttt cct ttg gga ggc atg    18037
Gly Val Glu Asp Glu Leu Pro Asn Tyr Cys Phe Pro Leu Gly Gly Met
            890                 895                 900

gca gta acc gac acc tac tcg cct ata aag gtt aat gga gga ggc aat    18085
Ala Val Thr Asp Thr Tyr Ser Pro Ile Lys Val Asn Gly Gly Gly Asn
        905                 910                 915

gga tgg gaa gcc aat aac ggc gtt ttc acc gaa aga gga gtg gaa ata    18133
Gly Trp Glu Ala Asn Asn Gly Val Phe Thr Glu Arg Gly Val Glu Ile
    920                 925                 930

ggt tca ggg aac atg ttt gcc atg gag att aac ctg caa gcc aac cta    18181
Gly Ser Gly Asn Met Phe Ala Met Glu Ile Asn Leu Gln Ala Asn Leu
935                 940                 945                 950

tgg cgt agc ttt ctg tac tcc aat att ggg ctg tac ctg cca gac tct    18229
Trp Arg Ser Phe Leu Tyr Ser Asn Ile Gly Leu Tyr Leu Pro Asp Ser
                955                 960                 965

ctc aaa atc act cct gac aac atc aca ctc cca gag aac aaa aac acc    18277
```

```
Leu Lys Ile Thr Pro Asp Asn Ile Thr Leu Pro Glu Asn Lys Asn Thr
            970                 975                 980

tat cag tat atg aac ggt cgc gtg acg cca ccc ggg ctg gtt gac acc    18325
Tyr Gln Tyr Met Asn Gly Arg Val Thr Pro Pro Gly Leu Val Asp Thr
        985                 990                 995

tac gtt  aac gtg ggc gcg cgc  tgg tcc ccc gat gtc  atg gac agt     18370
Tyr Val  Asn Val Gly Ala Arg  Trp Ser Pro Asp Val .Met Asp Ser
    1000                1005                1010

att aac  cct ttt aat cac cac  cgc aac gcc gga ctc  cgc tac cgt     18415
Ile Asn  Pro Phe Asn His His  Arg Asn Ala Gly Leu  Arg Tyr Arg
    1015                1020                1025

tcc atg  ctc ctg gga aac gga  cgc tac gtg ccc ttc  cac atc cag     18460
Ser Met  Leu Leu Gly Asn Gly  Arg Tyr Val Pro Phe  His Ile Gln
    1030                1035                1040

gtg ccc  cag aaa ttc ttt gca  att aaa aac ctg ctg  ctg ctc ccc     18505
Val Pro  Gln Lys Phe Phe Ala  Ile Lys Asn Leu Leu  Leu Leu Pro
    1045                1050                1055

ggt tcc  tac acc tac gag tgg  aac ttc cgc aag gac  gtg aac atg     18550
Gly Ser  Tyr Thr Tyr Glu Trp  Asn Phe Arg Lys Asp  Val Asn Met
    1060                1065                1070

atc ttg  cag agc tcg ctg ggc  aat gac ctg cga gtg  gac ggg gcc     18595
Ile Leu  Gln Ser Ser Leu Gly  Asn Asp Leu Arg Val  Asp Gly Ala
    1075                1080                1085

agc atc  cgc ttc gac agc atc  aac ctg tac gcc aac  ttt ttc ccc     18640
Ser Ile  Arg Phe Asp Ser Ile  Asn Leu Tyr Ala Asn  Phe Phe Pro
    1090                1095                1100

atg gcc  cac aac acg gcc tcc  acc ctg gaa gcc atg  ctg cgc aac     18685
Met Ala  His Asn Thr Ala Ser  Thr Leu Glu Ala Met  Leu Arg Asn
    1105                1110                1115

gac acc  aac gac caa tct ttc  aac gac tac ctg tgc  gcg gcc aac     18730
Asp Thr  Asn Asp Gln Ser Phe  Asn Asp Tyr Leu Cys  Ala Ala Asn
    1120                1125                1130

atg ctg  tac ccc atc ccc gcc  aac gcc acc agc gtg  ccc atc tcc     18775
Met Leu  Tyr Pro Ile Pro Ala  Asn Ala Thr Ser Val  Pro Ile Ser
    1135                1140                1145

att ccc  tct cgc aac tgg gca  gcc ttc agg ggc tgg  agt ttc acc     18820
Ile Pro  Ser Arg Asn Trp Ala  Ala Phe Arg Gly Trp  Ser Phe Thr
    1150                1155                1160

cgc ctc  aaa acc aag gag acc  ccc tcg ctg ggc tcc  ggg ttc gac     18865
Arg Leu  Lys Thr Lys Glu Thr  Pro Ser Leu Gly Ser  Gly Phe Asp
    1165                1170                1175

ccc tac  ttc gtc tac tcc ggc  tcc atc ccc tac ctg  gac ggc acc     18910
Pro Tyr  Phe Val Tyr Ser Gly  Ser Ile Pro Tyr Leu  Asp Gly Thr
```

```
    1180                1185                1190

ttc tac  ctc aac cat act ttc  aaa aag gtg tca atc  atg ttc gac     18955
Phe Tyr  Leu Asn His Thr Phe  Lys Lys Val Ser Ile  Met Phe Asp
    1195                1200                1205

tcc tcc  gtc agc tgg ccc ggc  aac gac cgt ctg ctg  acg ccc aac     19000
Ser Ser  Val Ser Trp Pro Gly  Asn Asp Arg Leu Leu  Thr Pro Asn
    1210                1215                1220

gag ttc  gaa atc aag cgt tcg  gtg gac ggt gaa ggg  tac aac gtg     19045
Glu Phe  Glu Ile Lys Arg Ser  Val Asp Gly Glu Gly  Tyr Asn Val
    1225                1230                1235

gct cag  agc aac atg acc aag  gac tgg ttc ctg att  cag atg ctc     19090
Ala Gln  Ser Asn Met Thr Lys  Asp Trp Phe Leu Ile  Gln Met Leu
    1240                1245                1250

agc cac  tac aac atc ggc tac  cag ggc ttc tac gtg  ccc gaa aat     19135
Ser His  Tyr Asn Ile Gly Tyr  Gln Gly Phe Tyr Val  Pro Glu Asn
    1255                1260                1265

tac aag  gac cgc atg tac tct  ttc ttc aga aac ttc  caa ccc atg     19180
Tyr Lys  Asp Arg Met Tyr Ser  Phe Phe Arg Asn Phe  Gln Pro Met
    1270                1275                1280

agc cgc  caa att gta gat tca  acg gct tac act aat  tat cag gat     19225
Ser Arg  Gln Ile Val Asp Ser  Thr Ala Tyr Thr Asn  Tyr Gln Asp
    1285                1290                1295

gtg aaa  ctg cca tac cag cat  aac aac tca ggg ttc  gtg ggc tac     19270
Val Lys  Leu Pro Tyr Gln His  Asn Asn Ser Gly Phe  Val Gly Tyr
    1300                1305                1310

atg gga  ccc acc atg cga gag  ggg cag gcc tac ccg  gcc aac tat     19315
Met Gly  Pro Thr Met Arg Glu  Gly Gln Ala Tyr Pro  Ala Asn Tyr
    1315                1320                1325

ccc tat  ccc ctg att ggg gcc  acc gcc gtg ccc agc  ctc acg cag     19360
Pro Tyr  Pro Leu Ile Gly Ala  Thr Ala Val Pro Ser  Leu Thr Gln
    1330                1335                1340

aaa aag  ttc ctc tgc gac cgg  gtg atg tgg agg atc  ccc ttc tct     19405
Lys Lys  Phe Leu Cys Asp Arg  Val Met Trp Arg Ile  Pro Phe Ser
    1345                1350                1355

agc aac  ttc atg tct atg ggc  tcc ctc acc gac ctg  ggg cag aac     19450
Ser Asn  Phe Met Ser Met Gly  Ser Leu Thr Asp Leu  Gly Gln Asn
    1360                1365                1370

atg ctg  tac gcc aac tcc gct  cac gcc ttg gat atg  acc ttt gag     19495
Met Leu  Tyr Ala Asn Ser Ala  His Ala Leu Asp Met  Thr Phe Glu
    1375                1380                1385

gtg gat  ccc atg gat gag ccc  acg ctt ctc tat gtt  ctg ttt gaa     19540
Val Asp  Pro Met Asp Glu Pro  Thr Leu Leu Tyr Val  Leu Phe Glu
    1390                1395                1400
```

```
gtc ttc  gac gtg gtg cgc atc  cac cag ccg cac cgc  ggc gtc atc     19585
Val Phe  Asp Val Val Arg Ile  His Gln Pro His Arg  Gly Val Ile
    1405                 1410                 1415

gag gcc  gtc tac ctg cgc aca  cct ttc tct gcc ggt  aac gcc acc     19630
Glu Ala  Val Tyr Leu Arg Thr  Pro Phe Ser Ala Gly  Asn Ala Thr
    1420                 1425                 1430

acc taa agaagccgat gggctccagc gaacaggagc tgcaggccat tgttcgcgac     19686
Thr

ctgggctgcg ggccctactt tttgggcacc ttcgacaagc gttttcccgg cttcatgtcc  19746

ccccacaagc cggcctgtgc catcgttaac acggccggac gggagaccgg gggggtccac  19806

tggctcgcct tcgcctggaa cccgcgtaac cgcacctgct acctgttcga ccctttttggt  19866

ttctccgacg aaaggctgaa gcagatctac cagttcgagt acgaggggct cctcaagcgc  19926

agcgctctgg cctccacgcc cgaccactgc gtcaccctgg aaaagtccac ccaaacggtc  19986

caggggcccc tctcggccgc ctgcgggctc ttctgttgca tgtttttgca cgccttcgtg  20046

cactggcctc acacccccat ggatcacaac cccaccatgg atctgctcac cggagtgccc  20106

aacagcatgc ttcacagccc ccaggtcgcc cccaccctgc gccgtaacca ggaacacctg  20166

tatcgctttc tggggaaaca ctctgcctat tttcgccgcc accggcagcg catcgaacgg  20226

gccacggcct tcgaaagcat gagccaaaga gtgtaatcaa taaaaaacat ttttatttga  20286

catgatacgc gcttctggcg ttttattaaa aatcgaaggg ttcgagggag gggtcctcgt  20346

gcccgctggg gagggacacg ttgcgatact ggaaacgggc gctccaacga aactcgggga  20406

tcaccagccg cggcaggggc acgtcttcta ggttctgctt ccaaaactgc cgcaccagct  20466

gcagggctcc catgacgtcg ggcgccgata tcttgaagtc gcagttaggg ccggagctcc  20526

cgcggctgtt gcggaacacg gggttggcac actggaacac cagcacgccg gggttgtgga  20586

tactggccag ggccgtcggg tcggtcacct ccgacgcatc cagatcctcg gcgttgctca  20646

gggcaaacgg ggtcagcttg cacatctgcc gcccaatctg gggtactagg tcgcgcttgt  20706

tgaggcagtc gcagcgcaga gggatcagga tgcgtcgctg cccgcgttgc atgatagggt  20766

aactcgccgc caggaactcc tccatttgac ggaaggccat ctgggctttg ccgccctcgg  20826

tgtagaatag cccgcaggac ttgctagaga atacgttatg accgcagttg acgtcctccg  20886

cgcagcagcg ggcgtcttcg ttcttcagct gaaccacgtt gcggccccaa cggttctgga  20946

ccaccttggc tctagtgggg tgctccttca gcgcccgctg tccgttctcg ctggttacat  21006
```

ccatttccaa cacgtgctcc ttgcagacca tctccactcc gtggaagcaa aacaggacgc 21066

cctcctgctg ggtactgcga tgctcccata cggcgcatcc ggtgggctcc cagctcttgt 21126

gttttacccc cgcgtaggct tccatgtaag ccataaggaa tctgcccatc agctcggtga 21186

aggtcttctg gttggtgaag gttagcggca ggccgcggtg ctcctcgttc aaccaagttt 21246

gacagatctt gcggtacacc gctccctggt cgggcagaaa cttaaaagcc gctctgctgt 21306

cgttgtctac gtggaacttc tccattaaca tcatcatggt ttccataccc ttctcccacg 21366

ctgtcaccag tggtttgctg tcggggttct tcaccaacac ggcggtagag gggccctcgc 21426

cggccccgac gtccttcatg gtcattcttt gaaactccac ggagccgtcc gcgcgacgta 21486

ctctgcgcac cggagggtag ctgaagccca cctccaccac ggtgccttcg ccctcgctgt 21546

cggagacaat ctccggggat ggcggcggcg cgggtgtcgc cttgcgagcc ttcttcttgg 21606

gagggagctg aggcgcctcc tgctcgcgct cggggctcat ctcccgcaag taggggtaa 21666

tggagctgcc tgcttggttc tgacggttgg ccattgtatc ctaggcagaa agacatggag 21726

cttatgcgcg aggaaacttt aaccgccccg tcccccgtca gcgacgaaga tgtcatcgtc 21786

gaacaggacc cgggctacgt tacgccgccc gaggatctgg aggggcctga ccggcgcgac 21846

gctagtgagc ggcaggaaaa tgagaaagag gaggcctgct acctcctgga aggcgacgtt 21906

ttgctaaagc atttcgccag gcagagcacc atagttaagg aggccttgca agaccgctcc 21966

gaggtgccct tggacgtcgc cgcgctctcc caggcctacg aggcgaacct tttctcgcct 22026

cgagtgcctc cgaagagaca gcccaacggc acctgcgagc ccaacccgcg actcaacttc 22086

taccccgtgt tcgccgtacc agaggcgctg gccacctatc acatttttt caaaaaccaa 22146

cgcatccccc tatcgtgccg ggccaaccgc accgcggccg ataggaatct caggcttaaa 22206

aacggagcca acatacctga tatcacgtcg ctggaggaag tgcccaagat tttcgagggt 22266

ctgggtcgag atgagaagcg ggcggcgaac gctctgcaga aagaacagaa agagagtcag 22326

aacgtgctgg tggagctgga gggggacaac gcgcgtctgg ccgtcctcaa acgctgcata 22386

gaagtctccc acttcgccta ccccgccctc aacttgccac ccaaagttat gaaatcggtc 22446

atggatcagc tgctcatcaa gagagctgag cccctggatc ccgaccaccc cgaggcggaa 22506

aactcagagg acggaaagcc cgtcgtcagc gacgaggagc tcgagcggtg gctggaaacc 22566

agggaccccc aacagttgca agagaggcgc aagatgatga tggcggccgt gctggtcacc 22626

gtggagctgg aatgcctgca acggtttttc agcgacgtgg agacgctacg caaaatcggg 22686

gaatccctgc actacacctt ccgccagggc tacgtccgcc aggcctgcaa gatctccaac 22746

gtggagctca gcaacctggt ctcctacatg ggcatcctcc acgagaaccg gctggggcag 22806

agcgtgctgc actgcacctt gcaaggcgag gcgcggcggg actacgtgcg agactgcatc 22866

tacctcttcc tcaccctcac ctggcagacc gccatgggcg tctggcagca gtgcttggaa 22926

gagagaaacc tcaaagagct agacaaactc ctctgccgcc agcggcgcgc cctgtggtcc 22986

ggtttcagcg agcgcacggt cgccagcgct ctggcggaca tcatcttccc ggagcgcctg 23046

atgaaaacct tgcaaaacgg cctgccggat ttcatcagtc aaagcatttt gcaaaacttc 23106

cgctcttttg tcctggaacg ctccgggatc ttgcccgcca tgagctgcgc gctaccttct 23166

gactttgtcc ccctctccta ccgcgagtgc cctcccccac tgtggagcca ctgctacctc 23226

ttccaactgg ccaactttct ggcctaccac tccgacctca tggaagacgt aagcggagag 23286

ggtttactgg agtgccactg ccgctgcaac ctgtgcaccc cccacagatc gctggcctgc 23346

aacaccgagc tactcagcga aacccaggtc ataggtacct tcgagatcca ggggccccag 23406

cagcaagagg gtgcttccgg cttgaagctc actccggcgc tgtggacctc ggcttactta 23466

cgcaaatttg tagccgagga ctaccacgcc cacaaaattc agttttacga agaccaatct 23526

cgaccaccga aagccccccc cacggcctgc gtcatcaccc agagcaagat cctggcccaa 23586

ttgcaatcca tcaaccaagc gcgccgcgat ttccttttga aaaagggtcg gggggtgtac 23646

ctggaccccc agaccggcga ggaactcaac ccgtccacac tctccgtcga agcagccccc 23706

ccgagacatg ccgcccaagg gaaccgccaa gcagctgatc gctcggcaga gagcgaagaa 23766

gcaagagctg ctccagcagc aggtggagga cgaggaagag atgtgggaca gccaggcaga 23826

ggaggtgtca gaggacgagg aggagatgga aagctgggac agcctagacg aggaggagga 23886

cgagctttca gaggaagagg cgaccgaaga aaaaccacct gcatccagcg cgccttctct 23946

gagccgacag ccgaagcccc ggccccccgac gcccccggcc ggctcactca aagccagccg 24006

taggtgggac gccaccgaat ctccagcggc agcggcaacg gcagcgggta aggccaaacg 24066

cgagcggcgg gggtattgct cctggcgggc ccacaaaagc agtattgtga actgcttgca 24126

acactgcggg ggaaacatct cctttgcccg acgctacctc ctcttccatc acggtgtggc 24186

cttccctcgc aacgttctct attattaccg tcatctctac agcccctacg aaacgctcgg 24246

agaaaaaagc taaggcctcc tccgccgcga ggaaaaactc cgccgccgct gccgccgcca 24306

aggatccacc ggccaccgaa gagctgagaa agcgcatctt tcccactctg tatgctatct 24366

ttcagcaaag ccgcgggcag caccctcagc gcgaactgaa aataaaaaac cgctccttcc 24426

gctcgctcac ccgcagctgt ctgtaccaca agagagaaga ccagctgcag cgcaccctgg 24486

acgacgccga agcactgttc agcaaatact gctcagcgtc tcttaaagac taaaagaccc 24546

gcgctttttc cccctcggcc gccaaaaccc acgtcatcgc cagcatgagc aaggagattc 24606

ccacccccta catgtggagc tatcagcccc agatgggcct ggccgcgggg gccgcccagg 24666

actactccag caagatgaac tggctcagcg ccggcccccа catgatctca cgagttaacg 24726

gcatccgagc ccaccgaaac cagattctct tagaacaggc ggcaatcacc gccacacccc 24786

ggcgccaact caacccgcct agttggcccg ccgcccaggt gtatcaggaa aatccccgcc 24846

cgaccacagt cctcctgcca cgcgacgcgg aggccgaagt cctcatgact aactctgggg 24906

tacaattagc gggcgggtcc aggtacgcca ggtacagagg tcgggccgct ccttactctc 24966

ccgggagtat aaagagggtg atcattcgag gccgaggtat ccagctcaac gacgagacgg 25026

tgagctcctc aaccggtctc agacctgacg gagtcttcca gctcggagga gcgggccgct 25086

cttccttcac cactcgccag gcctacctga ccctgcagag ctcttcctcg cagccgcgct 25146

ccgggggaat cggcactctc cagttcgtgg aagagttcgt tccctccgtc tacttcaacc 25206

ccttctccgg ctcgcctgga cgctacccgg acgccttcat tcccaacttt gacgcagtga 25266

gtgaatccgt ggacggctac gactgatgac agatggtgcg gccgtgagag ctcggctgcg 25326

acatctgcat cactgccgtc agcctcgctg ctacgctcgg gaggcgatcg tcttcagcta 25386

ctttgagctg ccggacgagc accctcaggg tccggctcac gggttgaaac tcgagatcga 25446

gaacgcgctc gagtctcgcc tcatcgacac cttcaccgcc cgacctctcc tggtagaaat 25506

ccaacggggg atcactacca tcaccctgtt ctgcatctgc cccacgcccg gattacatga 25566

agatctgtgt tgtcatcttt gcgctcagtt taataaaaac tgaacttttt gccgcacctt 25626

caacgccatc tgtgatttct acaacaaaaa gttcttctgg caaaggtaca caaactgtat 25686

tttattctaa ttctacctca tctatcgtgc tgaactgcgc ctgcactaac gaacttatcc 25746

agtggattgc aaacggtagt gtgtgcaagt acttttgggg gaacgatata gttagtagaa 25806

ataacagcct ttgcgagcac tgcaactcct ccacactaat cctttatccc ccatttgtta 25866

ctggatggta tatgtgcgtt ggctccggtt taaatcctag ttgctttcat aagtggtttc 25926

tacaaaaaga gacccttccc aacaattctg tttctttttt cgccctatcc tactgctgtt 25986

ctccctctgg ttactctttc aaacctctaa ttggtatttt agctttgata ctcataatct 26046

ttattaactt tataataatt aacaacttac agtaaacatg cttgttctac tgctcgccac 26106

atctttcgct ctctctcacg ccagaacaag tattgttggc gcaggttaca atgcaactct 26166

tcaatctgct tacatgccag attccgacca gataccccat attacgtggt acttacaaac 26226

ctccaaacct aattcttcat tttatgaagg aaacaaactc tgcgatgact ccgacaacag 26286

aacgcacaca tttccccacc cttcactaca attcgaatgc gtaaacaaaa gcttgaagct 26346

ttacaactta aagccttcag attctggctt gtaccatgct gtagttgaaa aaagtaattt 26406

agaagtccac agtgattaca ttgaattgac ggttgtggac ctgccacctc caaaatgtga 26466

ggtttcctcc tcttaccttg aagttcaagg cgtggatgcc tactgcctca tacacattaa 26526

ctgcagcaac tctaaatatc cagctagaat ttactataat ggacaggaaa gtaatctttt 26586

ttattattta acaacaagcg ctggtaacgg taaacagtta cctgactatt ttactgctgt 26646

tgttgaattt tccacctaca gagaaacgta tgccaagcgg ccttacaatt tctcataccc 26706

gtttaacgac ctttgcaatg aaatacaagc gctcgaaact ggaactgatt ttactccaat 26766

tttcattgct gccattgttg taagcttaat taccattatt gtcagcctag cattttactg 26826

cttttacaag cccaaaaacc ctaagtttga aaaacttaaa ctaaaacctg tcattcaaca 26886

agtgtgattt tgttttccag catggtagct gcatttctac ttctcctctg tctacccatc 26946

attttcgtct cttcaacttt cgccgcagtt tcccacctgg aaccagagtg cctaccgcct 27006

tttgacgtgt atctgattct caccttttgtt tgttgtatat ccatttgcag tatagcctgc 27066

ttttttataa caatctttca agccgccgac tatttttacg tgcgaattgc ttactttaga 27126

caccatcctg aatacagaaa tcaaaacgtt gcctccttac tttgtttggc atgattaagt 27186

tattgctgat acttaattat ttacccctaa tcaactgtaa ttgtccattc accaaaccct 27246

ggtcattcta cacctgttat gataaaatcc ccgacactcc tgttgcttgg ctttacgcag 27306

ccaccgccgc tttggtattt atatctactt gccttggagt aaaattgtat tttattttac 27366

acactgggtg gctacatccc agagaagatt tacctagata tcctcttgta aacgcttttc 27426

aattacagcc tctgcctcct cctgatcttc ttcctcgagc tccctctatt gtgagctact 27486

ttcaactcac cggtggagat gactgactct caggacatta atattagtgt ggaaagaata 27546

gctgctcagc gtcagcgaga aacgcgagtg ttggaatacc tggaactaca gcaacttaaa 27606

gagtcccact ggtgtgagaa aggagtgctg tgccatgtta agcaggcagc cctttcctac 27666

gatgtcagcg ttcagggaca tgaactgtct tacactttgc ctttgcagaa acaaaccttc 27726

tgcaccatga tgggctctac ctccatcaca atcacccaac aagccgggcc tgtagagggg 27786

gctatcctct gtcactgtca cgcacctgat tgcatgtcca aactaatcaa aactctctgt 27846

```
gctttaggtg atatttttaa ggtgtaaatc aataataaac ttaccttaaa tttgacaaca  27906

aatttctggt gacatcattc agcagcacca ctttaccctc ttcccagctc tcgtatggga  27966

tgcgatagtg ggtggcaaac ttcctccaaa ccctaaaaga aatattggta tccacttcct  28026

tgtcctcacc cacaattttc atcttttcat ag atg  aaa aga acc aga gtt  gat  28079
                                    Met  Lys Arg Thr Arg Val  Asp
                                    1435                 1440

gaa gac ttc aac  ccc gtc tac ccc tat  gac acc aca acc act  cct     28124
Glu Asp Phe Asn  Pro Val Tyr Pro Tyr  Asp Thr Thr Thr Thr  Pro
            1445                 1450                 1455

gca gtt ccc ttt  ata tca ccc ccc ttt  gta aac agc gat ggt  ctt     28169
Ala Val Pro Phe  Ile Ser Pro Pro Phe  Val Asn Ser Asp Gly  Leu
            1460                 1465                 1470

cag gaa aac ccc  cca ggt gtt tta agt  ctg cga ata gct aaa  ccc     28214
Gln Glu Asn Pro  Pro Gly Val Leu Ser  Leu Arg Ile Ala Lys  Pro
            1475                 1480                 1485

cta tat ttc gac  atg gag aga aaa cta  gcc ctt tca ctt gga  aga     28259
Leu Tyr Phe Asp  Met Glu Arg Lys Leu  Ala Leu Ser Leu Gly  Arg
            1490                 1495                 1500

ggg ttg aca att  acc gcc gcc gga caa  tta gaa agt acg cag  agc     28304
Gly Leu Thr Ile  Thr Ala Ala Gly Gln  Leu Glu Ser Thr Gln  Ser
            1505                 1510                 1515

gta caa acc aac  cca ccg ttg ata att  acc aac aac aac aca  ctg     28349
Val Gln Thr Asn  Pro Pro Leu Ile Ile  Thr Asn Asn Asn Thr  Leu
            1520                 1525                 1530

acc cta cgt cat  tct ccc ccc tta aac  cta act gac aat agc  tta     28394
Thr Leu Arg His  Ser Pro Pro Leu Asn  Leu Thr Asp Asn Ser  Leu
            1535                 1540                 1545

gtg cta ggc tac  tcg agt cct ctc cgc  gtc aca gac aac aaa  ctt     28439
Val Leu Gly Tyr  Ser Ser Pro Leu Arg  Val Thr Asp Asn Lys  Leu
            1550                 1555                 1560

aca ttt aac ttc  aca tca cca ctc cgt  tat gaa aat gaa aac  ctt     28484
Thr Phe Asn Phe  Thr Ser Pro Leu Arg  Tyr Glu Asn Glu Asn  Leu
            1565                 1570                 1575

act ttt aac tat  aca gag cct ctt aaa  ctt ata aat aac agc  ctt     28529
Thr Phe Asn Tyr  Thr Glu Pro Leu Lys  Leu Ile Asn Asn Ser  Leu
            1580                 1585                 1590

gcc att gac atc  aat tcc tca aaa ggc  ctt agt agc gtc gga  ggc     28574
Ala Ile Asp Ile  Asn Ser Ser Lys Gly  Leu Ser Ser Val Gly  Gly
            1595                 1600                 1605

tca cta gct gta  aac ctg agt tca gac  tta aag ttt gac agc  aac     28619
Ser Leu Ala Val  Asn Leu Ser Ser Asp  Leu Lys Phe Asp Ser  Asn
            1610                 1615                 1620
```

```
gga tcc ata gct  ttt ggc ata caa acc  ctg tgg acc gct ccg  acc      28664
Gly Ser Ile Ala  Phe Gly Ile Gln Thr  Leu Trp Thr Ala Pro  Thr
            1625                 1630                 1635

tcg act ggc aac  tgc acc gtc tac agc  gag ggc gat tcc cta  ctt      28709
Ser Thr Gly Asn  Cys Thr Val Tyr Ser  Glu Gly Asp Ser Leu  Leu
            1640                 1645                 1650

agt ctc tgt tta  acc aaa tgc gga gct  cac gtc tta gga agt  gta      28754
Ser Leu Cys Leu  Thr Lys Cys Gly Ala  His Val Leu Gly Ser  Val
            1655                 1660                 1665

agt tta acc ggt  tta aca gga acc ata  acc caa atg act gat  att      28799
Ser Leu Thr Gly  Leu Thr Gly Thr Ile  Thr Gln Met Thr Asp  Ile
            1670                 1675                 1680

tct gtc acc att  caa ttt aca ttt gac  aac aat ggt aag cta  cta      28844
Ser Val Thr Ile  Gln Phe Thr Phe Asp  Asn Asn Gly Lys Leu  Leu
            1685                 1690                 1695

agc tct cca ctt  ata aac aac gcc ttt  agt att cga cag aat  gac      28889
Ser Ser Pro Leu  Ile Asn Asn Ala Phe  Ser Ile Arg Gln Asn  Asp
            1700                 1705                 1710

agt acg gcc tca  aac cct acc tac aac  gcc ctg gcg ttt atg  cct      28934
Ser Thr Ala Ser  Asn Pro Thr Tyr Asn  Ala Leu Ala Phe Met  Pro
            1715                 1720                 1725

aac agt acc ata  tat gca aga ggg gga  ggt ggt gaa cca cga  aac      28979
Asn Ser Thr Ile  Tyr Ala Arg Gly Gly  Gly Gly Glu Pro Arg  Asn
            1730                 1735                 1740

aac tac tac gtc  caa acg tat ctt agg  gga aat gtt caa aaa  cca      29024
Asn Tyr Tyr Val  Gln Thr Tyr Leu Arg  Gly Asn Val Gln Lys  Pro
            1745                 1750                 1755

atc att ctt act  gta acc tac aac tca  gtc gcc aca gga tat  tcc      29069
Ile Ile Leu Thr  Val Thr Tyr Asn Ser  Val Ala Thr Gly Tyr  Ser
            1760                 1765                 1770

tta tct ttt aag  tgg act gct ctt gca  cgt gaa aag ttt gca  acc      29114
Leu Ser Phe Lys  Trp Thr Ala Leu Ala  Arg Glu Lys Phe Ala  Thr
            1775                 1780                 1785

cca aca acc tcg  ttt tgc tac att aca  gaa caa taa aaccgtgtac      29160
Pro Thr Thr Ser  Phe Cys Tyr Ile Thr  Glu Gln
            1790                 1795

cccaccgttt cgttttttc ag atg aaa cgg  gcg aga gtt gat gaa  gac      29209
                        Met Lys Arg  Ala Arg Val Asp Glu  Asp
                                1800                 1805

ttc aac cca gtg  tac cct tat gac ccc  cca cat gct cct gtt  atg      29254
Phe Asn Pro Val  Tyr Pro Tyr Asp Pro  Pro His Ala Pro Val  Met
            1810                 1815                 1820
```

```
ccc ttc att act  cca cct ttt acc tcc  tcg gat ggg ttg cag  gaa     29299
Pro Phe Ile Thr  Pro Pro Phe Thr Ser  Ser Asp Gly Leu Gln  Glu
            1825                 1830                 1835

aaa cca ctt gga  gtg tta agt tta aac  tac aga gat ccc att  act     29344
Lys Pro Leu Gly  Val Leu Ser Leu Asn  Tyr Arg Asp Pro Ile  Thr
            1840                 1845                 1850

acg caa aat gag  tct ctt aca att aaa  cta gga aac ggc ctc  act     29389
Thr Gln Asn Glu  Ser Leu Thr Ile Lys  Leu Gly Asn Gly Leu  Thr
            1855                 1860                 1865

cta gac aac cag  gga caa cta aca tca  acc gct ggc gaa gta  gaa     29434
Leu Asp Asn Gln  Gly Gln Leu Thr Ser  Thr Ala Gly Glu Val  Glu
            1870                 1875                 1880

cct cca ctc act  aac gct aac aac aaa  ctt gca ctg gtc tat  agc     29479
Pro Pro Leu Thr  Asn Ala Asn Asn Lys  Leu Ala Leu Val Tyr  Ser
            1885                 1890                 1895

gat cct tta gca  gta aag cgc aac agc  cta acc tta tcg cac  acc     29524
Asp Pro Leu Ala  Val Lys Arg Asn Ser  Leu Thr Leu Ser His  Thr
            1900                 1905                 1910

gct ccc ctt gtt  att gct gat aac tct  tta gca ttg caa gtt  tca     29569
Ala Pro Leu Val  Ile Ala Asp Asn Ser  Leu Ala Leu Gln Val  Ser
            1915                 1920                 1925

gag cct att ttt  ata aat gac aag gac  aaa cta gcc ctg caa  aca     29614
Glu Pro Ile Phe  Ile Asn Asp Lys Asp  Lys Leu Ala Leu Gln  Thr
            1930                 1935                 1940

gcc gcg ccc ctt  gta act aac gct ggc  acc ctt cgc tta caa  agc     29659
Ala Ala Pro Leu  Val Thr Asn Ala Gly  Thr Leu Arg Leu Gln  Ser
            1945                 1950                 1955

gcc gcc cct tta  ggc att gca gac caa  acc cta aaa ctc ctg  ttt     29704
Ala Ala Pro Leu  Gly Ile Ala Asp Gln  Thr Leu Lys Leu Leu  Phe
            1960                 1965                 1970

acc aac cct ttg  tac ttg cag aat aac  ttt ctc acg tta gcc  att     29749
Thr Asn Pro Leu  Tyr Leu Gln Asn Asn  Phe Leu Thr Leu Ala  Ile
            1975                 1980                 1985

gaa cga ccc ctt  gcc att acc aat act  gga aag ctg gct cta  cag     29794
Glu Arg Pro Leu  Ala Ile Thr Asn Thr  Gly Lys Leu Ala Leu  Gln
            1990                 1995                 2000

ctc tcc cca ccg  cta caa aca gca gac  aca ggc ttg act ttg  caa     29839
Leu Ser Pro Pro  Leu Gln Thr Ala Asp  Thr Gly Leu Thr Leu  Gln
            2005                 2010                 2015

acc aac gtg cca  tta act gta agc aac  ggg acc cta ggc tta  gcc     29884
Thr Asn Val Pro  Leu Thr Val Ser Asn  Gly Thr Leu Gly Leu  Ala
            2020                 2025                 2030

ata aag cgc cca  ctt att att cag gac  aac aac ttg ttt ttg  gac     29929
```

```
Ile Lys Arg Pro  Leu Ile Ile Gln Asp  Asn Asn Leu Phe Leu  Asp
            2035                 2040                 2045

ttc aga gct ccc  ctg cgt ctt ttc aac  agc gac cca gta cta  ggg     29974
Phe Arg Ala Pro  Leu Arg Leu Phe Asn  Ser Asp Pro Val Leu  Gly
            2050                 2055                 2060

ctt aac ttt tac  acc cct ctt gcg gta  cgc gat gag gcg ctc  act     30019
Leu Asn Phe Tyr  Thr Pro Leu Ala Val  Arg Asp Glu Ala Leu  Thr
            2065                 2070                 2075

gtt aac aca ggc  cgc ggc ctc aca gtg  agt tac gat ggt tta  att     30064
Val Asn Thr Gly  Arg Gly Leu Thr Val  Ser Tyr Asp Gly Leu  Ile
            2080                 2085                 2090

tta aat ctt ggt  aag gat ctt cgc ttt  gac aac aac acc gtt  tct     30109
Leu Asn Leu Gly  Lys Asp Leu Arg Phe  Asp Asn Asn Thr Val  Ser
            2095                 2100                 2105

gtc gct ctt agt  gct gct ttg cct tta  caa tac act gat cag  ctt     30154
Val Ala Leu Ser  Ala Ala Leu Pro Leu  Gln Tyr Thr Asp Gln  Leu
            2110                 2115                 2120

cgc ctt aac gtg  ggc gct ggg ctg cgt  tac aat cca gtg agt  aag     30199
Arg Leu Asn Val  Gly Ala Gly Leu Arg  Tyr Asn Pro Val Ser  Lys
            2125                 2130                 2135

aaa ttg gac gtg  aac ccc aat caa aac  aag ggt tta acc tgg  gaa     30244
Lys Leu Asp Val  Asn Pro Asn Gln Asn  Lys Gly Leu Thr Trp  Glu
            2140                 2145                 2150

aat gac tac ctc  att gta aag cta gga  aat gga tta ggt ttt  gat     30289
Asn Asp Tyr Leu  Ile Val Lys Leu Gly  Asn Gly Leu Gly Phe  Asp
            2155                 2160                 2165

ggc gat gga aac  ata gct gtt tct cct  caa gtt aca tcg cct  gac     30334
Gly Asp Gly Asn  Ile Ala Val Ser Pro  Gln Val Thr Ser Pro  Asp
            2170                 2175                 2180

acc tta tgg acc  act gcc gac cca tcc  ccc aat tgt tcc atc  tac     30379
Thr Leu Trp Thr  Thr Ala Asp Pro Ser  Pro Asn Cys Ser Ile  Tyr
            2185                 2190                 2195

act gat tta gat  gcc aaa atg tgg ctc  tcg ttg gta aaa caa  ggg     30424
Thr Asp Leu Asp  Ala Lys Met Trp Leu  Ser Leu Val Lys Gln  Gly
            2200                 2205                 2210

ggt gtg gtt cac  ggt tct gtt gct tta  aaa gca ttg aaa gga  acc     30469
Gly Val Val His  Gly Ser Val Ala Leu  Lys Ala Leu Lys Gly  Thr
            2215                 2220                 2225

cta ttg agt cct  acg gaa agc gcc att  gtt att ata cta cat  ttt     30514
Leu Leu Ser Pro  Thr Glu Ser Ala Ile  Val Ile Ile Leu His  Phe
            2230                 2235                 2240

gac aat tat gga  gtg cga att ctc aat  tat ccc act ttg ggc  act     30559
Asp Asn Tyr Gly  Val Arg Ile Leu Asn  Tyr Pro Thr Leu Gly  Thr
```

```
            2245                2250                2255

caa ggc acg ttg  gga aat aat gca act  tgg ggt tat agg cag  gga     30604
Gln Gly Thr Leu  Gly Asn Asn Ala Thr  Trp Gly Tyr Arg Gln  Gly
            2260                2265                2270

gaa tct gca gac  act aat gta ctc aat  gca cta gca ttt atg  ccc     30649
Glu Ser Ala Asp  Thr Asn Val Leu Asn  Ala Leu Ala Phe Met  Pro
            2275                2280                2285

agt tca aaa agg  tac cca aga ggg cgt  gga agc gaa gtt cag  aat     30694
Ser Ser Lys Arg  Tyr Pro Arg Gly Arg  Gly Ser Glu Val Gln  Asn
            2290                2295                2300

caa act gtg ggc  tac act tgt ata cag  ggt gac ttt tct atg  ccc     30739
Gln Thr Val Gly  Tyr Thr Cys Ile Gln  Gly Asp Phe Ser Met  Pro
            2305                2310                2315

gta ccg tac caa  ata cag tac aac tat  gga cca act ggc tac  tcc     30784
Val Pro Tyr Gln  Ile Gln Tyr Asn Tyr  Gly Pro Thr Gly Tyr  Ser
            2320                2325                2330

ttt aaa ttt att  tgg aga act gtt tca  aga caa cca ttt gac  atc     30829
Phe Lys Phe Ile  Trp Arg Thr Val Ser  Arg Gln Pro Phe Asp  Ile
            2335                2340                2345

cca tgc tgt ttt  ttc tct tac att acg  gaa gaa taa aacaactttt      30875
Pro Cys Cys Phe  Phe Ser Tyr Ile Thr  Glu Glu
            2350                2355

tctttttatt ttctttttat tttacacgca cagtaaggct tcctccaccc ttccatctca  30935

cagcatacac cagcctctcc cccttcatgg cagtaaactg ttgtgagtca gtccggtatt  30995

tgggagttaa gatccaaaca gtctctttgg tgatgaaaca tggatccgtg atggacacaa  31055

atccctggga caggttctcc aacgtttcgg taaaaaactg catgccgccc tacaaaacaa  31115

acaggttcag gctctccacg ggttatctcc ccgatcaaac tcagacagag taaaggtgcg  31175

atgatgttcc actaaaccac gcaggtggcg ctgtctgaac ctctcggtgc gactcctgtg  31235

aggctggtaa gaagttagat tgtccagcag cctcacagca tggatcatca gtctacgagt  31295

gcgtctggcg cagcagcgca tctgaatctc actgagattc cggcaagaat cgcacaccat  31355

cacaatcagg ttgttcatga tcccatagct gaacacgctc cagccaaagc tcattcgctc  31415

caacagcgcc accgcgtgtc cgtccaacct tactttaaca taaatcaggt gtctgccgcg  31475

tacaaacatg ctacccgcat acagaacctc ccggggcaaa cccctgttca ccacctgcct  31535

gtaccaggga aacctcacat ttatcaggga gccatagata gccattttaa accaattagc  31595

taacaccgcc ccaccagctc tacactgaag agaaccggga gagttacaat gacagtgaat  31655

aatccatctc tcataacccc taatggtctg atggaaatcc agatctaacg tggcacagca  31715
```

gatacacact ttcatataca ttttcatcac atgtttttcc caggccgtta aaatacaatc 31775

ccaatacacg ggccactcct gcagtacaat aaagctaata caagatggta tactcctcac 31835

ctcactaaca ttgtgcatgt tcatattttc acattctaag taccgagagt tctcctctac 31895

aacagcactg ccgcggtcct cacaaggtgg tagctggtga cgattgtaag gagccagtct 31955

gcagcgatac cgtctgtcgc gttgcatcgt agaccaggga ccgacgcact tcctcgtact 32015

tgtagtagca gaaccacgtc cgctgccagc acgtctccaa gtaacgccgg tccctgcgtc 32075

gctcacgctc cctcctcaac gcaaagtgca accactcttg taatccacac agatccctct 32135

cggcctccgg ggcgatgcac acctcaaacc tacagatgtc tcggtacagt tccaaacacg 32195

tagtgagggc gagttccaac caagacagac agcctgatct atcccgacac actggaggtg 32255

gaggaagaca cggaagaggc atgttattcc aagcgattca ccaacgggtc gaaatgaaga 32315

tcccgaagat gacaacggtc gcctccggag ccctgatgga atttaacagc cagatcaaac 32375

attatgcgat tttccaggct atcaatcgcg gcctccaaaa gagcctggac ccgcacttcc 32435

acaaacacca gcaaagcaaa agcgttatta tcaaactctt cgatcatcaa gctgcaggac 32495

tgtacaatgc ccaagtaatt ttcatttctc cactcgcgaa tgatgtcgcg gcaaatagtc 32555

tgaaggttca tgccgtgcat attaaaaagc tccgaaaggg cgccctctat agccatgcgt 32615

agacacacca tcatgactgc aagatatcgg gctcctgaga cacctgcagc agatttaaca 32675

gacccaggtc aggttgctct ccgcgatcgc gaatctccat ccgcaaagtc atttgcaaat 32735

aattaaatag atctgcgccg actaaatctg ttaactccgc gctaggaact aaatcaggtg 32795

tggctacgca gcacaaaagt tccagggatg gcgccaaact cactagaacc gctcccgagt 32855

agcaaaactg atgaatggga gtaacacagt gtaaaatgtt cagccaaaaa tcactaagct 32915

gctcctttaa aaagtccagt acttctatat tcagttcgtg caagtactga agcaactgtg 32975

cgggaatatg cacagcaaaa aaaatagggc ggctcagata catgttgacc taaaataaaa 33035

agaatcatta aactaaagaa gcctggcgaa cggtgggata tatgacacgc tccagcagca 33095

ggcaagcaac cggctgtccc cgggaaccgc ggtaaaattc atccgaatga ttaaaaagaa 33155

caacagagac ttcccaccat gtactcggtt ggatctcctg agcacagagc aataccccc 33215

tcacattcat atccgctaca gaaaaaaaac gtcccagata cccagcggga atatccaacg 33275

acagctgcaa agacagcaaa acaatccctc tgggagcaat cacaaaatcc tccggtgaaa 33335

aaagcacata catattagaa taaccctgtt gctggggcaa aaaggcccgt cgtcccagca 33395

aatgcacata aatatgttca tcagccattg ccccgtctta ccgcgtaaac agccacgaaa 33455

aaatcgagct aaaatccacc caacagccta tagctatata tacactccac ccaatgacgc 33515

taataccgca ccacccacga ccaaagttca cccacaccca caaaacccgc gaaaatccag 33575

cgccgtcagc acttccgcaa tttcagtctc acaacgtcac ttccgcgcgc cttttcactt 33635

tcccacacac gcccttcgcc cgcccgccct cgcgccaccc cgcgtcaccc cacgtcaccg 33695

cacgtcaccc cggccccgcc tcgctcctcc ccgctcatta tcatattggc acgtttccag 33755

aataaggtat attattgatg cagcaaaaca atccctctgg gagcaatcac aaaatcctcc 33815

ggtgaaaaaa gcacatacat attagaataa ccctgttgct ggggcaaaaa ggcccgtcgt 33875

cccagcaaat gcacataaat atgttcatca gccattgccc cgtcttaccg cgtaaacagc 33935

cacgaaaaaa tcgagctaaa atccacccaa cagcctatag ctatatatac actccaccca 33995

atgacgctaa taccgcacca cccacgacca aagttcaccc acacccacaa aacccgcgaa 34055

aatccagcgc cgtcagcact tccgcaattt cagtctcaca acgtcacttc cgcgcgcctt 34115

ttcactttcc cacacacgcc cttcgcccgc ccgccctcgc gccaccccgc gtcaccccac 34175

gtcaccgcac gtcaccccgg ccccgcctcg ctcctccccg ctcattatca tattggcacg 34235

tttccagaat aaggtatatt attgatgca 34264

<210> 25
<211> 503
<212> PRT
<213> simian adenovirus SV-1

<400> 25

Met Arg Arg Ala Val Arg Val Thr Pro Ala Ala Tyr Glu Gly Pro Pro
1 5 10 15

Pro Ser Tyr Glu Ser Val Met Gly Ser Ala Asn Val Pro Ala Thr Leu
20 25 30

Glu Ala Pro Tyr Val Pro Pro Arg Tyr Leu Gly Pro Thr Glu Gly Arg
35 40 45

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

Val Tyr Leu Val Asp Asn Lys Ser Ala Asp Ile Ala Ser Leu Asn Tyr
65 70 75 80

```
Gln Asn Asp His Ser Asn Phe Leu Thr Thr Val Val Gln Asn Asn Asp
                85                  90                  95

Phe Thr Pro Thr Glu Ala Gly Thr Gln Thr Ile Asn Phe Asp Glu Arg
            100                 105                 110

Ser Arg Trp Gly Gly Gln Leu Lys Thr Ile Leu His Thr Asn Met Pro
        115                 120                 125

Asn Ile Asn Glu Phe Met Ser Thr Asn Lys Phe Arg Ala Arg Leu Met
    130                 135                 140

Val Lys Lys Ala Glu Asn Gln Pro Pro Glu Tyr Glu Trp Phe Glu Phe
145                 150                 155                 160

Thr Ile Pro Glu Gly Asn Tyr Ser Glu Thr Met Thr Ile Asp Leu Met
                165                 170                 175

Asn Asn Ala Ile Val Asp Asn Tyr Leu Gln Val Gly Arg Gln Asn Gly
            180                 185                 190

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

Leu Gly Trp Asp Pro Val Thr Lys Leu Val Met Pro Gly Val Tyr Thr
    210                 215                 220

Asn Glu Ala Phe His Pro Asp Ile Val Leu Leu Pro Gly Cys Gly Val
225                 230                 235                 240

Asp Phe Thr Gln Ser Arg Leu Ser Asn Leu Leu Gly Ile Arg Lys Arg
                245                 250                 255

Arg Pro Phe Gln Glu Gly Phe Gln Ile Met Tyr Glu Asp Leu Glu Gly
            260                 265                 270

Gly Asn Ile Pro Gly Leu Leu Asp Val Pro Ala Tyr Glu Glu Ser Val
        275                 280                 285

Lys Gln Ala Glu Ala Gln Gly Arg Glu Ile Arg Gly Asp Thr Phe Ala
    290                 295                 300
```

```
Thr Glu Pro His Glu Leu Val Ile Lys Pro Leu Glu Gln Asp Ser Lys
305                 310                 315                 320

Lys Arg Ser Tyr Asn Ile Ile Ser Gly Thr Met Asn Thr Leu Tyr Arg
                325                 330                 335

Ser Trp Phe Leu Ala Tyr Asn Tyr Gly Asp Pro Glu Lys Gly Val Arg
            340                 345                 350

Ser Trp Thr Ile Leu Thr Thr Thr Asp Val Thr Cys Gly Ser Gln Gln
        355                 360                 365

Val Tyr Trp Ser Leu Pro Asp Met Met Gln Asp Pro Val Thr Phe Arg
    370                 375                 380

Pro Ser Thr Gln Val Ser Asn Phe Pro Val Val Gly Thr Glu Leu Leu
385                 390                 395                 400

Pro Val His Ala Lys Ser Phe Tyr Asn Glu Gln Ala Val Tyr Ser Gln
                405                 410                 415

Leu Ile Arg Gln Ser Thr Ala Leu Thr His Val Phe Asn Arg Phe Pro
            420                 425                 430

Glu Asn Gln Ile Leu Val Arg Pro Pro Ala Pro Thr Ile Thr Thr Val
        435                 440                 445

Ser Glu Asn Val Pro Ala Leu Thr Asp His Gly Thr Leu Pro Leu Arg
    450                 455                 460

Ser Ser Ile Ser Gly Val Gln Arg Val Thr Ile Thr Asp Ala Arg Arg
465                 470                 475                 480

Arg Thr Cys Pro Tyr Val Tyr Lys Ala Leu Gly Val Val Ala Pro Lys
                485                 490                 495

Val Leu Ser Ser Arg Thr Phe
            500
```

<210> 26
<211> 931
<212> PRT
<213> simian adenovirus SV-1

<400> 26

Met Ala Thr Pro Ser Met Met Pro Gln Trp Ser Tyr Met His Ile Ala
1 5 10 15

Gly Gln Asp Ala Ser Glu Tyr Leu Ser Pro Gly Leu Val Gln Phe Ala
20 25 30

Arg Ala Thr Asp Thr Tyr Phe Ser Leu Gly Asn Lys Phe Arg Asn Pro
35 40 45

Thr Val Ala Pro Thr His Asp Val Thr Thr Asp Arg Ser Gln Arg Leu
50 55 60

Thr Leu Arg Phe Val Pro Val Asp Arg Glu Asp Thr Ala Tyr Ser Tyr
65 70 75 80

Lys Val Arg Tyr Thr Leu Ala Val Gly Asp Asn Arg Val Leu Asp Met
85 90 95

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

Phe Lys Pro Tyr Ser Gly Thr Ala Tyr Asn Ser Leu Ala Pro Lys Gly
115 120 125

Ala Pro Asn Pro Ala Glu Trp Thr Asn Ser Asp Ser Lys Val Lys Val
130 135 140

Arg Ala Gln Ala Pro Phe Val Ser Ser Tyr Gly Ala Thr Ala Ile Thr
145 150 155 160

Lys Glu Gly Ile Gln Val Gly Val Thr Leu Thr Asp Ser Gly Ser Thr
165 170 175

Pro Gln Tyr Ala Asp Lys Thr Tyr Gln Pro Glu Pro Gln Ile Gly Glu
180 185 190

Leu Gln Trp Asn Ser Asp Val Gly Thr Asp Asp Lys Ile Ala Gly Arg
195 200 205

Val Leu Lys Lys Thr Thr Pro Met Phe Pro Cys Tyr Gly Ser Tyr Ala
210 215 220

```
Arg Pro Thr Asn Glu Lys Gly Gly Gln Ala Thr Pro Ser Ala Ser Gln
225                 230                 235                 240

Asp Val Gln Asn Pro Glu Leu Gln Phe Phe Ala Ser Thr Asn Val Ala
                245                 250                 255

Asn Thr Pro Lys Ala Val Leu Tyr Ala Glu Asp Val Ser Ile Glu Ala
            260                 265                 270

Pro Asp Thr His Leu Val Phe Lys Pro Thr Val Thr Glu Gly Ile Thr
        275                 280                 285

Ser Ser Glu Ala Leu Leu Thr Gln Gln Ala Ala Pro Asn Arg Pro Asn
    290                 295                 300

Tyr Ile Ala Phe Arg Asp Asn Phe Ile Gly Leu Met Tyr Tyr Asn Ser
305                 310                 315                 320

Thr Gly Asn Met Gly Val Leu Ala Gly Gln Ala Ser Gln Leu Asn Ala
                325                 330                 335

Val Val Asp Leu Gln Asp Arg Asn Thr Glu Leu Ser Tyr Gln Leu Met
            340                 345                 350

Leu Asp Ala Leu Gly Asp Arg Ser Arg Tyr Phe Ser Met Trp Asn Gln
        355                 360                 365

Ala Val Asp Ser Tyr Asp Pro Asp Val Arg Ile Ile Glu Asn His Gly
    370                 375                 380

Val Glu Asp Glu Leu Pro Asn Tyr Cys Phe Pro Leu Gly Gly Met Ala
385                 390                 395                 400

Val Thr Asp Thr Tyr Ser Pro Ile Lys Val Asn Gly Gly Gly Asn Gly
                405                 410                 415

Trp Glu Ala Asn Asn Gly Val Phe Thr Glu Arg Gly Val Glu Ile Gly
            420                 425                 430

Ser Gly Asn Met Phe Ala Met Glu Ile Asn Leu Gln Ala Asn Leu Trp
        435                 440                 445
```

```
Arg Ser Phe Leu Tyr Ser Asn Ile Gly Leu Tyr Leu Pro Asp Ser Leu
    450                 455                 460

Lys Ile Thr Pro Asp Asn Ile Thr Leu Pro Glu Asn Lys Asn Thr Tyr
465                 470                 475                 480

Gln Tyr Met Asn Gly Arg Val Thr Pro Pro Gly Leu Val Asp Thr Tyr
                485                 490                 495

Val Asn Val Gly Ala Arg Trp Ser Pro Asp Val Met Asp Ser Ile Asn
            500                 505                 510

Pro Phe Asn His His Arg Asn Ala Gly Leu Arg Tyr Arg Ser Met Leu
        515                 520                 525

Leu Gly Asn Gly Arg Tyr Val Pro Phe His Ile Gln Val Pro Gln Lys
    530                 535                 540

Phe Phe Ala Ile Lys Asn Leu Leu Leu Leu Pro Gly Ser Tyr Thr Tyr
545                 550                 555                 560

Glu Trp Asn Phe Arg Lys Asp Val Asn Met Ile Leu Gln Ser Ser Leu
                565                 570                 575

Gly Asn Asp Leu Arg Val Asp Gly Ala Ser Ile Arg Phe Asp Ser Ile
            580                 585                 590

Asn Leu Tyr Ala Asn Phe Phe Pro Met Ala His Asn Thr Ala Ser Thr
        595                 600                 605

Leu Glu Ala Met Leu Arg Asn Asp Thr Asn Asp Gln Ser Phe Asn Asp
    610                 615                 620

Tyr Leu Cys Ala Ala Asn Met Leu Tyr Pro Ile Pro Ala Asn Ala Thr
625                 630                 635                 640

Ser Val Pro Ile Ser Ile Pro Ser Arg Asn Trp Ala Ala Phe Arg Gly
                645                 650                 655

Trp Ser Phe Thr Arg Leu Lys Thr Lys Glu Thr Pro Ser Leu Gly Ser
            660                 665                 670

Gly Phe Asp Pro Tyr Phe Val Tyr Ser Gly Ser Ile Pro Tyr Leu Asp
```

675 680 685

Gly Thr Phe Tyr Leu Asn His Thr Phe Lys Lys Val Ser Ile Met Phe
690 695 700

Asp Ser Ser Val Ser Trp Pro Gly Asn Asp Arg Leu Leu Thr Pro Asn
705 710 715 720

Glu Phe Glu Ile Lys Arg Ser Val Asp Gly Glu Gly Tyr Asn Val Ala
725 730 735

Gln Ser Asn Met Thr Lys Asp Trp Phe Leu Ile Gln Met Leu Ser His
740 745 750

Tyr Asn Ile Gly Tyr Gln Gly Phe Tyr Val Pro Glu Asn Tyr Lys Asp
755 760 765

Arg Met Tyr Ser Phe Phe Arg Asn Phe Gln Pro Met Ser Arg Gln Ile
770 775 780

Val Asp Ser Thr Ala Tyr Thr Asn Tyr Gln Asp Val Lys Leu Pro Tyr
785 790 795 800

Gln His Asn Asn Ser Gly Phe Val Gly Tyr Met Gly Pro Thr Met Arg
805 810 815

Glu Gly Gln Ala Tyr Pro Ala Asn Tyr Pro Tyr Pro Leu Ile Gly Ala
820 825 830

Thr Ala Val Pro Ser Leu Thr Gln Lys Lys Phe Leu Cys Asp Arg Val
835 840 845

Met Trp Arg Ile Pro Phe Ser Ser Asn Phe Met Ser Met Gly Ser Leu
850 855 860

Thr Asp Leu Gly Gln Asn Met Leu Tyr Ala Asn Ser Ala His Ala Leu
865 870 875 880

Asp Met Thr Phe Glu Val Asp Pro Met Asp Glu Pro Thr Leu Leu Tyr
885 890 895

Val Leu Phe Glu Val Phe Asp Val Val Arg Ile His Gln Pro His Arg
900 905 910

Gly Val Ile Glu Ala Val Tyr Leu Arg Thr Pro Phe Ser Ala Gly Asn
915 920 925

Ala Thr Thr
930

<210> 27
<211> 363
<212> PRT
<213> simian adenovirus SV-1

<400> 27

Met Lys Arg Thr Arg Val Asp Glu Asp Phe Asn Pro Val Tyr Pro Tyr
1 5 10 15

Asp Thr Thr Thr Thr Pro Ala Val Pro Phe Ile Ser Pro Pro Phe Val
20 25 30

Asn Ser Asp Gly Leu Gln Glu Asn Pro Pro Gly Val Leu Ser Leu Arg
35 40 45

Ile Ala Lys Pro Leu Tyr Phe Asp Met Glu Arg Lys Leu Ala Leu Ser
50 55 60

Leu Gly Arg Gly Leu Thr Ile Thr Ala Ala Gly Gln Leu Glu Ser Thr
65 70 75 80

Gln Ser Val Gln Thr Asn Pro Pro Leu Ile Ile Thr Asn Asn Asn Thr
85 90 95

Leu Thr Leu Arg His Ser Pro Pro Leu Asn Leu Thr Asp Asn Ser Leu
100 105 110

Val Leu Gly Tyr Ser Ser Pro Leu Arg Val Thr Asp Asn Lys Leu Thr
115 120 125

Phe Asn Phe Thr Ser Pro Leu Arg Tyr Glu Asn Glu Asn Leu Thr Phe
130 135 140

Asn Tyr Thr Glu Pro Leu Lys Leu Ile Asn Asn Ser Leu Ala Ile Asp
145 150 155 160

```
Ile Asn Ser Ser Lys Gly Leu Ser Ser Val Gly Gly Ser Leu Ala Val
                165                 170                 175

Asn Leu Ser Ser Asp Leu Lys Phe Asp Ser Asn Gly Ser Ile Ala Phe
            180                 185                 190

Gly Ile Gln Thr Leu Trp Thr Ala Pro Thr Ser Thr Gly Asn Cys Thr
        195                 200                 205

Val Tyr Ser Glu Gly Asp Ser Leu Leu Ser Leu Cys Leu Thr Lys Cys
    210                 215                 220

Gly Ala His Val Leu Gly Ser Val Ser Leu Thr Gly Leu Thr Gly Thr
225                 230                 235                 240

Ile Thr Gln Met Thr Asp Ile Ser Val Thr Ile Gln Phe Thr Phe Asp
                245                 250                 255

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

Ile Arg Gln Asn Asp Ser Thr Ala Ser Asn Pro Thr Tyr Asn Ala Leu
        275                 280                 285

Ala Phe Met Pro Asn Ser Thr Ile Tyr Ala Arg Gly Gly Gly Gly Glu
    290                 295                 300

Pro Arg Asn Asn Tyr Tyr Val Gln Thr Tyr Leu Arg Gly Asn Val Gln
305                 310                 315                 320

Lys Pro Ile Ile Leu Thr Val Thr Tyr Asn Ser Val Ala Thr Gly Tyr
                325                 330                 335

Ser Leu Ser Phe Lys Trp Thr Ala Leu Ala Arg Glu Lys Phe Ala Thr
            340                 345                 350

Pro Thr Thr Ser Phe Cys Tyr Ile Thr Glu Gln
        355                 360
```

<210> 28
<211> 560
<212> PRT
<213> simian adenovirus SV-1

<400> 28

```
Met Lys Arg Ala Arg Val Asp Glu Asp Phe Asn Pro Val Tyr Pro Tyr
1               5                  10                  15

Asp Pro Pro His Ala Pro Val Met Pro Phe Ile Thr Pro Pro Phe Thr
            20                  25                  30

Ser Ser Asp Gly Leu Gln Glu Lys Pro Leu Gly Val Leu Ser Leu Asn
        35                  40                  45

Tyr Arg Asp Pro Ile Thr Thr Gln Asn Glu Ser Leu Thr Ile Lys Leu
    50                  55                  60

Gly Asn Gly Leu Thr Leu Asp Asn Gln Gly Gln Leu Thr Ser Thr Ala
65                  70                  75                  80

Gly Glu Val Glu Pro Pro Leu Thr Asn Ala Asn Asn Lys Leu Ala Leu
                85                  90                  95

Val Tyr Ser Asp Pro Leu Ala Val Lys Arg Asn Ser Leu Thr Leu Ser
            100                 105                 110

His Thr Ala Pro Leu Val Ile Ala Asp Asn Ser Leu Ala Leu Gln Val
        115                 120                 125

Ser Glu Pro Ile Phe Ile Asn Asp Lys Asp Lys Leu Ala Leu Gln Thr
    130                 135                 140

Ala Ala Pro Leu Val Thr Asn Ala Gly Thr Leu Arg Leu Gln Ser Ala
145                 150                 155                 160

Ala Pro Leu Gly Ile Ala Asp Gln Thr Leu Lys Leu Leu Phe Thr Asn
                165                 170                 175

Pro Leu Tyr Leu Gln Asn Asn Phe Leu Thr Leu Ala Ile Glu Arg Pro
            180                 185                 190

Leu Ala Ile Thr Asn Thr Gly Lys Leu Ala Leu Gln Leu Ser Pro Pro
        195                 200                 205

Leu Gln Thr Ala Asp Thr Gly Leu Thr Leu Gln Thr Asn Val Pro Leu
    210                 215                 220
```

```
Thr Val Ser Asn Gly Thr Leu Gly Leu Ala Ile Lys Arg Pro Leu Ile
225                 230                 235                 240

Ile Gln Asp Asn Asn Leu Phe Leu Asp Phe Arg Ala Pro Leu Arg Leu
                245                 250                 255

Phe Asn Ser Asp Pro Val Leu Gly Leu Asn Phe Tyr Thr Pro Leu Ala
            260                 265                 270

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

Ser Tyr Asp Gly Leu Ile Leu Asn Leu Gly Lys Asp Leu Arg Phe Asp
    290                 295                 300

Asn Asn Thr Val Ser Val Ala Leu Ser Ala Ala Leu Pro Leu Gln Tyr
305                 310                 315                 320

Thr Asp Gln Leu Arg Leu Asn Val Gly Ala Gly Leu Arg Tyr Asn Pro
                325                 330                 335

Val Ser Lys Lys Leu Asp Val Asn Pro Asn Gln Asn Lys Gly Leu Thr
            340                 345                 350

Trp Glu Asn Asp Tyr Leu Ile Val Lys Leu Gly Asn Gly Leu Gly Phe
        355                 360                 365

Asp Gly Asp Gly Asn Ile Ala Val Ser Pro Gln Val Thr Ser Pro Asp
    370                 375                 380

Thr Leu Trp Thr Thr Ala Asp Pro Ser Pro Asn Cys Ser Ile Tyr Thr
385                 390                 395                 400

Asp Leu Asp Ala Lys Met Trp Leu Ser Leu Val Lys Gln Gly Gly Val
                405                 410                 415

Val His Gly Ser Val Ala Leu Lys Ala Leu Lys Gly Thr Leu Leu Ser
            420                 425                 430

Pro Thr Glu Ser Ala Ile Val Ile Ile Leu His Phe Asp Asn Tyr Gly
        435                 440                 445
```

Val Arg Ile Leu Asn Tyr Pro Thr Leu Gly Thr Gln Gly Thr Leu Gly
450 455 460

Asn Asn Ala Thr Trp Gly Tyr Arg Gln Gly Glu Ser Ala Asp Thr Asn
465 470 475 480

Val Leu Asn Ala Leu Ala Phe Met Pro Ser Ser Lys Arg Tyr Pro Arg
485 490 495

Gly Arg Gly Ser Glu Val Gln Asn Gln Thr Val Gly Tyr Thr Cys Ile
500 505 510

Gln Gly Asp Phe Ser Met Pro Val Pro Tyr Gln Ile Gln Tyr Asn Tyr
515 520 525

Gly Pro Thr Gly Tyr Ser Phe Lys Phe Ile Trp Arg Thr Val Ser Arg
530 535 540

Gln Pro Phe Asp Ile Pro Cys Cys Phe Phe Ser Tyr Ile Thr Glu Glu
545 550 555 560

<210> 29
<211> 31044
<212> DNA
<213> simian adenovirus SV-25

<220>
<221> CDS
<222> (12284)..(13801)
<223> Penton

<220>
<221> CDS
<222> (16681)..(19446)
<223> Hexon

<220>
<221> CDS
<222> (25380)..(26423)
<223> Fiber #2

<220>
<221> CDS
<222> (26457)..(28136)
<223> Fiber #1

<400> 29

catcatcaat aatatacctt attctggaaa cgtgccaata tgataatgag cggggaggag 60

cgaggcgggg ccggggtgac gtgcggtgac gcggggtggc gcgagggcgg ggcgaagggc 120

gcgggtgtgt gtgtgggagg cgcttagttt ttacgtatgc ggaaggaggt tttataccgg 180

aagatgggta atttgggcgt atacttgtaa gttttgtgta atttggcgcg aaaactgggt 240

aatgaggaag ttgaggttaa tatgtacttt ttatgactgg gcggaatttc tgctgatcag 300

cagtgaactt tgggcgctga cggggaggtt tcgctacgtg acagtaccac gagaaggctc 360

aaaggtccca tttattgtac tcttcagcgt tttcgctggg tatttaaacg ctgtcagatc 420

atcaagaggc cactcttgag tgctggcgag aagagttttc tcctccgtgc tgccacgatg 480

aggctggtcc ccgagatgta cggtgttttt agcgacgaga cggtgcgtaa ctcagatgac 540

ctgctgaatt cagacgcgct ggaaatttcc aattcgcctg tgctttcgcc gccgtcactt 600

cacgacctgt ttgtgttttg gctcaacgct tagcaacgtg ttatataggg tcaagaagga 660

gcaggagacg cagtttgcta ggctgttggc cgatactcct ggagtttttg tggctctgga 720

tctaggccat cactctcttt tccaagagaa aattatcaaa aacttaactt ttacgtctcc 780

tggtcgcacg gttgcttccg ctgcctttat tacctatatt ttggatcaat ggagcaacag 840

cgacagccac ctgtcgtggg agtacatgct ggattacatg tcgatggcgc tgtggagggc 900

catgctgcgg aggagggttt gcatttactt gcgggcgcag cctccgcggc tggaccgagt 960

ggaggaggag gacgagccgg gggagaccga gaacctgagg gccgggctgg accctccaac 1020

ggaggactag gtgctgagga tgatcccgaa gaggggacta gtggggctag gaagaagcaa 1080

aagactgagt ctgaacctcg aaacttttg aatgagttga ctgtgagttt gatgaatcgt 1140

cagcgtccgg agacaatttt ctggtctgaa ttggaggagg aattcaggag gggggaactg 1200

aacctgctat acaagtatgg gtttgaacag ttaaaaactc actggttgga gccgtgggag 1260

gattttgaaa ccgccttgga cacttttgct aaagtggctc tgcggccgga taaggtttac 1320

actatccgcc gcactgttaa cataaagaag agtgtttatg ttataggcca tggagctctg 1380

gtgcaggtgc aaaccgtcga ccgggtggcc tttagttgcg gtatgcaaaa tctgggcccc 1440

ggggtgatag gcttaaatgg tgtaacattt cacaatgtaa ggtttactgg tgaaagtttt 1500

aacggctctg tgtttgcaaa taacacacag ctgacgctcc acggcgttta ctttttaac 1560

tttaataaca catgtgtgga gtcgtggggc agggtgtctt tgaggggctg ctgttttcac 1620

ggctgctgga aggcggtggt gggaagactt aaaagtgtaa catctgtaaa aaaatgcgtg 1680

```
tttgagcggt gtgtgttggc tttaactgtg gagggctgtg gacgcattag gaataatgcg   1740
gcgtctgaga atggatgttt tcttttgcta aaaggcacgg ctagtattaa gcataacatg   1800
atatgcggca gcggtctgta cccttcacag ctgttaactt gcgcggatgg aaactgtcag   1860
accttgcgca ccgtgcacat agcgtcccac cagcgccgcg cctggccaac attcgagcac   1920
aatatgctta tgcgttgtgc cgtccacttg ggccctaggc gaggcgtgtt tgtgccttac   1980
cagtgtaact ttagccatac caagatttta ctagaacctg ataccttctc tcgagtgtgt   2040
ttcaatgggg tgtttgacat gtcaatggaa ctgtttaaag tgataagata tgatgaatcc   2100
aagtctcgtt gtcgcccatg tgaatgcgga gctaatcatc tgaggttgta tcctgtaacc   2160
ctaaacgtta ccgaggagct gaggacggat caccacatgt tgtcctgcct gcgcaccgac   2220
tatgaatcca gcgacgagga gtgaggtgag gggcggagcc acaaagggta taaaggggcg   2280
tgaggggtgg gtgtgatgat tcaaaatgag cgggacgacg gacggcaacg cgtttgaggg   2340
tggagtgttc agcccttatc tgacatctcg tcttccttcc tgggcaggag tgcgtcagaa   2400
tgtagtgggc tccaccgtgg acggacgacc ggtcgccccc gcaaattccg ccaccctcac   2460
ctatgccacc gtgggatcat cgttggacac tgccgcggca gctgccgctt ctgctgccgc   2520
ttctactgct cgcggcatgg cggctgattt tggactgtat aaccaactgg ccactgcagc   2580
tgtggcgtct cggtctctgg ttcaagaaga tgccctgaat gtgatcctga ctcgcctgga   2640
gatcatgtca cgtcgcttgg acgaactggc tgcgcagata tcccaagcta accccgatac   2700
cacttcagaa tcctaaaata aagacaaaca aatatgttga aaagtaaaat ggctttattt   2760
gtttttttttg gctcggtagg ctcgggtcca cctgtctcgg tcgttaagaa ctttgtgtat   2820
gttttccaaa acacggtaca gatgggcttg gatgttcaag tacatgggca tgaggccatc   2880
tttggggtga agataggacc attgaagagc gtcatgctcc ggggtggtgt tgtaaattac   2940
ccagtcgtag cagggtttct gggcgtggaa ctggaagatg tcctttagga gtaggctgat   3000
ggccaagggc aggcccttag tgtaggtgtt tacaaagcgg ttaagctggg agggatgcat   3060
gcgggggggag atgatatgca tcttggcttg gatcttgagg ttagctatgt taccacccag   3120
gtctctgcgg gggttcatgt tatgaaggac caccagcacg gtgtagccgg tgcatttggg   3180
gaacttgtca tgcagtttgg aggggaaggc gtggaagaat ttagagaccc ccttgtggcc   3240
ccctaggttt tccatgcact catccataat gatggcaatg ggacccctgg cggccgcttt   3300
ggcaaacacg ttttgggggt tggaaacatc atagttttgc tctagagtga gctcatcata   3360
ggccatctta acaaagcggg gtaggagggt gcccgactgg gggatgatag ttccatctgg   3420
```

gcctggggcg tagttaccct cacagatctg catctcccag gccttaattt ccgaggggg 3480

tatcatgtcc acctgggggg caataaagaa cacggtttct ggcgggggat tgatgagctg 3540

ggtggaaagc aagttacgca gcagttgaga tttgccacag ccggtggggc cgtagatgac 3600

cccgatgacg ggttgcagct ggtagttgag agaggaacag ctgccgtcgg ggcgcaggag 3660

gggggctacc tcattcatca tgcttctaac atgtttattt tcactcacta agttttgcaa 3720

gagcctctcc ccacccaggg ataagagttc ttccaggctg ttgaagtgtt tcagcggttt 3780

taggccgtcg gccatgggca tcttttcgag cgactgacga agcaagtaca gtcggtccca 3840

gagctcggtg acgtgctcta tggaatctcg atccagcaga cttcttggtt gcgggggttg 3900

ggtcgacttt cgctgtaggg caccagccgg tgggcgtcca gggccgcgag ggttctgtcc 3960

ttccagggtc tcagcgtccg ggtgagggtg gtctcggtga cggtgaaggg atgagccccg 4020

ggctgggcgc ttgcgagggt gcgcttcagg ctcatcctgc tggtgctgaa gcggacgtcg 4080

tctccctgtg agtcggccag atagcaacga agcatgaggt cgtagctgag ggactcggcc 4140

gcgtgtccct tggcgcgcag ctttcccttg gaaacgtgct gacatttggt gcagtgcaga 4200

cattggaggg cgtagagttt gggggccagg aagaccgact cgggcgagta ggcgtcggct 4260

ccgcactgag cgcagacggt ctcgcactcc actagccacg tgagctcggg tttagcggga 4320

tcaaaaacca agttgcctcc attttttttg atgcgtttct taccttgcgt ttccatgagt 4380

ttgtggcccg cttccgtgac aaaaaggctg tcggtgtctc cgtagacaga cttgaggggg 4440

cgatcttcca aaggtgttcc gaggtcttcc gcgtacagga actgggacca ctccgagacg 4500

aaggctctgg tccaggctaa cacgaaggag gcaatctgcg aggggtatct gtcgttttca 4560

atgagggggt ccacctttc cagggtgtgc agacacaggt cgtcctcctc cgcgtccacg 4620

aaggtgattg gcttgtaagt gtaggtcacg tgatctgcac cccccaaagg ggtataaaag 4680

ggggcgtgcc caccctctcc gtcactttct tccgcatcgc tgtggaccag agccagctgt 4740

tcgggtgagt aggccctctc aaaagccggc atgatctcgg cgctcaagtt gtcagtttct 4800

acaaacgagg tggatttgat attcacgtgc cccgcggcga tgcttttgat ggtggagggg 4860

tccatctgat cagaaaacac gatctttttg ttgtcaagtt tggtggcgaa agacccgtag 4920

agggcgttgg aaagcaactt ggcgatggag cgcagggtct gatttttctc ccgatcggcc 4980

ctctccttgg cggcgatgtt gagttgcacg tactcccggg ccgcgcaccg ccactcgggg 5040

aacacggcgg tgcgctcgtc gggcaggatg cgcacgcgcc agccgcgatt gtgcagggtg 5100

atgaggtcca cgctggtagc cacctccccg cggaggggct cgttggtcca acacaatcgc 5160

ccccctttc tggagcagaa cggaggcagg ggatctagca agttggcggg cgggggtcg 5220

gcgtcgatgg tgaagatacc gggtagcagg atcttattaa aataatcgat ttcggtgtcc 5280

gtgtcttgca acgcgtcttc ccacttcttc accgccaggg ccctttcgta gggattcagg 5340

ggcggtcccc agggcatggg gtgggtcagg gccgaggcgt acatgccgca gatgtcatac 5400

acgtacaggg gttccctcaa caccccgatg taagtggggt aacagcgccc cccgcggatg 5460

ctggctcgca cgtagtcgta catctcgcgc gagggagcca tgaggccgtc tcccaagtgg 5520

gtcttgtggg gtttttcggc ccggtagagg atctgtctga agatggcgtg ggagttggaa 5580

gagatggtgg ggcgttggaa gacgttaaag ttggccccgg gtagtcccac ggagtcttgg 5640

atgaactggg cgtaggattc ccggagtttg tccaccaggg cggcggtcac cagcacgtcg 5700

agagcgcagt agtccaacgt ctcgcggacc aggttgtagg ccgtctcttg ttttttctcc 5760

cacagttcgc ggttgaggag gtattcctcg cggtctttcc agtactcttc ggcgggaaat 5820

ccttttcgt ccgctcggta agaacctaac atgtaaaatt cgttcaccgc tttgtatgga 5880

caacagcctt tttctaccgg cagggcgtac gcttgagcgg cctttctgag agaggtgtgg 5940

gtgagggcga aggtgtcccg caccatcact ttcaggtact gatgtttgaa gtccgtgtcg 6000

tcgcaggcgc cctgttccca cagcgtgaag tcggtgcgct ttttctgcct gggattgggg 6060

agggcgaagg tgacatcgtt aaagagtatt ttcccggcgc ggggcatgaa gttgcgagag 6120

atcctgaagg gcccgggcac gtccgagcgg ttgttgatga cctgcgccgc caggacgatc 6180

tcgtcgaagc cgttgatgtt gtgacccacg atgtaaagtt cgatgaagcg cggctgtccc 6240

ttgagggccg gcgctttttt caactcctcg taggtgagac agtccggcga ggagagaccc 6300

agctcagccc gggcccagtc ggagagttga ggattagccg caaggaagga gctccataga 6360

tccaaggcca ggagagtttg caagcggtcg cggaactcgc ggaacttttt ccccacggcc 6420

attttctccg gtgtcactac gtaaaaggtg ttggggcggt tgttccacac gtcccatcgg 6480

agctctaggg ccagctcgca ggcttggcga acgagggtct cctcgccaga gacgtgcatg 6540

accagcataa agggtaccaa ctgtttcccg aacgagccca tccatgtgta ggtttctacg 6600

tcgtaggtga caaagagccg ctgggtgcgc gcgtgggagc cgatcggaaa gaagctgatc 6660

tcctgccacc agctggagga atgggtgtta atgtggtgga agtagaagtc ccgccggcgc 6720

acagagcatt cgtgctgatg tttgtaaaag cgaccgcagt agtcgcagcg ctgcacgctc 6780

tgtatctcct gaacgagatg cgcttttcgc ccgcgcacca gaaaccggag ggggaagttg 6840

agacgggggg ctggtggggc gacatcccct tcgccttggc ggtgggagtc tgcgtctgcg 6900

tcctccttct ctgggtggac gacggtgggg acgacgacgc cccgggtgcc gcaagtccag 6960

atctccgcca cggaggggtg caggcgctgc aggaggggac gcagctgccc gctgtccagg 7020

gagtcgaggg aagtcgcgct gaggtcggcg ggaagcgttt gcaagttcac tttcagaaga 7080

ccggtaagag cgtgagccag gtgcagatgg tacttgattt ccaggggggt gttggatgaa 7140

gcgtccacgg cgtagaggag tccgtgtccg cgcggggcca ccaccgtgcc ccgaggaggt 7200

tttatctcac tcgtcgaggg cgagcgccgg ggggtagagg cggctctgcg ccgggggcca 7260

gcggaggcag aggcacgttt tcgtgaggat tcggcagcgg ttgatgacga gcccggagac 7320

tgctggcgtg ggcgacgacg cggcggttga ggtcctggat gtgccgtctc tgcgtgaaga 7380

ccaccggccc ccgggtcctg aacctaaaga gagttccaca gaatcaatgt ctgcatcgtt 7440

aacggcggcc tgcctgagga tctcctgcac gtcgcccgag ttgtcctgat aggcgatctc 7500

ggccatgaac tgttccactt cttcctcgcg gaggtcaccg tggcccgctc gctccacggt 7560

ggcggccagg tcgttggaga tgcggcgcat gagttgagag aaggcgttga ggccgttctc 7620

gttccacacg cggctgtaca ccacgtttcc gaaggagtcg cgcgctcgca tgaccacctg 7680

ggccacgttg agttccacgt ggcgggcgaa gacggcgtag tttctgaggc gctggaagag 7740

gtagttgagc gtggtggcga tgtgctcgca gacgaagaag tacataatcc agcgccgcag 7800

ggtcatctcg ttgatgtctc cgatggcttc gagacgctcc atggcctcgt agaagtcgac 7860

ggcgaagttg aaaaattggg agttgcgggc ggccaccgtg agttcttctt gcaggaggcg 7920

gatgagatcg gcgaccgtgt cgcgcacctc ctgttcgaaa gcgccccgag gcgcctctgc 7980

ttcttcctcc ggctcctcct cttccagggg ctcgggttcc tccggcagct ctgcgacggg 8040

gacggggcgg cgacgtcgtc gtctgaccgg caggcggtcc acgaagcgct cgatcatttc 8100

gccgcgccgg cgacgcatgg tctcggtgac ggcgcgtccg ttttcgcgag gtcgcagttc 8160

gaagacgccg ccgcgcagag cgcccccgtg cagggagggt aagtggttag ggccgtcggg 8220

cagggacacg gcgctgacga tgcattttat caattgctgc gtaggcactc cgtgcaggga 8280

tctgagaacg tcgaggtcga cgggatccga gaacttctct aggaaagcgt ctatccaatc 8340

gcaatcgcaa ggtaagctga gaacggtggg tcgctggggg gcgttcgcgg gcagttggga 8400

ggtgatgctg ctgatgatgt aattaaagta ggcggtcttc aggcggcgga tggtggcgag 8460

gaggaccacg tctttgggcc cggcctgttg aatgcgcagg cgctcggcca tgccccaggc 8520

ctcgctctga cagcgacgca ggtctttgta gaagtcttgc atcagtctct ccaccggaac 8580

ctctgcttct cccctgtctg ccatgcgagt cgagccgaac ccccgcaggg gctgcagcaa 8640

cgctaggtcg gccacgaccc tttcggccag cacggcctgt tgaatctgcg tgagggtggc 8700

ctggaagtcg tccaggtcca cgaagcggtg ataggccccc gtgttgatgg tgtaggtgca 8760

gttggccatg acggaccagt tgacgacttg catgccgggt tgggtgatct ccgtgtactt 8820

gaggcgcgag taggccctgg actcgaacac gtagtcgttg catgtgcgca ccagatactg 8880

gtagccgacc aggaagtgag gaggcggctc tcggtacagg ggccagccaa cggtggcggg 8940

ggcgccgggg gacaggtcgt ccagcatgag gcggtggtag tggtagatgt agcgggagag 9000

ccaggtgatg ccggccgagg tggttgcggc cctggtgaat tcgcggacgc ggttccagat 9060

gttgcgcagg ggaccaaagc gctccatggt gggcacgctc tgccccgtga ggcgggcgca 9120

atcttgtacg ctctagatgg aaaaaagaca gggcggtcat cgactccttt ccgtagcttg 9180

gggggtaaag tcgcaagggt gcggcggcgg ggaaccccgg ttcgagaccg gccggatccg 9240

ccgctcccga tgcgcctggc cccgcatcca cgacgtccgc gccgagaccc agccgcgacg 9300

ctccgcccca atacggaggg gagtcttttg gtgtttttttc gtagatgcat ccggtgctgc 9360

ggcagatgcg accccagacg cccactacca ccgccgtggc ggcagtaaac ctgagcggag 9420

gcggtgacag ggaggaggaa gagctggctt tagacctgga agagggagag gggctggccc 9480

ggctgggagc gccatcccca gagagacacc ctagggttca gctcgtgagg gacgccaggc 9540

aggcttttgt gccgaagcag aacctgttta gggaccgcag cggtcaggag gcggaggaga 9600

tgcgcgattg caggtttcgg gcgggcagag agctcagggc gggcttcgat cgggagcggc 9660

tcctgagggc ggaggatttc gagcccgacg agcgttctgg ggtgagcccg gcccgcgctc 9720

acgtatcggc ggccaacctg gtgagcgcgt acgagcagac ggtgaacgag gagcgcaact 9780

tccaaaagag ctttaacaat cacgtgagga ccctgatcgc gagggaggag gtgaccatcg 9840

ggctgatgca tctgtgggac ttcgtggagg cctacgtgca gaacccggct agcaaacccc 9900

tgacggccca gctgttcctg atcgtgcagc acagccgcga caacgagacg ttccgcgacg 9960

ccatgttgaa catcgcggag cccgagggtc gctggctctt ggatctgatt aacatcctgc 10020

agagcatcgt ggtgcaggag aggggcctga gtttagcgga caaggtggcg gccattaact 10080

attcgatgca gagcctgggg aagttctacg ctcgcaagat ctacaagagc ccttacgtgc 10140

ccatagacaa ggaggtgaag atagacagct tttacatgcg catggcgctg aaggtgctga 10200

cgctgagcga cgacctcggc gtgtaccgta acgacaagat ccacaaggcg gtgagcgcca 10260

gccgccggcg ggagctgagc gacagggagc tgatgcacag cctgcagagg gcgctggcgg 10320

gcgccgggga cgaggagcgc gaggcttact tcgacatggg agccgatctg cagtggcgtc 10380

ccagcgcgcg cgccttggag gcggcgggtt atcccgacga ggaggatcgg gacgatttgg 10440

aggaggcagg cgagtacgag gacgaagcct gaccgggcag gtgttgtttt agatgcagcg 10500

gccggcggac gggaccaccg cggatcccgc acttttggca tccatgcaga gtcaaccttc 10560

gggcgtgacc gcctccgatg actgggcggc ggccatggac cgcatcatgg cgctgaccac 10620

ccgcaacccc gaggctttta ggcagcaacc ccaggccaac cgtttttcgg ccatcttgga 10680

agcggtggtg ccgtcgcgca ccaacccgac gcacgagaaa gtcctgacta tcgtgaacgc 10740

cctggtagac agcaaggcca tccgccgtga cgaggcgggc ttgatttaca acgctctttt 10800

ggaacgcgtg gcgcgctaca acagcactaa cgtgcagacc aatctggacc gcctcaccac 10860

cgacgtgaag gaggcgctgg cgcagaagga gcggtttctg agggacagta atctgggctc 10920

tctggtggca ctgaacgcct tcctgagctc acagccggcc aacgtgcccc gcgggcagga 10980

ggattacgtg agcttcatca gcgctctgag actgctggtg tccgaggtgc cccagagcga 11040

ggtgtaccag tctgggccgg attacttttt ccagacgtcc cgacagggct tgcaaacggt 11100

gaacctgact caggccttta aaaacttgca aggcatgtgg ggggtcaagg ccccggtggg 11160

cgatcgcgcc actatctcca gtctgctgac ccccaacact cgcctgctgc tgctcttgat 11220

cgcaccgttt accaacagta gcactatcag ccgtgactcg tacctgggtc atctcatcac 11280

tctgtaccgc gaggccatcg gccaggctca gatcgacgag catacgtatc aggagattac 11340

taacgtgagc cgtgccctgg gtcaggaaga taccggcagc ctggaagcca cgttgaactt 11400

tttgctaacc aaccggaggc aaaaaatacc ctcccagttc acgttaagcg ccgaggagga 11460

gaggattctg cgatacgtgc agcagtccgt gagcctgtac ttgatgcgcg agggcgccac 11520

cgcttccacg gctttagaca tgacggctcg gaacatggaa ccgtcctttt actccgccca 11580

ccggccgttc attaaccgtc tgatggacta cttccatcgc gcggccgcca tgaacgggga 11640

gtacttcacc aatgccatcc tgaatccgca ttggatgccc ccgtccggct tctacaccgg 11700

ggagtttgac ctgcccgaag ccgacgacgg ctttctgtgg gacgacgtgt ccgatagcat 11760

tttcacgccg gctaatcgcc gattccagaa gaaggagggc ggagacgagc tccccctctc 11820

cagcgtggaa gcggcctcaa ggggagagag tccctttcca agtctgtctt ccgccagtag 11880

cggtcgggta acgcgtccac ggttgccggg ggagagcgac tacctgaacg acccccttgct 11940

```
gcgaccggct agaaagaaaa attttcccaa taacggggtg gaaagcttgg tggataaaat  12000

gaatcgttgg aagacgtacg cccaggagca gcgggagtgg gaggacagtc agccgcggcc  12060

gctggtaccg ccgcattggc gtcgccagag agaagacccg gacgactccg cagacgatag  12120

tagcgtgttg gacctgggag ggagcggagc caaccccttt gctcacttgc aacccaaggg  12180

gcgctcgagt cgcctgtatt aataaaaaag acgcggaaac ttaccagagc catggccaca  12240

gcgtgtgtgc tttcttcctc tctttcttcc tcggcgcggc aga atg aga aga gcg     12295
                                              Met Arg Arg Ala
                                               1

gtg aga gtc acg ccg gcg gcg tat gag ggc ccg ccc cct tct tac gaa     12343
Val Arg Val Thr Pro Ala Ala Tyr Glu Gly Pro Pro Pro Ser Tyr Glu
5                   10                  15                  20

agc gtg atg gga tca gcg aac gtg ccg gcc acg ctg gag gcg cct tac     12391
Ser Val Met Gly Ser Ala Asn Val Pro Ala Thr Leu Glu Ala Pro Tyr
                25                  30                  35

gtt cct ccc aga tac ctg gga cct acg gag ggc aga aac agc atc cgt     12439
Val Pro Pro Arg Tyr Leu Gly Pro Thr Glu Gly Arg Asn Ser Ile Arg
            40                  45                  50

tac tcc gag ctg gcg ccc ctg tac gat acc acc aag gtg tac ctg gtg     12487
Tyr Ser Glu Leu Ala Pro Leu Tyr Asp Thr Thr Lys Val Tyr Leu Val
        55                  60                  65

gac aac aag tcg gcg gac atc gcc tcc ctg aat tac caa aac gat cac     12535
Asp Asn Lys Ser Ala Asp Ile Ala Ser Leu Asn Tyr Gln Asn Asp His
    70                  75                  80

agt aac ttt ctg act acc gtg gtg cag aac aat gac ttc acc ccg acg     12583
Ser Asn Phe Leu Thr Thr Val Val Gln Asn Asn Asp Phe Thr Pro Thr
85                  90                  95                  100

gag gcg ggc acg cag acc att aac ttt gac gag cgt tcc cgc tgg ggc     12631
Glu Ala Gly Thr Gln Thr Ile Asn Phe Asp Glu Arg Ser Arg Trp Gly
                105                 110                 115

ggt cag ctg aaa acc atc ctg cac acc aac atg ccc aac atc aac gag     12679
Gly Gln Leu Lys Thr Ile Leu His Thr Asn Met Pro Asn Ile Asn Glu
            120                 125                 130

ttc atg tcc acc aac aag ttc agg gct aag ctg atg gta gaa aaa agt     12727
Phe Met Ser Thr Asn Lys Phe Arg Ala Lys Leu Met Val Glu Lys Ser
        135                 140                 145

aat gcg gaa act cgg cag ccc cga tac gag tgg ttc gag ttt acc att     12775
Asn Ala Glu Thr Arg Gln Pro Arg Tyr Glu Trp Phe Glu Phe Thr Ile
    150                 155                 160

cca gag ggc aac tat tcc gaa act atg act atc gat ctc atg aat aac     12823
Pro Glu Gly Asn Tyr Ser Glu Thr Met Thr Ile Asp Leu Met Asn Asn
165                 170                 175                 180
```

```
gcg atc gtg gac aat tac ctg caa gtg ggg aga cag aac ggg gtg ctg    12871
Ala Ile Val Asp Asn Tyr Leu Gln Val Gly Arg Gln Asn Gly Val Leu
                185                 190                 195

gaa agc gat atc ggc gtg aaa ttc gat acc aga aac ttc cga ctg ggg    12919
Glu Ser Asp Ile Gly Val Lys Phe Asp Thr Arg Asn Phe Arg Leu Gly
            200                 205                 210

tgg gat ccc gtg acc aag ctg gtg atg cca ggc gtg tac acc aac gag    12967
Trp Asp Pro Val Thr Lys Leu Val Met Pro Gly Val Tyr Thr Asn Glu
        215                 220                 225

gct ttt cac ccg gac atc gtg ctg ctg ccg ggg tgc ggt gtg gac ttc    13015
Ala Phe His Pro Asp Ile Val Leu Leu Pro Gly Cys Gly Val Asp Phe
    230                 235                 240

act cag agc cgt ttg agt aac ctg tta gga att aga aag cgc cgc ccc    13063
Thr Gln Ser Arg Leu Ser Asn Leu Leu Gly Ile Arg Lys Arg Arg Pro
245                 250                 255                 260

ttc caa gag ggc ttt caa atc atg tat gag gac ctg gag gga ggt aat    13111
Phe Gln Glu Gly Phe Gln Ile Met Tyr Glu Asp Leu Glu Gly Gly Asn
                265                 270                 275

ata ccc gcc tta ctg gac gtg tcg aag tac gaa gct agc ata caa cgc    13159
Ile Pro Ala Leu Leu Asp Val Ser Lys Tyr Glu Ala Ser Ile Gln Arg
            280                 285                 290

gcc aaa gcg gag ggt aga gag att cgg gga gac acc ttt gcg gta gct    13207
Ala Lys Ala Glu Gly Arg Glu Ile Arg Gly Asp Thr Phe Ala Val Ala
        295                 300                 305

ccc cag gac ctg gaa ata gtg cct tta act aaa gac agc aaa gac aga    13255
Pro Gln Asp Leu Glu Ile Val Pro Leu Thr Lys Asp Ser Lys Asp Arg
    310                 315                 320

agc tac aat att ata aac aac acg acg gac acc ctg tat cgg agc tgg    13303
Ser Tyr Asn Ile Ile Asn Asn Thr Thr Asp Thr Leu Tyr Arg Ser Trp
325                 330                 335                 340

ttt ctg gct tac aac tac gga gac ccc gag aaa gga gtg aga tca tgg    13351
Phe Leu Ala Tyr Asn Tyr Gly Asp Pro Glu Lys Gly Val Arg Ser Trp
                345                 350                 355

acc ata ctc acc acc acg gac gtg acc tgt ggc tcg cag caa gtg tac    13399
Thr Ile Leu Thr Thr Thr Asp Val Thr Cys Gly Ser Gln Gln Val Tyr
            360                 365                 370

tgg tcc ctg ccg gat atg atg caa gac ccg gtc acc ttc cgc ccc tcc    13447
Trp Ser Leu Pro Asp Met Met Gln Asp Pro Val Thr Phe Arg Pro Ser
        375                 380                 385

acc caa gtc agc aac ttc ccg gtg gtg ggc acc gag ctg ctg ccc gtc    13495
Thr Gln Val Ser Asn Phe Pro Val Val Gly Thr Glu Leu Leu Pro Val
    390                 395                 400
```

```
cat gcc aag agc ttc tac aac gag cag gcc gtc tac tcg caa ctt att    13543
His Ala Lys Ser Phe Tyr Asn Glu Gln Ala Val Tyr Ser Gln Leu Ile
405                 410                 415                 420

cgc cag tcc acc gcg ctt acc cac gtg ttc aat cgc ttt ccc gag aac    13591
Arg Gln Ser Thr Ala Leu Thr His Val Phe Asn Arg Phe Pro Glu Asn
                425                 430                 435

cag att ctg gtg cgc cct ccc gct cct acc att acc acc gtc agt gaa    13639
Gln Ile Leu Val Arg Pro Pro Ala Pro Thr Ile Thr Thr Val Ser Glu
            440                 445                 450

aac gtt ccc gcc ctc aca gat cac gga acc ctg ccg ctg cgc agc agt    13687
Asn Val Pro Ala Leu Thr Asp His Gly Thr Leu Pro Leu Arg Ser Ser
        455                 460                 465

atc agt gga gtt cag cgc gtg acc atc acc gac gcc aga cgt cga acc    13735
Ile Ser Gly Val Gln Arg Val Thr Ile Thr Asp Ala Arg Arg Arg Thr
    470                 475                 480

tgc ccc tac gtt tac aaa gcg ctt ggc gtg gtg gct cct aaa gtt ctt    13783
Cys Pro Tyr Val Tyr Lys Ala Leu Gly Val Val Ala Pro Lys Val Leu
485                 490                 495                 500

tct agt cgc acc ttc taa aaacatgtcc atcctcatct ctcccgataa           13831
Ser Ser Arg Thr Phe
                505

caacaccggc tggggactgg gctccggcaa gatgtacggc ggagccaaaa ggcgctccag  13891

tcagcaccca gttcgagttc ggggccactt ccgcgctcct tggggagctt acaagcgagg  13951

actctcgggt cgaacggctg tagacgatac catagatgcc gtgattgccg acgcccgccg  14011

gtacaacccc ggaccggtcg ctagcgccgc ctccaccgtg gattccgtga tcgacagcgt  14071

ggtagccggc gctcgggcct atgctcgccg caagaggcgg ctgcatcgga gacgtcgccc  14131

caccgccgcc atgctggcag ccagggccgt gctgaggcgg gcccggaggg caggcagaag  14191

ggctatgcgc cgcgctgccg ccaacgccgc cgccgggagg gcccgccgac aggctgcccg  14251

ccaggctgcc gctgccatcg ctagcatggc cagacccagg agagggaacg tgtactgggt  14311

gcgtgattct gtgacgggag tccgagtgcc ggtgcgcagc cgacctcccc gaagttagaa  14371

gatccaagct gcgaagacgg cggtactgag tctccctgtt gttatcagcc caacatgagc  14431

aagcgcaagt ttaaagaaga actgctgcag acgctggtgc ctgagatcta tggccctccg  14491

gacgtgaagc cagacattaa gccccgcgat atcaagcgtg ttaaaaagcg ggaaaagaaa  14551

gaggaactcg cggtggtaga cgatggcgga gtggaattta ttaggagttt cgccccgcga  14611

cgcagggttc aatggaaagg gcggcgggta caacgcgttt tgaggccggg caccgcggta  14671

gtttttaccc cgggagagcg gtcggccgtt aggggtttca aaaggcagta cgacgaggtg  14731
```

tacggcgacg aggacatatt ggaacaggcg gctcaacaga tcggagaatt tgcctacgga 14791

aagcgttcgc gtcgcgaaga cctggccatc gccttagaca gcggcaaccc cacgcccagc 14851

ctcaaacccg tgacgctgca gcaggtgctt cccgtgagcg ccagcacgga cagcaagagg 14911

gggattaaga gagaaatgga agatctgcat cccaccatcc aactcatggt ccctaaacgg 14971

cagaggctgg aagaggtcct ggagaagatg aaagtggacc ccagcataga gccggatgta 15031

aaagtcagac ctattaagga agtggccccc ggtcttgggg tgcaaacggt ggacattcaa 15091

atccccgtca ccaccgcttc aaccgccgtg gaagctatgg aaacgcaaac ggagacccct 15151

gccgcgatcg gtaccaggga agtggcgttg caaacggagc cttggtacga atacgcagcc 15211

cctcggcgtc agaggcgttc cgctcgttac ggccccgcca acgccatcat gccagaatat 15271

gcgctgcatc cgtctattct gcccactccc ggataccggg gtgtgacgta tcgcccgtct 15331

ggaacccgcc gccgaacccg tcgccgccgc cgctcccgtc gcgctctggc ccccgtgtcg 15391

gtgcggcgtg tgacccgccg gggaaagaca gtcgtcattc ccaacccgcg ttaccaccct 15451

agcatccttt aataactctg ccgttttgca gatggctctg acttgccgcg tgcgccttcc 15511

cgttccgcac tatcgaggaa gatctcgtcg taggagaggc atgacgggca gtggtcgccg 15571

gcgggctttg cgcaggcgca tgaaaggcgg aattttaccc gccctgatac ccataattgc 15631

cgccgccatc ggtgccatac ccggcgttgc ttcagtggcg ttgcaagcag ctcgtaataa 15691

ataaacaaag gcttttgcac ttatgacctg gtcctgacta ttttatgcag aaagagcatg 15751

gaagacatca attttacgtc gctggctccg cggcacggct cgcggccgct catgggcacc 15811

tggaacgaca tcggcaccag tcagctcaac gggggcgctt tcaattgggg gagcctttgg 15871

agcggcatta aaaactttgg ctccacgatt aaatcctacg gcagcaaagc ctggaacagt 15931

agtgctggtc agatgctccg agataaactg aaggacacca acttccaaga aaaagtggtc 15991

aatggggtgg tgaccggcat ccacggcgcg gtagatctcg ccaaccaagc ggtgcagaaa 16051

gagattgaca ggcgtttgga aagctcgcgg gtgccgccgc agagagggga tgaggtggag 16111

gtcgaggaag tagaagtaga ggaaaagctg cccccgctgg agaaagttcc cggtgcgcct 16171

ccgagaccgc agaagcggcc caggccagaa ctagaagaga ctctggtgac ggagagcaag 16231

gagcctccct cgtacgagca agccttgaaa gagggcgcct ctccaccctc ctacccgatg 16291

actaagccga tcgcacccat ggctcgaccg gtgtacggca aggattacaa gcccgtcacg 16351

ctagagctgc ccccaccgcc ccccacgcgc ccgaccgtcc ccccccctgcc gactccgtcg 16411

```
gcggccgcgg cgggacccgt gtccgcacca tccgctgtgc ctctgccagc cgcccgtcca  16471

gtggccgtgg ccactgccag aaaccccaga ggccagagag gagccaactg gcaaagcacg  16531

ctgaacagca tcgtgggcct gggagtgaaa agcctgaaac gccgccgttg ctattattaa  16591

aaaagtgtag ctaaaaagtc tcccgttgta tacgcctcct atgttaccgc cagagacgag  16651

tgactgtcgc cgcgagcgcc gctttcaag atg gcc acc cca tcg atg atg ccg     16704
                               Met Ala Thr Pro Ser Met Met Pro
                                                   510

cag tgg tct tac atg cac atc gcc ggc cag gac gcc tcg gag tac ctg    16752
Gln Trp Ser Tyr Met His Ile Ala Gly Gln Asp Ala Ser Glu Tyr Leu
    515                 520                 525

agt ccc ggc ctc gtg cag ttt gcc cgc gcc acc gac acc tac ttc agc    16800
Ser Pro Gly Leu Val Gln Phe Ala Arg Ala Thr Asp Thr Tyr Phe Ser
530                 535                 540                 545

ttg gga aac aag ttt aga aac ccc acc gtg gcc ccc acc cac gat gtg    16848
Leu Gly Asn Lys Phe Arg Asn Pro Thr Val Ala Pro Thr His Asp Val
                550                 555                 560

acc acg gac cgc tcg cag agg ctg acc ctg cgc ttt gtg ccc gta gac    16896
Thr Thr Asp Arg Ser Gln Arg Leu Thr Leu Arg Phe Val Pro Val Asp
            565                 570                 575

cgg gag gac acc gcg tac tct tac aaa gtg cgc tac acg ttg gcc gta    16944
Arg Glu Asp Thr Ala Tyr Ser Tyr Lys Val Arg Tyr Thr Leu Ala Val
        580                 585                 590

ggg gac aac cga gtg ctg gac atg gcc agc acc tac ttt gac atc cgg    16992
Gly Asp Asn Arg Val Leu Asp Met Ala Ser Thr Tyr Phe Asp Ile Arg
    595                 600                 605

ggg gtg ctg gat cgg ggt ccc agc ttc aag ccc tat tcc ggc acc gct    17040
Gly Val Leu Asp Arg Gly Pro Ser Phe Lys Pro Tyr Ser Gly Thr Ala
610                 615                 620                 625

tac aac tcc ctg gcc ccc aag gga gct ccc aac ccc tcg gaa tgg acg    17088
Tyr Asn Ser Leu Ala Pro Lys Gly Ala Pro Asn Pro Ser Glu Trp Thr
                630                 635                 640

gac act tcc gac aac aaa ctt aaa gca tat gct cag gct ccc tac cag    17136
Asp Thr Ser Asp Asn Lys Leu Lys Ala Tyr Ala Gln Ala Pro Tyr Gln
            645                 650                 655

agt caa gga ctt aca aag gat ggt att cag gtt ggg cta gtt gtg aca    17184
Ser Gln Gly Leu Thr Lys Asp Gly Ile Gln Val Gly Leu Val Val Thr
        660                 665                 670

gag tca gga caa aca ccc caa tat gca aac aaa gtg tac caa ccc gag    17232
Glu Ser Gly Gln Thr Pro Gln Tyr Ala Asn Lys Val Tyr Gln Pro Glu
    675                 680                 685

cca caa att ggg gaa aac caa tgg aat tta gaa caa gaa gat aaa gcg    17280
```

```
Pro Gln Ile Gly Glu Asn Gln Trp Asn Leu Glu Gln Glu Asp Lys Ala
690                 695                 700                 705

gcg gga aga gtc cta aag aaa gat acc cct atg ttt ccc tgc tat ggg   17328
Ala Gly Arg Val Leu Lys Lys Asp Thr Pro Met Phe Pro Cys Tyr Gly
                710                 715                 720

tca tat gcc agg ccc aca aac gaa caa gga ggg cag gca aaa aac caa   17376
Ser Tyr Ala Arg Pro Thr Asn Glu Gln Gly Gly Gln Ala Lys Asn Gln
            725                 730                 735

gaa gta gat tta cag ttt ttt gcc act ccg ggc gac acc cag aac acg   17424
Glu Val Asp Leu Gln Phe Phe Ala Thr Pro Gly Asp Thr Gln Asn Thr
        740                 745                 750

gct aaa gtg gta ctt tat gct gaa aat gtc aac ctg gaa act cca gat   17472
Ala Lys Val Val Leu Tyr Ala Glu Asn Val Asn Leu Glu Thr Pro Asp
    755                 760                 765

act cac tta gtg ttt aaa ccc gat gac gac agc acc agt tca aaa ctt   17520
Thr His Leu Val Phe Lys Pro Asp Asp Asp Ser Thr Ser Ser Lys Leu
770                 775                 780                 785

ctt ctt ggg cag cag gct gca cct aac aga ccc aac tac ata ggt ttt   17568
Leu Leu Gly Gln Gln Ala Ala Pro Asn Arg Pro Asn Tyr Ile Gly Phe
                790                 795                 800

aga gat aat ttt att ggt tta atg tac tac aat agc act gga aac atg   17616
Arg Asp Asn Phe Ile Gly Leu Met Tyr Tyr Asn Ser Thr Gly Asn Met
            805                 810                 815

ggc gtg ctg gcc gga cag gct tct caa ttg aat gcc gta gtc gac ttg   17664
Gly Val Leu Ala Gly Gln Ala Ser Gln Leu Asn Ala Val Val Asp Leu
        820                 825                 830

cag gac aga aac acc gag ttg tcc tac cag ctg atg ctg gac gca ctg   17712
Gln Asp Arg Asn Thr Glu Leu Ser Tyr Gln Leu Met Leu Asp Ala Leu
    835                 840                 845

ggg gat cgc agc cga tat ttt tca atg tgg aat cag gca gta gac agc   17760
Gly Asp Arg Ser Arg Tyr Phe Ser Met Trp Asn Gln Ala Val Asp Ser
850                 855                 860                 865

tat gac cca gac gtt aga att ata gaa aac cac gga gtg gaa gac gaa   17808
Tyr Asp Pro Asp Val Arg Ile Ile Glu Asn His Gly Val Glu Asp Glu
                870                 875                 880

ctg cca aac tat tgt ttt cct ctg gga gga atg gtg gtg act gac aat   17856
Leu Pro Asn Tyr Cys Phe Pro Leu Gly Gly Met Val Val Thr Asp Asn
            885                 890                 895

tac aac tct gtg acg cct caa aat gga ggc agt gga aat aca tgg cag   17904
Tyr Asn Ser Val Thr Pro Gln Asn Gly Gly Ser Gly Asn Thr Trp Gln
        900                 905                 910

gca gac aat act aca ttt agt caa aga gga gcg cag att ggc tcc gga   17952
Ala Asp Asn Thr Thr Phe Ser Gln Arg Gly Ala Gln Ile Gly Ser Gly
```

```
    915                 920                 925

aac atg ttt gcc ctg gaa att aac cta cag gcc aac ctc tgg cgc ggc    18000
Asn Met Phe Ala Leu Glu Ile Asn Leu Gln Ala Asn Leu Trp Arg Gly
930                 935                 940                 945

ttc ttg tat tcc aat att ggg ttg tat ctt cca gac tct ctg aaa atc    18048
Phe Leu Tyr Ser Asn Ile Gly Leu Tyr Leu Pro Asp Ser Leu Lys Ile
                950                 955                 960

acc ccc gac aac atc acg ctg cca gaa aac aaa aac act tat cag tac    18096
Thr Pro Asp Asn Ile Thr Leu Pro Glu Asn Lys Asn Thr Tyr Gln Tyr
            965                 970                 975

atg aac ggt cgc gta acg cca ccc ggg ctc ata gac acc tat gta aac    18144
Met Asn Gly Arg Val Thr Pro Pro Gly Leu Ile Asp Thr Tyr Val Asn
        980                 985                 990

gtg ggc gcg cgc tgg tcc ccc  gat gtc atg gac agc  att aac ccc ttc  18192
Val Gly Ala Arg Trp Ser Pro  Asp Val Met Asp Ser  Ile Asn Pro Phe
    995                 1000                1005

aac  cac cac cgt aac gcg  ggc ttg cgc tac cgc  tcc atg ctc ttg     18237
Asn  His His Arg Asn Ala  Gly Leu Arg Tyr Arg  Ser Met Leu Leu
1010                1015                1020

ggc  aac ggc cgt tat gtg  cct ttt cac att cag  gtg ccc caa aaa     18282
Gly  Asn Gly Arg Tyr Val  Pro Phe His Ile Gln  Val Pro Gln Lys
1025                1030                1035

ttc  ttt gcc att aaa aac  ctg ctg ctt ctc ccc  ggt tcc tat acc     18327
Phe  Phe Ala Ile Lys Asn  Leu Leu Leu Leu Pro  Gly Ser Tyr Thr
1040                1045                1050

tat  gag tgg aac ttc cgc  aag gat gtc aac atg  atc ctg cag agc     18372
Tyr  Glu Trp Asn Phe Arg  Lys Asp Val Asn Met  Ile Leu Gln Ser
1055                1060                1065

tcg  ctg ggt aat gac ctg  cga gtg gac ggg gcc  agc ata cgc ttt     18417
Ser  Leu Gly Asn Asp Leu  Arg Val Asp Gly Ala  Ser Ile Arg Phe
1070                1075                1080

gac  agc att aac ctg tat  gcc aac ttt ttt ccc  atg gcc cac aac     18462
Asp  Ser Ile Asn Leu Tyr  Ala Asn Phe Phe Pro  Met Ala His Asn
1085                1090                1095

acg  gcc tct acc ctg gaa  gcc atg ctg cgc aac  gac acc aat gac     18507
Thr  Ala Ser Thr Leu Glu  Ala Met Leu Arg Asn  Asp Thr Asn Asp
1100                1105                1110

cag  tcc ttc aac gac tac  ctg tgc gcg gct aac  atg ctg tac ccc     18552
Gln  Ser Phe Asn Asp Tyr  Leu Cys Ala Ala Asn  Met Leu Tyr Pro
1115                1120                1125

atc  ccc gcc aac gcc acc  agc gtg ccc att tct  att cct tct cgg     18597
Ile  Pro Ala Asn Ala Thr  Ser Val Pro Ile Ser  Ile Pro Ser Arg
1130                1135                1140
```

```
aac  tgg gct gcc ttc agg  ggc tgg agt ttt act  cgc ctc aaa acc     18642
Asn  Trp Ala Ala Phe Arg  Gly Trp Ser Phe Thr  Arg Leu Lys Thr
1145                1150                1155

aag  gag act ccc tcg ctg  ggc tcc ggt ttt gac  ccc tac ttt gtt     18687
Lys  Glu Thr Pro Ser Leu  Gly Ser Gly Phe Asp  Pro Tyr Phe Val
1160                1165                1170

tac  tcc ggc tcc att ccc  tac cta gat ggc acc  ttt tac ctc aac     18732
Tyr  Ser Gly Ser Ile Pro  Tyr Leu Asp Gly Thr  Phe Tyr Leu Asn
1175                1180                1185

cac  act ttc aaa aag gtg  tct att atg ttt gac  tcc tcg gtt agc     18777
His  Thr Phe Lys Lys Val  Ser Ile Met Phe Asp  Ser Ser Val Ser
1190                1195                1200

tgg  ccc ggc aac gac cgc  ctg cta acg ccc aac  gag ttc gaa att     18822
Trp  Pro Gly Asn Asp Arg  Leu Leu Thr Pro Asn  Glu Phe Glu Ile
1205                1210                1215

aag  cgt tcc gtg gac ggt  gaa ggg tac aac gtg  gcc cag agc aac     18867
Lys  Arg Ser Val Asp Gly  Glu Gly Tyr Asn Val  Ala Gln Ser Asn
1220                1225                1230

atg  acc aag gac tgg ttt  cta att caa atg ctc  agt cac tat aat     18912
Met  Thr Lys Asp Trp Phe  Leu Ile Gln Met Leu  Ser His Tyr Asn
1235                1240                1245

ata  ggt tac cag ggc ttc  tat gtg ccc gag aac  tac aag gac cgc     18957
Ile  Gly Tyr Gln Gly Phe  Tyr Val Pro Glu Asn  Tyr Lys Asp Arg
1250                1255                1260

atg  tac tcc ttc ttc cgc  aac ttc caa cca atg  agc cgg cag gtg     19002
Met  Tyr Ser Phe Phe Arg  Asn Phe Gln Pro Met  Ser Arg Gln Val
1265                1270                1275

gta  gat acc gtg act tat  aca gac tac aaa gat  gtc aag ctc ccc     19047
Val  Asp Thr Val Thr Tyr  Thr Asp Tyr Lys Asp  Val Lys Leu Pro
1280                1285                1290

tac  caa cac aac aac tca  ggg ttc gtg ggc tac  atg gga ccc acc     19092
Tyr  Gln His Asn Asn Ser  Gly Phe Val Gly Tyr  Met Gly Pro Thr
1295                1300                1305

atg  cga gag gga cag gcc  tac ccg gcc aac tat  ccc tac ccc ctg     19137
Met  Arg Glu Gly Gln Ala  Tyr Pro Ala Asn Tyr  Pro Tyr Pro Leu
1310                1315                1320

atc  gga gag act gcc gta  ccc agc ctc acg cag  aaa aag ttc ctc     19182
Ile  Gly Glu Thr Ala Val  Pro Ser Leu Thr Gln  Lys Lys Phe Leu
1325                1330                1335

tgc  gac cgg gtg atg tgg  agg ata ccc ttc tct  agc aac ttt atg     19227
Cys  Asp Arg Val Met Trp  Arg Ile Pro Phe Ser  Ser Asn Phe Met
1340                1345                1350
```

```
tcg  atg ggc tcc ctc acc  gac ctg ggg cag aac  atg ctg tac gcc     19272
Ser  Met Gly Ser Leu Thr  Asp Leu Gly Gln Asn  Met Leu Tyr Ala
1355                 1360                 1365

aac  tcc gct cac gcc ttg  gac atg act ttt gag  gtg gat ccc atg     19317
Asn  Ser Ala His Ala Leu  Asp Met Thr Phe Glu  Val Asp Pro Met
1370                 1375                 1380

gat  gag ccc acg ctt ctc  tat gtt ctg ttt gaa  gtc ttc gac gtg     19362
Asp  Glu Pro Thr Leu Leu  Tyr Val Leu Phe Glu  Val Phe Asp Val
1385                 1390                 1395

gtg  cgc atc cac cag ccg  cac cgc ggc gtc atc  gag gcc gtc tac     19407
Val  Arg Ile His Gln Pro  His Arg Gly Val Ile  Glu Ala Val Tyr
1400                 1405                 1410

ctg  cgc aca cct ttc tct  gcc ggt aac gcc acc  acc taa agaagctgat  19456
Leu  Arg Thr Pro Phe Ser  Ala Gly Asn Ala Thr  Thr
1415                 1420                 1425

gggttccagc gaacaggagt tgcaggccat tgttcgcgac ctgggctgcg ggccctgctt  19516

tttgggcacc ttcgacaagc gttttcccgg attcatgtcc ccccacaagc cggcctgcgc  19576

catcgttaac acggccggac gggagacagg gggggtgcac tggctcgcct tcgcctggaa  19636

cccgcgcaac cgcacctgct acctgttcga ccctttttggt ttctccgacg aaaggctgaa  19696

gcagatctac caattcgagt acgaggggct cctcaagcgc agcgctctgg cctccacgcc  19756

cgaccactgc gtcaccctgg aaaagtccac ccagacggtc caggggcccc tctcggccgc  19816

ctgcgggctt ttctgttgca tgtttttgca cgccttcgtg cactggcctc acacccccat  19876

ggagcgcaac cccaccatgg atctgctcac cggagtgccc aacagcatgc ttcacagtcc  19936

ccaggtcgcc cccaccctgc gtcgcaatca ggaccacctg tatcgctttc tggggaaaca  19996

ctctgcctat ttccgccgcc accggcagcg catcgaacag gccacggcct tcgaaagcat  20056

gagccaaaga gtgtaatcaa taaaaaccgt ttttatttga catgatacgc gcttctggcg  20116

tttttattaa aaatcgaagg gttcgaggga ggggtcctcg tgcccgctgg ggagggacac  20176

gttgcggtac tggaatcggg cgctccaacg aaactcgggg atcaccagcc gcggcagggc  20236

cacgtcttcc atgttctgct tccaaaactg tcgcaccagc tgcagggctc ccatcacgtc  20296

gggcgctgag atcttgaagt cgcagttagg gccggagccc ccgcggctgt tgcggaacac  20356

ggggttggca cactggaaca ccaacacgct ggggttgtgg atactagcca gggccgtcgg  20416

gtcggtcacc tccgatgcat ccagatcctc ggcattgctc agggcgaacg gggtcagctt  20476

gcacatctgc cgcccgatct ggggtaccag gtcgcgcttg ttgaggcagt cgcagcgcag  20536

agggatgagg atgcgacgct gcccgcgttg catgatgggg taactcgccg ccaggaactc  20596
```

ctctatctga cggaaggcca tctgggcctt gacgccctcg gtgaaaaata gcccacagga 20656

cttgctggaa aacacgttat tgccacagtt gatgtcttcc gcgcagcagc gcgcatcttc 20716

gttcttcagc tgaaccacgt tgcgacccca gcggttctga accaccttgg ctttcgtggg 20776

atgctccttc agcgcccgct gtccgttctc gctggtcaca tccatttcca ccacgtgctc 20836

cttgcagacc atctccactc cgtggaaaca gaacagaatg ccctcctgtt gggtattgcg 20896

atgctcccac acggcgcacc cggtggactc ccagctcttg tgtttcaccc ccgcgtaggc 20956

ttccatgtaa gccattagaa atctgcccat cagctcagtg aaggtcttct ggttggtgaa 21016

ggttagcggc aggccgcggt gttcctcgtt caaccaagtt tgacagatct tgcggtacac 21076

ggctccctgg tcgggcagaa acttaaaagt cgttctgctc tcgttgtcca cgtggaactt 21136

ctccatcaac atcgtcatga cttccatgcc cttctcccag gcagtcacca gcggcgcgct 21196

ctcggggttc ttcaccaaca cggcggtgga ggggccctcg ccggccccga cgtccttcat 21256

ggacattttt tgaaactcca cggtgccgtc cgcgcggcgt actctgcgca tcggagggta 21316

gctgaagccc acctccatga cggtgctttc gccctcgctg tcggagacga tctccgggga 21376

gggcggcgga acgggggcag acttgcgagc cttcttcttg ggagggagcg gaggcacctc 21436

ctgctcgcgc tcgggactca tctcccgcaa gtagggggtg atggagcttc ctggttggtt 21496

ctgacggttg gccattgtat cctaggcaga aagacatgga gcttatgcgc gaggaaactt 21556

taaccgcccc gtcccccgtc agcgacgaag aggtcatcgt cgaacaggac ccgggctacg 21616

ttacgccgcc cgaggatctg gaggggccct tagacgaccg gcgcgacgct agtgagcggc 21676

aggaaaatga gaaagaggag gaggagggct gctacctcct ggaaggcgac gttttgctaa 21736

agcatttcgc caggcagagc accatactca aggaggcctt gcaagaccgc tccgaggtgc 21796

ccttggacgt cgccgcgctc tcccaggcct acgaggcgaa ccttttctcg ccccgagtgc 21856

ctccgaagag acagcccaac ggcacctgcg agcccaaccc gcgactcaac ttctaccccg 21916

tgttcgccgt gcccgaggcg ctggccacct accacatctt tttcaaaaac cagcgcattc 21976

ccctttcctg ccgggccaac cgcaccgcgg ccgataggaa gctaacactc agaaacggag 22036

tcagcatacc tgatatcacg tcactggagg aagtgcctaa gatcttcgag ggtctgggtc 22096

gagatgagaa gcgggcggcg aacgctctgc agaaagaaca gaaagagagt cagaacgtgc 22156

tggtggagct ggagggggac aacgcgcgtc tgaccgtcct caaacgttgc atagaagttt 22216

cccacttcgc ctacccggcc ctcaacctgc cgcccaaagt tatgaaatcg gtcatggacc 22276

agctactcat caagagagct gagcccctga atcccgacca ccctgaggcg gaaaactcag 22336

aggacggaaa gcccgtcgtc agcgacgagg agctcgagcg gtggctggaa accagggacc 22396

cccagcagtt gcaagagagg cgcaagatga tgatggcggc cgtgctggtc acggtggagc 22456

tagaatgcct gcaacggttt ttcagcgacg tggagacgct acgcaaaatc ggggagtccc 22516

tgcactacac cttccgccag ggctacgttc gccaggcctg caaaatctcc aacgtagagc 22576

tcagcaacct ggtttcctac atgggcatcc tccacgagaa ccggctgggg cagagcgtgc 22636

tgcactgcac cttgcaaggc gaggcgcgaa gggactacgt ccgagactgc gtctacctct 22696

tcctcaccct cacctggcag accgccatgg gcgtgtggca gcagtgcttg gaagagagaa 22756

acctcaaaga gctggacaaa ctcctctgcc gccagcggcg ggccctctgg accggcttca 22816

gcgagcgcac ggtcgcctgc gccctggcag acatcatttt cccagaacgc ctgatgaaaa 22876

ccttgcagaa cggcctgccg gatttcatca gtcagagcat cttgcaaaac ttccgctcct 22936

tcgtcctgga gcgctccggg atcttgcccg ccatgagctg cgcgctgcct tctgactttg 22996

tccccctttc ctaccgcgag tgccctcccc cactgtggag ccactgctac ctcttccaac 23056

tggccaactt tctggcctac cactccgacc tcatggaaga cgtgagcgga gaggggctgc 23116

tcgagtgcca ctgccgctgc aacctctgca ccccccacag atcgctggcc tgcaacaccg 23176

agctgctcag cgaaacccag gtcataggta ccttcgagat ccaggggccc cagcagcaag 23236

agggtgcttc cggcttgaag ctcactccgg cgctgtggac ctcggcttac ttacgcaaat 23296

ttgtagccga ggactaccac gcccacaaaa ttcagtttta cgaagaccaa tctcgaccac 23356

cgaaagcccc cctcacggcc tgcgtcatca cccagagcaa aatcctggcc caattgcaat 23416

ccatcaacca agcgcgccga gatttccttt tgaaaaaggg tcgggggggtg tacctggacc 23476

cccagaccgg cgaggaactc aacccgtcca cactttccgt cgaagcagcc cccccgagac 23536

atgccaccca agggaaccgc caagcagctg atcgctcggc agagagcgaa gaagcaagag 23596

ctgctccagc agcaggtgga ggacgaggaa gagctgtggg acagccaggc agaggaggtg 23656

tcagaggacg aggaggagat ggaaagctgg gacagcctag acgaggagga cgagctttca 23716

gaggaagagg cgaccgaaga aaaaccacct gcatccagcg cgccttctct gagccgacag 23776

ccgaagcccc ggcccccgac gcccccggcc ggctcactca aagccagccg taggtgggac 23836

gccaccggat ctccagcggc agcggcaacg gcagcgggta aggccaaacg cgagcggcgg 23896

gggtattgct cctggcggac ccacaaaagc agtatcgtga actgcttgca acactgcggg 23956

ggaaacatct cctttgcccg acgctacctc ctcttccatc acggtgtggc cttccctcgc 24016

```
aacgttctct attattaccg tcatctctac agcccctacg aaacgctcgg agaaaaaagc  24076

taaggcctcc tctgccgcga ggaaaaactc cgccgccgct gccgccaagg atccgccggc  24136

caccgaggag ctgagaaagc gcatctttcc cactctgtat gctatctttc agcaaagccg  24196

cgggcagcac cctcagcgcg aactgaaaat aaaaaaccgc tccttccgct cactcacccg  24256

cagctgtctg taccacaaga gagaagacca gctgcagcgc accctggacg acgccgaagc  24316

actgttcagc aaatactgct cagcgtctct taaagactaa aagacccgcg ctttttcccc  24376

ctcgggcgcc aaaacccacg tcatcgccag catgagcaag gagattccca ccccttacat  24436

gtggagctat cagccccaga tgggcctggc cgcgggggcc gcccaggact actccagcaa  24496

aatgaactgg ctcagcgccg gcccccacat gatctcacga gttaacggca tccgagccca  24556

ccgaaaccag atcctcttag aacaggcggc aatcaccgcc acacccggc gccaactcaa  24616

cccgcccagt tggcccgccg cccaggtgta tcaggaaact ccccgcccga ccacagtcct  24676

cctgccacgc gacgcggagg ccgaagtcct catgactaac tctggggtac aattagcggg  24736

cgggtccagg tacgccaggt acagaggtcg ggccgctcct tactctcccg ggagtataaa  24796

gagggtgatc attcgaggcc gaggtatcca gctcaacgac gaggcggtga gctcctcaac  24856

cggtctcaga cctgacggag tcttccagct cggaggagcg ggccgctctt ccttcaccac  24916

tcgccaggcc tacctgaccc tgcagagctc ttcctcgcag ccgcgctccg ggggaatcgg  24976

cactctccag ttcgtggaag agttcgtccc ctccgtctac ttcaacccgt tttccggctc  25036

acctggacgc tacccggacg ccttcattcc caactttgac gcagtgagtg aatccgtgga  25096

cggctacgac tgatgacaga tggtgcggcc gtgagagctc ggctgcgaca tctgcatcac  25156

tgccgccagc ctcgctgcta cgctcgggag gcgatcgtgt tcagctactt tgagctgccg  25216

gacgagcacc ctcagggacc ggctcacggg ttgaaactcg agattgagaa cgcgcttgag  25276

tctcacctca tcgacgcctt caccgcccgg cctctcctgg tagaaaccga acgcgggatc  25336

actaccatca ccctgttctg catctgcccc acgcccggat tac atg aag atc tgt     25391
                                          Met Lys Ile Cys
                                                        1430

gtt gtc atc ttt gcg  ctc agt tta ata aaa  act gaa ctt ttt gcc       25436
Val Val Ile Phe Ala  Leu Ser Leu Ile Lys  Thr Glu Leu Phe Ala
               1435                 1440                 1445

gta cct tca acg cca  cgc gtt gtt tct cct  tgt gaa aaa acc cca       25481
Val Pro Ser Thr Pro  Arg Val Val Ser Pro  Cys Glu Lys Thr Pro
               1450                 1455                 1460
```

```
gga gtc ctt aac tta  cac ata gca aaa ccc  ttg tat ttt acc ata      25526
Gly Val Leu Asn Leu  His Ile Ala Lys Pro  Leu Tyr Phe Thr Ile
                1465                1470                1475

gaa aaa caa cta gcc  ctt tca att gga aaa  ggg tta aca att tct      25571
Glu Lys Gln Leu Ala  Leu Ser Ile Gly Lys  Gly Leu Thr Ile Ser
                1480                1485                1490

gct aca gga cag ttg  gaa agc aca gca agc  gta cag gac agc gct      25616
Ala Thr Gly Gln Leu  Glu Ser Thr Ala Ser  Val Gln Asp Ser Ala
                1495                1500                1505

aca cca ccc cta cgt  ggt att tcc cct tta  aag ctg aca gac aac      25661
Thr Pro Pro Leu Arg  Gly Ile Ser Pro Leu  Lys Leu Thr Asp Asn
                1510                1515                1520

ggt tta aca tta agc  tat tca gat ccc ctg  cgt gtg gta ggt gac      25706
Gly Leu Thr Leu Ser  Tyr Ser Asp Pro Leu  Arg Val Val Gly Asp
                1525                1530                1535

caa ctt acg ttt aat  ttt act tct cca cta  cgt tac gaa aat ggc      25751
Gln Leu Thr Phe Asn  Phe Thr Ser Pro Leu  Arg Tyr Glu Asn Gly
                1540                1545                1550

agt ctt aca ttc aac  tac act tct ccc atg  aca cta ata aac aac      25796
Ser Leu Thr Phe Asn  Tyr Thr Ser Pro Met  Thr Leu Ile Asn Asn
                1555                1560                1565

agt ctt gct att aac  gtc aat acc tcc aaa  ggc ctc agt agt gac      25841
Ser Leu Ala Ile Asn  Val Asn Thr Ser Lys  Gly Leu Ser Ser Asp
                1570                1575                1580

aac ggc aca ctc gct  gta aat gtt act cca  gat ttt aga ttt aac      25886
Asn Gly Thr Leu Ala  Val Asn Val Thr Pro  Asp Phe Arg Phe Asn
                1585                1590                1595

agc tct ggt gcc tta  act ttt ggc ata caa  agt cta tgg act ttt      25931
Ser Ser Gly Ala Leu  Thr Phe Gly Ile Gln  Ser Leu Trp Thr Phe
                1600                1605                1610

cca acc aaa act cct  aac tgt acc gtg ttt  acc gaa agt gac tcc      25976
Pro Thr Lys Thr Pro  Asn Cys Thr Val Phe  Thr Glu Ser Asp Ser
                1615                1620                1625

ctg ctg agt ctt tgc  ttg act aaa tgc gga  gct cac gta ctt gga      26021
Leu Leu Ser Leu Cys  Leu Thr Lys Cys Gly  Ala His Val Leu Gly
                1630                1635                1640

agc gtg agt tta agc  gga gtg gca gga acc  atg cta aaa atg acc      26066
Ser Val Ser Leu Ser  Gly Val Ala Gly Thr  Met Leu Lys Met Thr
                1645                1650                1655

cac act tct gtt acc  gtt cag ttt tcg ttt  gat gac agt ggt aaa      26111
His Thr Ser Val Thr  Val Gln Phe Ser Phe  Asp Asp Ser Gly Lys
                1660                1665                1670

cta ata ttc tct cca  ctt gcg aac aac act  tgg ggt gtt cga caa      26156
```

```
Leu Ile Phe Ser Pro  Leu Ala Asn Asn Thr  Trp Gly Val Arg Gln
                1675                1680                1685

agc gag agt ccg ttg  ccc aac cca tcc ttc  aac gct ctc acg ttt      26201
Ser Glu Ser Pro Leu  Pro Asn Pro Ser Phe  Asn Ala Leu Thr Phe
                1690                1695                1700

atg cca aac agt acc  att tat tct aga gga  gca agt aac gaa cct      26246
Met Pro Asn Ser Thr  Ile Tyr Ser Arg Gly  Ala Ser Asn Glu Pro
                1705                1710                1715

caa aac aat tat tat  gtc cag acg tat ctt  aga ggc aac gtg cga      26291
Gln Asn Asn Tyr Tyr  Val Gln Thr Tyr Leu  Arg Gly Asn Val Arg
                1720                1725                1730

aag cca att cta cta  act gtt acc tac aac  tca gtt aat tca gga      26336
Lys Pro Ile Leu Leu  Thr Val Thr Tyr Asn  Ser Val Asn Ser Gly
                1735                1740                1745

tat tcc tta act ttt  aaa tgg gat gct gtc  gcc aat gaa aaa ttt      26381
Tyr Ser Leu Thr Phe  Lys Trp Asp Ala Val  Ala Asn Glu Lys Phe
                1750                1755                1760

gcc act cct aca tct  tcg ttt tgc tat gtt  gca gag caa taa          26423
Ala Thr Pro Thr Ser  Ser Phe Cys Tyr Val  Ala Glu Gln
                1765                1770

aaccctgtta ccccaccgtc tcgtttttttt cag atg aaa  cga gcg aga gtt      26474
                                     Met Lys  Arg Ala Arg Val
                                         1775

gat  gaa gac ttc aac cca  gtg tac cct tat gac  ccc cca tac gct      26519
Asp  Glu Asp Phe Asn Pro  Val Tyr Pro Tyr Asp  Pro Pro Tyr Ala
1780                1785                1790

ccc  gtc atg ccc ttc att  act ccg cct ttt acc  tcc tcg gat ggg      26564
Pro  Val Met Pro Phe Ile  Thr Pro Pro Phe Thr  Ser Ser Asp Gly
1795                1800                1805

ttg  cag gaa aaa cca ctt  gga gtg tta agt tta  aac tac agg gat      26609
Leu  Gln Glu Lys Pro Leu  Gly Val Leu Ser Leu  Asn Tyr Arg Asp
1810                1815                1820

ccc  att act aca caa aat  ggg tct ctc acg tta  aaa cta gga aac      26654
Pro  Ile Thr Thr Gln Asn  Gly Ser Leu Thr Leu  Lys Leu Gly Asn
1825                1830                1835

ggc  ctc act cta aac aac  cag gga cag tta aca  tca act gct ggc      26699
Gly  Leu Thr Leu Asn Asn  Gln Gly Gln Leu Thr  Ser Thr Ala Gly
1840                1845                1850

gaa  gtg gag cct ccg ctc  act aat gct aac aac  aaa ctt gca cta      26744
Glu  Val Glu Pro Pro Leu  Thr Asn Ala Asn Asn  Lys Leu Ala Leu
1855                1860                1865

gcc  tat agc gaa cca tta  gca gta aaa agc aac  cgc cta act cta      26789
Ala  Tyr Ser Glu Pro Leu  Ala Val Lys Ser Asn  Arg Leu Thr Leu
```

```
1870                1875                1880

tca  cac acc gct ccc ctt  gtc atc gct aat aat  tct tta gcg ttg     26834
Ser  His Thr Ala Pro Leu  Val Ile Ala Asn Asn  Ser Leu Ala Leu
1885                1890                1895

caa  gtt tca gag cct att  ttt gta aat gac gat  gac aag cta gcc     26879
Gln  Val Ser Glu Pro Ile  Phe Val Asn Asp Asp  Asp Lys Leu Ala
1900                1905                1910

ctg  cag aca gcc gcc ccc  ctt gta acc aac gct  ggc acc ctt cgc     26924
Leu  Gln Thr Ala Ala Pro  Leu Val Thr Asn Ala  Gly Thr Leu Arg
1915                1920                1925

tta  cag agc gct gcc cct  tta gga ttg gtt gaa  aat act ctt aaa     26969
Leu  Gln Ser Ala Ala Pro  Leu Gly Leu Val Glu  Asn Thr Leu Lys
1930                1935                1940

ctg  ctg ttt tct aaa ccc  ttg tat ttg caa aat  gat ttt ctt gca     27014
Leu  Leu Phe Ser Lys Pro  Leu Tyr Leu Gln Asn  Asp Phe Leu Ala
1945                1950                1955

tta  gcc att gaa cgc ccc  ctg gct gta gca gcc  gca ggt act ctg     27059
Leu  Ala Ile Glu Arg Pro  Leu Ala Val Ala Ala  Ala Gly Thr Leu
1960                1965                1970

acc  cta caa ctt act cct  cca tta aag act aac  gat gac ggg cta     27104
Thr  Leu Gln Leu Thr Pro  Pro Leu Lys Thr Asn  Asp Asp Gly Leu
1975                1980                1985

aca  cta tcc aca gtc gag  cca tta act gta aaa  aac gga aac cta     27149
Thr  Leu Ser Thr Val Glu  Pro Leu Thr Val Lys  Asn Gly Asn Leu
1990                1995                2000

ggc  ttg caa ata tcg cgc  cct tta gtt gtt caa  aac aac ggc ctt     27194
Gly  Leu Gln Ile Ser Arg  Pro Leu Val Val Gln  Asn Asn Gly Leu
2005                2010                2015

tcg  ctt gct att acc ccc  ccg ctg cgt ttg ttt  aac agc gac ccc     27239
Ser  Leu Ala Ile Thr Pro  Pro Leu Arg Leu Phe  Asn Ser Asp Pro
2020                2025                2030

gtt  ctt ggt ttg ggc ttc  act ttt ccc cta gct  gtc aca aac aac     27284
Val  Leu Gly Leu Gly Phe  Thr Phe Pro Leu Ala  Val Thr Asn Asn
2035                2040                2045

ctc  ctc tcc tta aac atg  gga gac gga gtt aaa  ctt acc tat aat     27329
Leu  Leu Ser Leu Asn Met  Gly Asp Gly Val Lys  Leu Thr Tyr Asn
2050                2055                2060

aaa  cta aca gcc aat ttg  ggt agg gat tta caa  ttt gaa aac ggt     27374
Lys  Leu Thr Ala Asn Leu  Gly Arg Asp Leu Gln  Phe Glu Asn Gly
2065                2070                2075

gcg  att gcc gta acg ctt  act gcc gaa tta cct  ttg caa tac act     27419
Ala  Ile Ala Val Thr Leu  Thr Ala Glu Leu Pro  Leu Gln Tyr Thr
2080                2085                2090
```

```
aac  aaa ctt caa ctg aat  att gga gct ggc ctt  cgt tac aat gga     27464
Asn  Lys Leu Gln Leu Asn  Ile Gly Ala Gly Leu  Arg Tyr Asn Gly
2095                 2100                 2105

gcc  agc aga aaa cta gat  gta aac att aac caa  aat aaa ggc tta     27509
Ala  Ser Arg Lys Leu Asp  Val Asn Ile Asn Gln  Asn Lys Gly Leu
2110                 2115                 2120

act  tgg gac aac gat gca  gtt att ccc aaa cta  gga tcg ggc tta     27554
Thr  Trp Asp Asn Asp Ala  Val Ile Pro Lys Leu  Gly Ser Gly Leu
2125                 2130                 2135

caa  ttt gac cct aat ggc  aac atc gct gtt atc  cct gaa acc gtg     27599
Gln  Phe Asp Pro Asn Gly  Asn Ile Ala Val Ile  Pro Glu Thr Val
2140                 2145                 2150

aag  ccg caa acg tta tgg  acg act gca gat ccc  tcg cct aac tgc     27644
Lys  Pro Gln Thr Leu Trp  Thr Thr Ala Asp Pro  Ser Pro Asn Cys
2155                 2160                 2165

tca  gtg tac cag gac ttg  gat gcc agg ctg tgg  ctc gct ctt gtt     27689
Ser  Val Tyr Gln Asp Leu  Asp Ala Arg Leu Trp  Leu Ala Leu Val
2170                 2175                 2180

aaa  agt ggc gac atg gtg  cat gga agc att gcc  cta aaa gcc cta     27734
Lys  Ser Gly Asp Met Val  His Gly Ser Ile Ala  Leu Lys Ala Leu
2185                 2190                 2195

aaa  ggg acg ttg cta aat  cct aca gcc agc tac  att tcc att gtg     27779
Lys  Gly Thr Leu Leu Asn  Pro Thr Ala Ser Tyr  Ile Ser Ile Val
2200                 2205                 2210

ata  tat ttt tac agc aac  gga gtc agg cgt acc  aac tat cca acg     27824
Ile  Tyr Phe Tyr Ser Asn  Gly Val Arg Arg Thr  Asn Tyr Pro Thr
2215                 2220                 2225

ttt  gac aac gaa ggc acc  tta gct aac agc gcc  act tgg gga tac     27869
Phe  Asp Asn Glu Gly Thr  Leu Ala Asn Ser Ala  Thr Trp Gly Tyr
2230                 2235                 2240

cga  cag ggg caa tct gct  aac act aat gtg acc  aat gcc act gaa     27914
Arg  Gln Gly Gln Ser Ala  Asn Thr Asn Val Thr  Asn Ala Thr Glu
2245                 2250                 2255

ttt  atg ccc agc tca agc  agg tac ccc gtg aat  aaa gga gac aac     27959
Phe  Met Pro Ser Ser Ser  Arg Tyr Pro Val Asn  Lys Gly Asp Asn
2260                 2265                 2270

att  caa aat caa tct ttt  tca tac acc tgt att  aaa gga gat ttt     28004
Ile  Gln Asn Gln Ser Phe  Ser Tyr Thr Cys Ile  Lys Gly Asp Phe
2275                 2280                 2285

gct  atg cct gtc ccg ttc  cgt gta aca tat aat  cac gcc ctg gaa     28049
Ala  Met Pro Val Pro Phe  Arg Val Thr Tyr Asn  His Ala Leu Glu
2290                 2295                 2300
```

```
ggg  tat tcc ctt aag ttc  acc tgg cgc gtt gta  gcc aat cag gcc     28094
Gly  Tyr Ser Leu Lys Phe  Thr Trp Arg Val Val  Ala Asn Gln Ala
2305                 2310                 2315

ttt  gat att cct tgc tgt  tca ttt tca tac atc  aca gaa taa         28136
Phe  Asp Ile Pro Cys Cys  Ser Phe Ser Tyr Ile  Thr Glu
2320                 2325                 2330

aaaaccactt tttcatttta atttcttttt attttacacg aacagtgaga cttcctccac  28196

ccttccattt gacagcatac accagcctct ccccccttcat agcagtaaac tgttgtgaat  28256

cagtccggta tttgggagtt aaaatccaaa cagtctcttt ggtgatgaaa cgtcgatcag  28316

taatggacac aaatccctgg gacaggtttt ccaacgtttc ggtgaaaaac tgcacaccgc  28376

cctacaaaac aaacaggttc aggctctcca cgggttatct ccccgatcaa actcagacag  28436

ggtaaaggtg cggtggtgtt ccactaaacc acgcaggtgg cgctgtctga acctctcggt  28496

gcgactcctg tgaggctggt aagaagttag attgtccagt agcctcacag catgtatcat  28556

cagtctacga gtgcgtctgg cgcagcagcg catctgaatc tcactgagat tccggcaaga  28616

atcgcacacc atcacaatca ggttgttcat gatcccatag ctgaacacgc tccagccaaa  28676

gctcattcgc tccaacagcg ccaccgcgtg tccgtccaac cttactttaa cataaatcag  28736

gtgtctgccg cgtacaaaca tgctacccac atacagaact tcccggggca ggcccctgtt  28796

caccacctgt ctgtaccagg gaaacctcac atttatcagg gagccataga tggccatttt  28856

aaaccaatta gctaataccg ccccaccagc tctacactga agagaaccgg gagagttaca  28916

atgacagtga ataatccatc tctcataacc cctgatggtc tgatgaaaat ctagatctaa  28976

cgtggcacaa caaatacaca ctttcatata cattttcata acatgttttt cccaggccgt  29036

taaaatacaa tcccaataca cgggccactc ctgcagtaca ataaagctaa tacaagatgg  29096

tatactcctc acctcactga cactgtgcat gttcatattt tcacattcta agtaccgaga  29156

gttctcctct acagcagcac tgctgcggtc ctcacaaggt ggtagctggt gatgattgta  29216

gggggccagt ctgcagcgat accgtctgtc gcgttgcatc gtagaccagg aaccgacgca  29276

cctcctcgta cttgtggtag cagaaccacg tccgctgcca gcacgtctcc acgtaacgcc  29336

ggtccctgcg tcgctcacgc tccctcctca atgcaaagtg caaccactct tgtaatccac  29396

acagatccct ctcggcctcc ggggtgatgc acacctcaaa cctacagatg tctcggtaca  29456

gttccaaaca cgtagtgagg gcgagttcca accaagacag acagcctgat ctatcccgac  29516

acactggagg tggaggaaga cacggaagag gcatgttatt ccaagcgatt caccaacggg  29576

tcgaaatgaa gatcccgaag atgacaacgg tcgcctccgg agccctgatg gaatttaaca  29636
```

gccagatcaa acgttatgcg attctccaag ctatcgatcg ccgcttccaa aagagcctgg 29696

acccgcactt ccacaaacac cagcaaagca aaagcactat tatcaaactc ttcaatcatc 29756

aagctgcagg actgtacaat gcctaagtaa ttttcgtttc tccactcgcg aatgatgtcg 29816

cggcagatag tctgaaggtt catcccgtgc agggtaaaaa gctccgaaag ggcgccctct 29876

acagccatgc gtagacacac catcatgact gcaagatatc gggctcctga gacacctgca 29936

gcagatttaa cagatcaagg tcaggttgct ctccgcgatc acgaatctcc atccgcaagg 29996

tcatttgcaa aaaattaaat aaatctatgc cgactagatc tgtcaactcc gcattaggaa 30056

ccaaatcagg tgtggctacg cagcacaaaa gttccaggga tggtgccaaa ctcactagaa 30116

ccgctcccga gtaacaaaac tgatgaatgg gagtaacaca gtgtaaaatg tgcaaccaaa 30176

aatcactaag gtgctccttt aaaaagtcca gtacttctat attcagtccg tgcaagtact 30236

gaagcaactg tgcgggaata tgcacaacaa aaaaaatagg gcggctcaga tacatgttga 30296

cctaaaataa aaagaatcat taaactaaag aagcttggcg aacggtggga taaatgacac 30356

gctccagcag cagacaggca accggctgtc cccgggaacc gcggtaaaat tcatccgaat 30416

gattaaaaag aacaacagaa acttcccacc atgtactcgg ttggatctcc tgagcacaca 30476

gcaatacccc cctcacattc atgtccgcca cagaaaaaaa acgtcccaga tacccagcgg 30536

ggatatccaa cgacagctgc aaagacagca aaacaatccc tctgggagcg atcacaaaat 30596

cctccggtga aaaaagcaca tacatattag aataaccctg ttgctggggc aaaaaggccc 30656

ggcgtcccag caaatgcaca taaatatgtt catcagccat tgccccgtct taccgcgtaa 30716

tcagccacga aaaaatcgag ctaaaattca cccaacagcc tatagctata tatacactcc 30776

gcccaatgac gctaataccg caccacccac gaccaaagtt cacccacacc cacaaaaccc 30836

gcgaaaatcc agcgccgtca gcacttccgc aatttcagtc tcacaacgtc acttccgcgc 30896

gccttttcac attcccacac acacccgcgc ccttcgcccc gccctcgcgc caccccgcgt 30956

caccgcacgt caccccggcc ccgcctcgct cctccccgct cattatcata ttggcacgtt 31016

tccagaataa ggtatattat tgatgatg 31044

<210> 30
<211> 505
<212> PRT
<213> simian adenovirus SV-25

<400> 30

```
Met Arg Arg Ala Val Arg Val Thr Pro Ala Ala Tyr Glu Gly Pro Pro
1               5                  10                  15

Pro Ser Tyr Glu Ser Val Met Gly Ser Ala Asn Val Pro Ala Thr Leu
            20                  25                  30

Glu Ala Pro Tyr Val Pro Pro Arg Tyr Leu Gly Pro Thr Glu Gly Arg
        35                  40                  45

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

Val Tyr Leu Val Asp Asn Lys Ser Ala Asp Ile Ala Ser Leu Asn Tyr
65                  70                  75                  80

Gln Asn Asp His Ser Asn Phe Leu Thr Thr Val Val Gln Asn Asn Asp
                85                  90                  95

Phe Thr Pro Thr Glu Ala Gly Thr Gln Thr Ile Asn Phe Asp Glu Arg
            100                 105                 110

Ser Arg Trp Gly Gly Gln Leu Lys Thr Ile Leu His Thr Asn Met Pro
        115                 120                 125

Asn Ile Asn Glu Phe Met Ser Thr Asn Lys Phe Arg Ala Lys Leu Met
    130                 135                 140

Val Glu Lys Ser Asn Ala Glu Thr Arg Gln Pro Arg Tyr Glu Trp Phe
145                 150                 155                 160

Glu Phe Thr Ile Pro Glu Gly Asn Tyr Ser Glu Thr Met Thr Ile Asp
                165                 170                 175

Leu Met Asn Asn Ala Ile Val Asp Asn Tyr Leu Gln Val Gly Arg Gln
            180                 185                 190

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

Phe Arg Leu Gly Trp Asp Pro Val Thr Lys Leu Val Met Pro Gly Val
    210                 215                 220

Tyr Thr Asn Glu Ala Phe His Pro Asp Ile Val Leu Leu Pro Gly Cys
```

225 230 235 240

Gly Val Asp Phe Thr Gln Ser Arg Leu Ser Asn Leu Leu Gly Ile Arg
245 250 255

Lys Arg Arg Pro Phe Gln Glu Gly Phe Gln Ile Met Tyr Glu Asp Leu
260 265 270

Glu Gly Gly Asn Ile Pro Ala Leu Leu Asp Val Ser Lys Tyr Glu Ala
275 280 285

Ser Ile Gln Arg Ala Lys Ala Glu Gly Arg Glu Ile Arg Gly Asp Thr
290 295 300

Phe Ala Val Ala Pro Gln Asp Leu Glu Ile Val Pro Leu Thr Lys Asp
305 310 315 320

Ser Lys Asp Arg Ser Tyr Asn Ile Ile Asn Asn Thr Thr Asp Thr Leu
325 330 335

Tyr Arg Ser Trp Phe Leu Ala Tyr Asn Tyr Gly Asp Pro Glu Lys Gly
340 345 350

Val Arg Ser Trp Thr Ile Leu Thr Thr Thr Asp Val Thr Cys Gly Ser
355 360 365

Gln Gln Val Tyr Trp Ser Leu Pro Asp Met Met Gln Asp Pro Val Thr
370 375 380

Phe Arg Pro Ser Thr Gln Val Ser Asn Phe Pro Val Val Gly Thr Glu
385 390 395 400

Leu Leu Pro Val His Ala Lys Ser Phe Tyr Asn Glu Gln Ala Val Tyr
405 410 415

Ser Gln Leu Ile Arg Gln Ser Thr Ala Leu Thr His Val Phe Asn Arg
420 425 430

Phe Pro Glu Asn Gln Ile Leu Val Arg Pro Pro Ala Pro Thr Ile Thr
435 440 445

Thr Val Ser Glu Asn Val Pro Ala Leu Thr Asp His Gly Thr Leu Pro
450 455 460

```
Leu Arg Ser Ser Ile Ser Gly Val Gln Arg Val Thr Ile Thr Asp Ala
465                 470                 475                 480

Arg Arg Arg Thr Cys Pro Tyr Val Tyr Lys Ala Leu Gly Val Val Ala
                485                 490                 495

Pro Lys Val Leu Ser Ser Arg Thr Phe
            500                 505
```

<210> 31
<211> 921
<212> PRT
<213> simian adenovirus SV-25

<400> 31

```
Met Ala Thr Pro Ser Met Met Pro Gln Trp Ser Tyr Met His Ile Ala
1               5                   10                  15

Gly Gln Asp Ala Ser Glu Tyr Leu Ser Pro Gly Leu Val Gln Phe Ala
            20                  25                  30

Arg Ala Thr Asp Thr Tyr Phe Ser Leu Gly Asn Lys Phe Arg Asn Pro
        35                  40                  45

Thr Val Ala Pro Thr His Asp Val Thr Thr Asp Arg Ser Gln Arg Leu
    50                  55                  60

Thr Leu Arg Phe Val Pro Val Asp Arg Glu Asp Thr Ala Tyr Ser Tyr
65                  70                  75                  80

Lys Val Arg Tyr Thr Leu Ala Val Gly Asp Asn Arg Val Leu Asp Met
                85                  90                  95

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

Phe Lys Pro Tyr Ser Gly Thr Ala Tyr Asn Ser Leu Ala Pro Lys Gly
        115                 120                 125

Ala Pro Asn Pro Ser Glu Trp Thr Asp Thr Ser Asp Asn Lys Leu Lys
    130                 135                 140
```

```
Ala Tyr Ala Gln Ala Pro Tyr Gln Ser Gln Gly Leu Thr Lys Asp Gly
145                 150                 155                 160

Ile Gln Val Gly Leu Val Val Thr Glu Ser Gly Gln Thr Pro Gln Tyr
                165                 170                 175

Ala Asn Lys Val Tyr Gln Pro Glu Pro Gln Ile Gly Glu Asn Gln Trp
            180                 185                 190

Asn Leu Glu Gln Glu Asp Lys Ala Ala Gly Arg Val Leu Lys Lys Asp
        195                 200                 205

Thr Pro Met Phe Pro Cys Tyr Gly Ser Tyr Ala Arg Pro Thr Asn Glu
    210                 215                 220

Gln Gly Gly Gln Ala Lys Asn Gln Glu Val Asp Leu Gln Phe Phe Ala
225                 230                 235                 240

Thr Pro Gly Asp Thr Gln Asn Thr Ala Lys Val Val Leu Tyr Ala Glu
                245                 250                 255

Asn Val Asn Leu Glu Thr Pro Asp Thr His Leu Val Phe Lys Pro Asp
            260                 265                 270

Asp Asp Ser Thr Ser Ser Lys Leu Leu Leu Gly Gln Gln Ala Ala Pro
        275                 280                 285

Asn Arg Pro Asn Tyr Ile Gly Phe Arg Asp Asn Phe Ile Gly Leu Met
    290                 295                 300

Tyr Tyr Asn Ser Thr Gly Asn Met Gly Val Leu Ala Gly Gln Ala Ser
305                 310                 315                 320

Gln Leu Asn Ala Val Val Asp Leu Gln Asp Arg Asn Thr Glu Leu Ser
                325                 330                 335

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

Met Trp Asn Gln Ala Val Asp Ser Tyr Asp Pro Asp Val Arg Ile Ile
        355                 360                 365

Glu Asn His Gly Val Glu Asp Glu Leu Pro Asn Tyr Cys Phe Pro Leu
```

370 375 380

Gly Gly Met Val Val Thr Asp Asn Tyr Asn Ser Val Thr Pro Gln Asn
385 390 395 400

Gly Gly Ser Gly Asn Thr Trp Gln Ala Asp Asn Thr Thr Phe Ser Gln
405 410 415

Arg Gly Ala Gln Ile Gly Ser Gly Asn Met Phe Ala Leu Glu Ile Asn
420 425 430

Leu Gln Ala Asn Leu Trp Arg Gly Phe Leu Tyr Ser Asn Ile Gly Leu
435 440 445

Tyr Leu Pro Asp Ser Leu Lys Ile Thr Pro Asp Asn Ile Thr Leu Pro
450 455 460

Glu Asn Lys Asn Thr Tyr Gln Tyr Met Asn Gly Arg Val Thr Pro Pro
465 470 475 480

Gly Leu Ile Asp Thr Tyr Val Asn Val Gly Ala Arg Trp Ser Pro Asp
485 490 495

Val Met Asp Ser Ile Asn Pro Phe Asn His His Arg Asn Ala Gly Leu
500 505 510

Arg Tyr Arg Ser Met Leu Leu Gly Asn Gly Arg Tyr Val Pro Phe His
515 520 525

Ile Gln Val Pro Gln Lys Phe Phe Ala Ile Lys Asn Leu Leu Leu Leu
530 535 540

Pro Gly Ser Tyr Thr Tyr Glu Trp Asn Phe Arg Lys Asp Val Asn Met
545 550 555 560

Ile Leu Gln Ser Ser Leu Gly Asn Asp Leu Arg Val Asp Gly Ala Ser
565 570 575

Ile Arg Phe Asp Ser Ile Asn Leu Tyr Ala Asn Phe Phe Pro Met Ala
580 585 590

His Asn Thr Ala Ser Thr Leu Glu Ala Met Leu Arg Asn Asp Thr Asn
595 600 605

```
Asp Gln Ser Phe Asn Asp Tyr Leu Cys Ala Ala Asn Met Leu Tyr Pro
    610                 615                 620

Ile Pro Ala Asn Ala Thr Ser Val Pro Ile Ser Ile Pro Ser Arg Asn
625                 630                 635                 640

Trp Ala Ala Phe Arg Gly Trp Ser Phe Thr Arg Leu Lys Thr Lys Glu
                645                 650                 655

Thr Pro Ser Leu Gly Ser Gly Phe Asp Pro Tyr Phe Val Tyr Ser Gly
            660                 665                 670

Ser Ile Pro Tyr Leu Asp Gly Thr Phe Tyr Leu Asn His Thr Phe Lys
        675                 680                 685

Lys Val Ser Ile Met Phe Asp Ser Ser Val Ser Trp Pro Gly Asn Asp
    690                 695                 700

Arg Leu Leu Thr Pro Asn Glu Phe Glu Ile Lys Arg Ser Val Asp Gly
705                 710                 715                 720

Glu Gly Tyr Asn Val Ala Gln Ser Asn Met Thr Lys Asp Trp Phe Leu
                725                 730                 735

Ile Gln Met Leu Ser His Tyr Asn Ile Gly Tyr Gln Gly Phe Tyr Val
            740                 745                 750

Pro Glu Asn Tyr Lys Asp Arg Met Tyr Ser Phe Phe Arg Asn Phe Gln
        755                 760                 765

Pro Met Ser Arg Gln Val Val Asp Thr Val Thr Tyr Thr Asp Tyr Lys
    770                 775                 780

Asp Val Lys Leu Pro Tyr Gln His Asn Asn Ser Gly Phe Val Gly Tyr
785                 790                 795                 800

Met Gly Pro Thr Met Arg Glu Gly Gln Ala Tyr Pro Ala Asn Tyr Pro
                805                 810                 815

Tyr Pro Leu Ile Gly Glu Thr Ala Val Pro Ser Leu Thr Gln Lys Lys
            820                 825                 830
```

Phe Leu Cys Asp Arg Val Met Trp Arg Ile Pro Phe Ser Ser Asn Phe
835 840 845

Met Ser Met Gly Ser Leu Thr Asp Leu Gly Gln Asn Met Leu Tyr Ala
850 855 860

Asn Ser Ala His Ala Leu Asp Met Thr Phe Glu Val Asp Pro Met Asp
865 870 875 880

Glu Pro Thr Leu Leu Tyr Val Leu Phe Glu Val Phe Asp Val Val Arg
885 890 895

Ile His Gln Pro His Arg Gly Val Ile Glu Ala Val Tyr Leu Arg Thr
900 905 910

Pro Phe Ser Ala Gly Asn Ala Thr Thr
915 920

<210> 32
<211> 347
<212> PRT
<213> simian adenovirus SV-25

<400> 32

Met Lys Ile Cys Val Val Ile Phe Ala Leu Ser Leu Ile Lys Thr Glu
1 5 10 15

Leu Phe Ala Val Pro Ser Thr Pro Arg Val Val Ser Pro Cys Glu Lys
20 25 30

Thr Pro Gly Val Leu Asn Leu His Ile Ala Lys Pro Leu Tyr Phe Thr
35 40 45

Ile Glu Lys Gln Leu Ala Leu Ser Ile Gly Lys Gly Leu Thr Ile Ser
50 55 60

Ala Thr Gly Gln Leu Glu Ser Thr Ala Ser Val Gln Asp Ser Ala Thr
65 70 75 80

Pro Pro Leu Arg Gly Ile Ser Pro Leu Lys Leu Thr Asp Asn Gly Leu
85 90 95

Thr Leu Ser Tyr Ser Asp Pro Leu Arg Val Val Gly Asp Gln Leu Thr

```
            100                 105                 110

Phe Asn Phe Thr Ser Pro Leu Arg Tyr Glu Asn Gly Ser Leu Thr Phe
        115                 120                 125

Asn Tyr Thr Ser Pro Met Thr Leu Ile Asn Asn Ser Leu Ala Ile Asn
    130                 135                 140

Val Asn Thr Ser Lys Gly Leu Ser Ser Asp Asn Gly Thr Leu Ala Val
145                 150                 155                 160

Asn Val Thr Pro Asp Phe Arg Phe Asn Ser Ser Gly Ala Leu Thr Phe
                165                 170                 175

Gly Ile Gln Ser Leu Trp Thr Phe Pro Thr Lys Thr Pro Asn Cys Thr
            180                 185                 190

Val Phe Thr Glu Ser Asp Ser Leu Leu Ser Leu Cys Leu Thr Lys Cys
        195                 200                 205

Gly Ala His Val Leu Gly Ser Val Ser Leu Ser Gly Val Ala Gly Thr
    210                 215                 220

Met Leu Lys Met Thr His Thr Ser Val Thr Val Gln Phe Ser Phe Asp
225                 230                 235                 240

Asp Ser Gly Lys Leu Ile Phe Ser Pro Leu Ala Asn Asn Thr Trp Gly
                245                 250                 255

Val Arg Gln Ser Glu Ser Pro Leu Pro Asn Pro Ser Phe Asn Ala Leu
            260                 265                 270

Thr Phe Met Pro Asn Ser Thr Ile Tyr Ser Arg Gly Ala Ser Asn Glu
        275                 280                 285

Pro Gln Asn Asn Tyr Tyr Val Gln Thr Tyr Leu Arg Gly Asn Val Arg
    290                 295                 300

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

Ser Leu Thr Phe Lys Trp Asp Ala Val Ala Asn Glu Lys Phe Ala Thr
                325                 330                 335
```

Pro Thr Ser Ser Phe Cys Tyr Val Ala Glu Gln
340 345

<210> 33
<211> 559
<212> PRT
<213> simian adenovirus SV-25

<400> 33

Met Lys Arg Ala Arg Val Asp Glu Asp Phe Asn Pro Val Tyr Pro Tyr
1 5 10 15

Asp Pro Pro Tyr Ala Pro Val Met Pro Phe Ile Thr Pro Pro Phe Thr
20 25 30

Ser Ser Asp Gly Leu Gln Glu Lys Pro Leu Gly Val Leu Ser Leu Asn
35 40 45

Tyr Arg Asp Pro Ile Thr Thr Gln Asn Gly Ser Leu Thr Leu Lys Leu
50 55 60

Gly Asn Gly Leu Thr Leu Asn Asn Gln Gly Gln Leu Thr Ser Thr Ala
65 70 75 80

Gly Glu Val Glu Pro Pro Leu Thr Asn Ala Asn Asn Lys Leu Ala Leu
85 90 95

Ala Tyr Ser Glu Pro Leu Ala Val Lys Ser Asn Arg Leu Thr Leu Ser
100 105 110

His Thr Ala Pro Leu Val Ile Ala Asn Asn Ser Leu Ala Leu Gln Val
115 120 125

Ser Glu Pro Ile Phe Val Asn Asp Asp Asp Lys Leu Ala Leu Gln Thr
130 135 140

Ala Ala Pro Leu Val Thr Asn Ala Gly Thr Leu Arg Leu Gln Ser Ala
145 150 155 160

Ala Pro Leu Gly Leu Val Glu Asn Thr Leu Lys Leu Leu Phe Ser Lys
165 170 175

```
Pro Leu Tyr Leu Gln Asn Asp Phe Leu Ala Leu Ala Ile Glu Arg Pro
            180                 185                 190

Leu Ala Val Ala Ala Ala Gly Thr Leu Thr Leu Gln Leu Thr Pro Pro
        195                 200                 205

Leu Lys Thr Asn Asp Asp Gly Leu Thr Leu Ser Thr Val Glu Pro Leu
    210                 215                 220

Thr Val Lys Asn Gly Asn Leu Gly Leu Gln Ile Ser Arg Pro Leu Val
225                 230                 235                 240

Val Gln Asn Asn Gly Leu Ser Leu Ala Ile Thr Pro Pro Leu Arg Leu
                245                 250                 255

Phe Asn Ser Asp Pro Val Leu Gly Leu Gly Phe Thr Phe Pro Leu Ala
            260                 265                 270

Val Thr Asn Asn Leu Leu Ser Leu Asn Met Gly Asp Gly Val Lys Leu
        275                 280                 285

Thr Tyr Asn Lys Leu Thr Ala Asn Leu Gly Arg Asp Leu Gln Phe Glu
    290                 295                 300

Asn Gly Ala Ile Ala Val Thr Leu Thr Ala Glu Leu Pro Leu Gln Tyr
305                 310                 315                 320

Thr Asn Lys Leu Gln Leu Asn Ile Gly Ala Gly Leu Arg Tyr Asn Gly
                325                 330                 335

Ala Ser Arg Lys Leu Asp Val Asn Ile Asn Gln Asn Lys Gly Leu Thr
            340                 345                 350

Trp Asp Asn Asp Ala Val Ile Pro Lys Leu Gly Ser Gly Leu Gln Phe
        355                 360                 365

Asp Pro Asn Gly Asn Ile Ala Val Ile Pro Glu Thr Val Lys Pro Gln
    370                 375                 380

Thr Leu Trp Thr Thr Ala Asp Pro Ser Pro Asn Cys Ser Val Tyr Gln
385                 390                 395                 400

Asp Leu Asp Ala Arg Leu Trp Leu Ala Leu Val Lys Ser Gly Asp Met
```

405 410 415

Val His Gly Ser Ile Ala Leu Lys Ala Leu Lys Gly Thr Leu Leu Asn
420 425 430

Pro Thr Ala Ser Tyr Ile Ser Ile Val Ile Tyr Phe Tyr Ser Asn Gly
435 440 445

Val Arg Arg Thr Asn Tyr Pro Thr Phe Asp Asn Glu Gly Thr Leu Ala
450 455 460

Asn Ser Ala Thr Trp Gly Tyr Arg Gln Gly Gln Ser Ala Asn Thr Asn
465 470 475 480

Val Thr Asn Ala Thr Glu Phe Met Pro Ser Ser Ser Arg Tyr Pro Val
485 490 495

Asn Lys Gly Asp Asn Ile Gln Asn Gln Ser Phe Ser Tyr Thr Cys Ile
500 505 510

Lys Gly Asp Phe Ala Met Pro Val Pro Phe Arg Val Thr Tyr Asn His
515 520 525

Ala Leu Glu Gly Tyr Ser Leu Lys Phe Thr Trp Arg Val Val Ala Asn
530 535 540

Gln Ala Phe Asp Ile Pro Cys Cys Ser Phe Ser Tyr Ile Thr Glu
545 550 555

<210> 34
<211> 34115
<212> DNA
<213> simian adenovirus SV-39

<220>
<221> CDS
<222> (13448)..(14959)
<223> L2 Penton

<220>
<221> CDS
<222> (17785)..(20538)
<223> L3 Hexon

<220>

<221> CDS
<222> (29515)..(31116)
<223> L5 Fiber #1

<400> 34
catcatcaat ataacaccgc aagatggcga ccgagttaac atgcaaatga ggtgggcgga 60

gttacgcgac ctttgtcttg ggaacgcgga agtgggcgcg gcgggtttcg gggaggagcg 120

cggggcgggg cgggcgtgtc gcgcggcggt gacgcgccgg ggacccggaa attgagtagt 180

ttttattcat tttgcaagtt tttctgtaca ttttggcgcg aaaactgaaa cgaggaagtg 240

aaaagtgaaa aatgccgagg tagtcaccgg gtggagatct gacctttgcc gtgtggagtt 300

tacccgctga cgtgtgggtt tcggtctcta ttttttcact gtggttttcc gggtacggtc 360

aaaggtcccc attttatgac tccacgtcag ctgatcgcta gggtatttaa tgcgcctcag 420

accgtcaaga ggccactctt gagtgccggc gagaagagtt ttctcctccg cgttccgcca 480

actgtgaaaa aatgaggaac ttcttgctat ctccggggct gccagcgacc gtagccgccg 540

agctgttgga ggacattgtt accggagctc tgggagacga tcctcaggtg atttctcact 600

tttgtgaaga ttttagtctt catgatctct atgatattga tccgggtgtt gaggggcaag 660

aggatgaatg gctggagtct gtggatgggt tttttccgga cgctatgctg ctagaggctg 720

atttgccacc acctcacaac tctcacactg agcccgagtc agctgctatt cctgaattgt 780

catcaggtga acttgacttg gcttgttacg agactatgcc tccggagtcg gatgaggagg 840

acagcgggat cagcgatccc acggctttta tggtctctaa ggcgattgct atactaaaag 900

aagatgatga tggcgatgat ggatttcgac tggacgctcc ggcggtgccg gggagagact 960

gtaagtcctg tgaataccac cgggatcgta ccggagaccc gtctatgttg tgttctctgt 1020

gttatctccg tcttaacgct gcttttgtct acagtaagtg ttttgtgctt ttttaccctg 1080

tggctttgtt gagtttattt ttttctgtgt ctcatagggt gttgtttatt ataggtcctg 1140

tttcagatgt ggaggaacct gatagtacta ctggaaatga ggaggaaaag ccctccccgc 1200

cgaaactaac tcagcgctgc agacctaata ttttgagacc ctcggcccag cgtgtgtcat 1260

cccggaaacg tgctgctgtt aattgcatag aagatttatt ggaagagccc actgaacctt 1320

tggacttgtc cttaaagcga ccccgcccgc agtagggcgc ggtgccagtt ttttctctct 1380

agcttccggg tgactcagtg caataaaaat tttcttggca acaggtgtat gtgtttactt 1440

tacgggcggg aagggattag gggagtataa agctggaggg gaaaaatctg aggctgtcag 1500

atcgagtgag aagttccatg gacttgtacg agagcctaga gaatctaagt tctttgcgac 1560

gtttgctgga ggaggcctcc gacagaacct cttacatttg gaggtttctg ttcggttccc 1620

ctctgagtcg ctttttgcac cgggtgaagc gagagcacct gacggaattt gatgggcttt 1680

tagagcagct gcctggactg tttgattctt tgaatctcgg ccaccggacg ctgctagagg 1740

agaggctttt tccacaattg gacttttcct ctccaggccg tctgtgttca gcgcttgctt 1800

ttgctgtaca tctgttggac agatggaacg agcagacgca gctcagcccg ggttacactc 1860

tggacttcct gacgctatgc ctatggaagt tcggaatcag gaggggggagg aagctgtacg 1920

ggcgcttggt ggagaggcat ccgtctctgc gccagcagcg tctgcaagct caagtgctgc 1980

tgaggcggga ggatctggaa gccatttcgg aggaggagag cggcatggaa gagaagaatc 2040

cgagagcggg gctggaccct ccggcggagg agtaggggggg ataccggacc cttttcctga 2100

gttggctttg ggggcggtgg ggggcgcttc tgtggtacgt gaggatgaag aggggcgcca 2160

acgcggtcag aagagggagc attttgagtc ctcgactttc ttggctgatg taaccgtggc 2220

cctgatggcg aaaaacaggc tggaggtggt gtggtacccg gaagtatggg aggactttga 2280

gaaggggggac ttgcacctgc tggaaaaata taactttgag caggtgaaaa catactggat 2340

gaacccggat gaggactggg aggtggtttt gaaccgatac ggcaaggtag ctctgcgtcc 2400

cgactgtcgc taccaggttc gcgacaaggt ggtcctgcga cgcaacgtgt acctgttggg 2460

caacggcgcc accgtggaga tggtggaccc cagaaggggt ggttttgtgg ccaatatgca 2520

agaaatgtgc cctggggtgg tgggcttgtc tggggtgact tttcatagtg tgaggtttag 2580

cggtagcaat tttgggggtg tggttattac cgcgaacact cctgtggtcc tgcataattg 2640

ctacttttt ggcttcagca acacctgtgt ggaaatgagg gtgggaggca aagtgcgcgg 2700

gtgttccttt tacgcttgct ggaaggggt ggtgagccag ggtaaggcta aagtgtctgt 2760

tcacaagtgt atgttggaga gatgcacctt gggcatttcc agtgagggct tcctccacgc 2820

cagcgacaac gtggcttctg acaacggctg cgcctttctt atcaagggag ggggtcgcat 2880

ctgtcacaac atgatatgcg gccctgggga tgtcccccca aagccttacc agatggttac 2940

ctgcacagat ggcaaggtgc gcatgctcaa gcctgtgcac attgtgggcc accggcgcca 3000

ccgctggcca gagtttgaac acaatgtgat gacccgctgt agcttgtacc tgggaggcag 3060

gcgaggagtt ttcttgccca gacagtgtaa cctggcccac tgcaacgtga tcatggaaca 3120

atccgccgct acccaggttt gctttggagg aatatttgat ataagcatgg tggtgtataa 3180

gatcctgcgc tacgacgact gtcgggctcg tactcgaacc tgcgactgcg gagcctctca 3240

cctgtgtaac ctgactgtga tggggatggt gactgaggag gtgcgactgg accactgtca 3300

gcactcttgc ctgcgggagg agttttcttc ctcggacgag gaggactagg taggtggttg 3360

gggcgtggcc agcgagaggg tgggctataa aggggaggtg tcggctgacg ctgtcttctg 3420

tttttcaggt accatgagcg gatcaagcag ccagaccgcg ctgagcttcg acggggccgt 3480

gtacagcccc tttctgacgg ggcgcttgcc tgcctgggcc ggagtgcgtc agaatgttac 3540

cggttcgacc gtggacggac gtcccgtgga tccatctaac gctgcttcta tgcgctacgc 3600

tactatcagc acatctactc tggacagcgc cgctgccgcc gcagccgcca cctcagccgc 3660

tctctccgcc gccaagatca tggctattaa cccaagcctt tacagccctg tatccgtgga 3720

cacctcagcc ctggagcttt accggcgaga tctagctcaa gtggtggacc aactcgcagc 3780

cgtgagccaa cagttgcagc tggtgtcgac ccgagtggag caactttccc gccctcccca 3840

gtaaccgcaa aaattcaata aacagaattt aataaacagc acttgagaaa agtttaaact 3900

tgtggttgac tttattcctg gatagctggg gggagggaac ggcgggaacg gtaagacctg 3960

gtccatcgtt cccggtcgtt gagaacacgg tggatttttt ccaagacccg atagaggtgg 4020

gtctgaacgt tgagatacat gggcatgagc ccgtctcggg ggtggaggta ggcccactgc 4080

agggcctcgt tttcaggggt ggtgttgtaa atgatccagt cgtaggcccc ccgctgggcg 4140

tggtgctgga agatgtcctt cagcagcaag ctgatggcaa cgggaagacc cttggtgtag 4200

gtgttgacaa agcggttgag ttgggagggg tgcatgcggg gactgatgag gtgcattttg 4260

gcctggatct tgaggttggc tatgttgccg cccagatcgc gcctgggatt catgttatgc 4320

aagaccacca gcaccgagta accggtgcag cgggggaatt tgtcgtgcag cttggaaggg 4380

aaagcgtgga agaatttgga gacccctcgg tgcccgccta ggttttccat gcactcatcc 4440

atgatgatgg cgatgggccc ccgggaggca gcctgggcaa aaacgttgcg ggggtccgtg 4500

acatcgtagt tgtggtcctg ggtgagttca tcataggaca ttttgacaaa gcgcgggcag 4560

agggtcccag actggggaat gatggttcca tccggtccgg gggcgtagtt gccctcgcag 4620

atttgcattt cccaggcttt gatttcagag ggagggatca tgtcaacctg gggggcgatg 4680

aaaaaaatgg tctctggggc gggggtgatg agctgggtgg aaagcaggtt gcgcaagagc 4740

tgtgacttgc cgcagccggt gggcccgtag atgacagcta tgacgggttg cagggtgtag 4800

tttagagagc tacaactgcc atcatccttc aaaagcgggg ccacactgtt taaaagttct 4860

ctaacatgta agttttcccg cactaagtcc tgcaggagac gtgaccctcc tagggagaga 4920

agttcaggaa gcgaagcaaa gtttttaagt ggcttgaggc catcggccaa gggcaagttc 4980

ctgagagttt gactgagcag ttccagccgg tcccagagct cggttacgtg ctctacggca 5040

tctcgatcca gcagacctcc tcgtttcggg ggttggggcg gctctggctg tagggaatga 5100

ggcggtgggc gtccagctgg gccatggtgc ggtccctcca tgggcgcagg gttctcttca 5160

gggtggtctc ggtcacggtg aatgggtggg ccccgggctg ggcgctggcc agggtgcgct 5220

tgaggctgag gcggctggtg gcgaaccgtt gcttttcgtc tccctgcaag tcagccaaat 5280

agcaacggac catgagctca tagtccaggc tctctgcggc atgtcctttg gcgcgaagct 5340

tgcctttgga aacgtgcccg cagtttgagc agagcaagca ttttagcgcg tagagttttg 5400

gcgccaagaa cacggattcc ggggaataag catccccacc gcagttggag caaacggttt 5460

cgcattccac cagccaggtc agctgaggat cttttgggtc aaaaaccaag cgcccgccgt 5520

ttttttgat gcgcttccta cctcgggtct ccatgaggcg gtgcccgcgt tcggtgacga 5580

agaggctgtc ggtgtctccg tagacggagg tcagggcgcg ctcctccagg ggggtcccgc 5640

ggtcctcggc gtagagaaac tcgcaccact ctgacataaa cgcccgggtc caggctagga 5700

cgaatgaggc gatgtgggaa gggtaccggt cgttatcgat gagggggtcg gttttttcca 5760

aggtgtgcag gcacatgtcc ccctcgtccg cttccaaaaa tgtgattggc ttgtaggtgt 5820

aagtcacgtg atcctgtcct tccgcggggg tataaaaggg ggcgtttccc ccctcctcgt 5880

cactctcttc cggttcgctg tcgccaaagg ccagctgttg gggtacgtaa acgcgggtga 5940

aggcgggcat gacctgtgcg ctgaggttgt cagtttctat atacgaggaa gatttgatgg 6000

cgagcgcccc cgtggagatg cccttgaggt gctcggggcc catttggtca gaaaacacaa 6060

tctgtcggtt atcaagcttg gtggcaaaag acccgtagag ggcgttggag agcaacttgg 6120

cgatggagcg ctgggtttgg ttttttttccc ggtcggcttt ttccttggcc gcgatgttga 6180

gctggacgta ctccctggcc acgcacttcc agccgggaaa aacggccgtg cgctcgtccg 6240

gcaccagcct cacgctccat ccgcggttgt gcagggtgat gacgtcgatg ctggtggcca 6300

cctctccgcg caggggctcg ttggtccagc agaggcgacc gcccttgcga gagcagaagg 6360

ggggcagggg gtcaagcagg cgctcgtccg gggggtcggc gtcgatggta aagatggcgg 6420

gcagcaggtg tttgtcaaag taatcgatct gatgcccggg gcaacgcagg gcggtttccc 6480

agtcccgcac cgccaaggcg cgctcgtatg gactgagggg ggcgccccag ggcatgggat 6540

gcgtcagggc cgaggcgtac atgccgcaga tgtcatagac gtaaaggggc tcctccagga 6600

cgccgaggta ggtggggtag cagcgccccc cgcggatgct ggcccgtacg tagtcgtaga 6660

gctcgtgcga gggggccaga aggtggcggc tgaggtgagc gcgctggggc ttttcatctc 6720

ggaagaggat ctgcctgaag atggcgtggg agttggagga gatggtgggc cgctgaaaaa 6780

tgttgaagcg ggcgtcgggc agacccacgg cctcgccgat aaagtgggcg taggactctt 6840

gcagcttttc caccagggag gcggtgacca gcacgtccag agcgcagtag tccagggttt 6900

cccgcacgat gtcataatgc tcttcctttt tttccttcca gaggtctcgg ttgaagagat 6960

actcttcgcg gtctttccag tactcttgga gaggaaaccc gttttcgtct ccacggtaag 7020

agcccaacat gtaaaactgg ttgacggcct gatagggaca gcatcccttc tccacgggca 7080

gcgagtaggc cagggcggcc ttgcgcaggg aggtgtgagt cagggcaaag gtgtcgcgga 7140

ccataacttt tacaaactgg tacttaaagt cccggtcgtc gcacatgcct cgctcccagt 7200

ctgagtagtc tgtgcgcttt ttgtgcttgg ggttaggcag ggagtaggtg acgtcgttaa 7260

agaggatttt gccacatctg ggcataaagt tgcgagagat tctgaagggg ccgggcacct 7320

ccgagcggtt gttgatgact tgggcagcca ggagaatttc gtcgaagccg ttgatgttgt 7380

gccccacgac gtagaactct atgaaacgcg gagcgccgcg cagcaggggg cacttttcaa 7440

gttgctggaa agtaagttcc cgcggctcga cgccgtgttc cgtgcggctc cagtcctcca 7500

ccgggtttcg ctccacaaaa tcctgccaga tgtggtcgac tagcaagagc tgcagtcggt 7560

cgcgaaattc gcggaatttt ctgccgatgg cttgcttctg ggggttcaag caaaaaaagg 7620

tgtctgcgtg gtcgcgccag gcgtcccagc cgagctcgcg agccagattc agggccagca 7680

gcaccagagc cggctcaccg gtgattttca tgacgaggag aaagggcacc agctgttttc 7740

cgaacgcgcc catccaggtg taggtctcca cgtcgtaggt gagaaacaga cgttcggtcc 7800

gcgggtgcga tcccaggggg aaaaacttga tgggctgcca ccattgggag ctctgggcgt 7860

ggatgtgatg gaagtaaaag tcccggcggc gcgtggaaca ttcgtgctgg tttttgtaaa 7920

agcggccgca gtggtcgcag cgcgagacgg agtgaaggct gtgaatcagg tgaatcttgc 7980

gtcgctgagg gggccccaga gccaaaaagc ggagcgggaa cgaccgcgcg gccacttcgg 8040

cgtccgcagg caagatggat gagggttcca ccgttccccg cccgcggacc gaccagactt 8100

ccgccagctg cggcttcagt tcttgcacca gctctcgcag cgtttcgtcg ctgggcgaat 8160

cgtgaatacg gaagttgtcg ggtagaggcg ggaggcggtg gacttccagg aggtgtgtga 8220

gggccggcag gagatgcagg tggtacttga tttcccacgg atgacggtcg cgggcgtcca 8280

aggcgaagag atgaccgtgg ggccgcggcg ccaccagcgt tccgcggggg gtctttatcg 8340

gcggcgggga cgggctcccg gcggcagcgg cggctcggga cccgcgggca agtcgggcag 8400

cggcacgtcg gcgtggagct cgggcagggg ctggtgctgc gcgcggagct gactggcaaa 8460

ggctatcacc cggcgattga cgtcctggat ccggcggcgc tgcgtgaaga ccaccggacc 8520

cgtggtcttg aacctgaaag agagttcgac agaatcaatc tcggcatcgt taaccgcggc 8580

ctggcgcagg atttcggcca cgtccccgga gttgtcttga tacgcgattt ctgccatgaa 8640

ctggtcgatt tcctcttcct gcaagtctcc gtgaccggcg cgttcgacgg tggccgcgag 8700

atcgttggag atgcggccca tgagctggga aaaggcattg atgccgacct cgttccacac 8760

tcggctgtac accacctctc cgtgaacgtc gcgggcgcgc atcaccacct gggcgagatt 8820

gagttccacg tggcgggcga aaaccggata gtttcggagg cgctgataca gatagttgag 8880

ggtggtggcg gcgtgctcgg ccacaaaaaa atacatgatc cagcggcgga gggtcagctc 8940

gttgatgtcg cccagcgcct ccaggcgttc catggcctcg taaaagtcca cggcaaagtt 9000

gaaaaattgg ctgttcctgg ccgagaccgt gagctcttct tccaagagcc gaatgagatc 9060

cgccacggtg gccctgactt cgcgttcgaa agccccgggt gcctcctcca cctcttcctc 9120

ctcgacttct tcgaccgctt cgggcacctc ctcttcctcg accaccacct caggcggggc 9180

tcggcggcgc cggcggcgga cgggcaggcg gtcgacgaaa cgctcgatca tttccccccc 9240

ccgtcgacgc atggtctcgg tgacggcgcg accctgttcg cgaggacgca gggtgaaggc 9300

gccgccgccg agcggaggta acagggagat cgggggggcgg tcgtgggggga gactgacggc 9360

gctaactatg catctgatca atgtttgcgt agtgacctcg ggtcggagcg agctcagcgc 9420

ttgaaaatcc acgggatcgg aaaaccgttc caggaacgcg tctagccaat cacagtcgca 9480

aggtaagctg aggaccgtct cgggggcttg tctgttctgt cttcccgcgg tggtgctgct 9540

gatgaggtag ttgaagtagg cgctcttgag gcggcggatg gtggacagga gaaccacgtc 9600

tttgcgccca gcttgctgta tccgcaggcg gtcggccatg ccccacactt ctccttgaca 9660

gcggcggagg tccttgtagt attcttgcat cagcctttcc acgggcacct cgtcttcttc 9720

ttccgctcgg ccggacgaga gccgcgtcag gccgtacccg cgctgcccct gtggttggag 9780

cagggccagg tcggccacga cgcgctcggc cagcacggcc tgctggatgc gggtgagggt 9840

gtcctgaaag tcgtcgagat ccacaaagcg gtggtacgcg ccagtgttga tggtgtaggt 9900

gcagttgctc atgacggacc agtttacggt ctgggtgcca tggcccacgg tttccaggta 9960

gcggagacgc gagtaggccc gcgtctcgaa gatgtagtcg ttgcaggtcc gcagcaggta 10020

ctggtagccc accagcagat gcggcggcgg ctggcggtag aggggccacc gctgggtggc 10080

ggggggcgttg ggggcgagat cttccaacat gaggcggtga tagccgtaga tgtagcgcga 10140

catccaagtg atgccgctgg ccgtggtgct ggcgcgggcg tagtcgcgaa cgcggttcca 10200

gatgtttcgc agcggctgga agtactcgat ggtggggcga ctctgccccg tgaggcgggc 10260

gcagtcggcg atgctctacg gggaaaaaga agggccagtg aacaaccgcc ttccgtagcc 10320

ggaggagaac gcaagggggt caaagaccac cgaggctcgg gttcgaaacc cgggtggcgg 10380

cccgaatacg gagggcggtt ttttgctttt ttctcagatg catcccgtgc tgcggcagat 10440

gcgtccgaac gcggggtccc agtccccggc ggtgcctgcg gccgtgacgg cggcttctac 10500

ggccacgtcg cgctccaccc cgcctaccac ggcccaggcg gcggtggctc tgcgcggcgc 10560

aggggaaccc gaagcagagg cggtgttgga cgtggaggag ggccaggggt tggctcggct 10620

gggggccctg agtcccgagc ggcacccgcg cgtggctctg aagcgcgacg cggcggaggc 10680

gtacgtgccg cggagcaatc tgtttcgcga ccgcagcggc gaggaggccg aggagatgcg 10740

agacttgcgt tttcgggcgg ggagggagtt gcgtcacggg ctggaccggc agagggttct 10800

gagagaggag gactttgagg cggacgagcg cacgggggtg agtcccgcgc gggctcacgt 10860

ggcggccgcc aacctggtga gcgcgtacga gcagacggtc aaggaggaga tgaacttcca 10920

gaagagcttc aatcatcacg tgcgcacgct gattgcgcgc gaagaggtgg ccatcggcct 10980

catgcatctg tgggattttg tggaggcgta cgttcagaac cccagcagca agccgctgac 11040

ggctcagctg ttcctcatcg tgcaacatag tcgagacaac gaaacgttca gggaggccat 11100

gctgaacatt gcagagcctg aggggcgctg gctcttggat ctcattaaca tcttgcagag 11160

tatcgtagtg caggagcgct cgctgagcct ggccgacaag gtggctgcca tcaactacag 11220

catgctgtcg ctgggcaaat tttacgcccg caagatctac aagtctccgt tcgtccccat 11280

agacaaggag gtgaagatag acagctttta catgcgcatg gcgctcaagg tgctgactct 11340

aagcgacgac ctgggggtgt accgcaacga ccgcatacac aaggcggtga gcgccagccg 11400

ccggcgcgag ctgagcgacc gcgagctttt gcacagcctg catcgggcgt tgactggtgc 11460

cggcagcgcc gaggcggccg agtactttga cgccggagcg gacttgcgct ggcagccatc 11520

ccgacgcgcg ctggaggcgg ctggcgtcgg ggagtacggg gtcgaggacg acgatgaagc 11580

ggacgacgag ttgggcattg acttgtagcc gtttttcgtt agatatgtcg gcgaacgagc 11640

cgtctgcggc cgccatggtg acggcggcgg gcgcgcccca ggacccggcc acgcgcgcgg 11700

cgctgcagag tcagccttcc ggagtgacgc ccgcggacga ctggtccgag gccatgcgtc 11760

gcatcctggc gctgacggcg cgcaaccccg aggcttttcg gcagcagccg caggcaaacc 11820

```
ggtttgcggc cattttggaa gcggtggtgc cctccagacc caaccccacc cacgaaaagg  11880

tgctggccat cgtcaacgcc ctggcggaga ccaaggccat ccgcccagac gaggccgggc  11940

aggtttacaa cgcgctgcta gaaagggtgg gacgctacaa cagctccaac gtgcagacca  12000

atctggaccg cttggtgacg gacgtgaagg aggccgtagc ccagcgagag cggtttttca  12060

aggaagccaa tctgggctcg ctggtggccc tcaacgcctt cctgagcacg ctgccggcga  12120

acgtgccccg cggtcaggag gactacgtga actttctgag cgccctccgc ctgatggtgg  12180

ccgaggtgcc gcagagcgag gtgtaccagt ctggccccaa ctactacttc cagacctccc  12240

ggcagggcct gcagacggta aacctgacgc aggcctttca gaacctgcag ggcctttggg  12300

gggtgcgcgc tccgctgggc gaccgcagca cggtgtccag cctgctgacc cccaatgccc  12360

ggctgctctt gcttctcatt gctccgttca ccgacagcgg ttccatcagc cgcgactctt  12420

acctgggaca cctgctcacc ctgtaccggg aggccatcgg gcaggcgcgg gtggacgagc  12480

agacgtacca ggaaatcacc agcgtgagcc gcgcgctggg gcaggaggac acgggcagct  12540

tggaggcgac tctgaacttc ctgctgacca accggcggca gcgcctacct ccccagtacg  12600

cgctgaacgc ggaggaggag cgcatcctgc gtttcgtgca gcagagcacc gcgctgtact  12660

tgatgcggga aggcgcctct cccagcgctt cgctggacat gacggcggcc aacatggagc  12720

catcgttcta cgccgccaac cgtcccttcg tcaaccggct aatggactat ttgcatcggg  12780

cggcggccct gaacccggaa tactttacta acgtcatcct gaacgaccgt tggctgccac  12840

ctcccggctt ctacacgggg gagttcgacc tcccggaggc caacgacggt ttcatgtggg  12900

acgacgtgga cagcgtgttc ctgcccggca agaaggaggc gggtgactct cagagccacc  12960

gcgcgagcct cgcagacctg ggggcgaccg ggcccgcgtc tccgctgcct cgcctgccga  13020

gcgccagcag cgccagcgtg gggcgggtga gccgtccgcg cctcagcggt gaggaggact  13080

ggtggaacga tccgctgctc cgtccggccc gcaacaaaaa cttccccaac aacgggatag  13140

aggatttggt agacaaaatg aaccgttgga agacgtatgc ccaggagcat cgggagtggc  13200

aggcgaggca acccatgggc cctgttctgc cgccctctcg gcgcccgcgc agggacgaag  13260

acgccgacga ttcagccgat gacagcagcg tgttggatct gggcgggagc gggaacccct  13320

ttgcccacct gcaacctcgc ggcgtgggtc ggcggtggcg ctaggaaaaa aaattattaa  13380

aagcacttac cagagccatg gtaagaagag caacaaaggt gtgtcctgct ttcttcccgg  13440

tagcaaa atg cgt cgg gcg gtg gca gtt ccc tcc gcg gca atg gcg tta      13489
       Met Arg Arg Ala Val Ala Val Pro Ser Ala Ala Met Ala Leu
```

```
        1               5                  10

ggc ccg ccc cct tct tac gaa agc gtg atg gca gcg gcc acc ctg caa    13537
Gly Pro Pro Pro Ser Tyr Glu Ser Val Met Ala Ala Ala Thr Leu Gln
15                  20                  25                  30

gcg ccg ttg gag aat cct tac gtg ccg ccg cga tac ctg gag cct acg    13585
Ala Pro Leu Glu Asn Pro Tyr Val Pro Pro Arg Tyr Leu Glu Pro Thr
                35                  40                  45

ggc ggg aga aac agc att cgt tac tcg gag ctg acg ccc ctg tac gac    13633
Gly Gly Arg Asn Ser Ile Arg Tyr Ser Glu Leu Thr Pro Leu Tyr Asp
            50                  55                  60

acc acc cgc ctg tac ctg gtg gac aac aag tca gca gat atc gcc acc    13681
Thr Thr Arg Leu Tyr Leu Val Asp Asn Lys Ser Ala Asp Ile Ala Thr
        65                  70                  75

ttg aac tac cag aac gac cac agc aac ttt ctc acg tcc gtg gtg cag    13729
Leu Asn Tyr Gln Asn Asp His Ser Asn Phe Leu Thr Ser Val Val Gln
    80                  85                  90

aac agc gac tac acg ccc gcc gaa gcg agc acg cag acc att aac ttg    13777
Asn Ser Asp Tyr Thr Pro Ala Glu Ala Ser Thr Gln Thr Ile Asn Leu
95                  100                 105                 110

gac gac cgc tcg cgc tgg ggc ggg gac ttg aaa acc att ctg cac act    13825
Asp Asp Arg Ser Arg Trp Gly Gly Asp Leu Lys Thr Ile Leu His Thr
                115                 120                 125

aac atg ccc aac gtg aac gag ttc atg ttt acc aac tcg ttc agg gct    13873
Asn Met Pro Asn Val Asn Glu Phe Met Phe Thr Asn Ser Phe Arg Ala
            130                 135                 140

aaa ctt atg gtg gcg cac gag gcc gac aag gac ccg gtt tat gag tgg    13921
Lys Leu Met Val Ala His Glu Ala Asp Lys Asp Pro Val Tyr Glu Trp
        145                 150                 155

gtg cag ctg acg ctg ccg gag ggg aac ttt tca gag att atg acc ata    13969
Val Gln Leu Thr Leu Pro Glu Gly Asn Phe Ser Glu Ile Met Thr Ile
    160                 165                 170

gac ctg atg aac aac gcc att atc gac cac tac ctg gcg gta gcc aga    14017
Asp Leu Met Asn Asn Ala Ile Ile Asp His Tyr Leu Ala Val Ala Arg
175                 180                 185                 190

cag cag ggg gtg aaa gaa agc gag atc ggc gtc aag ttt gac acg cgc    14065
Gln Gln Gly Val Lys Glu Ser Glu Ile Gly Val Lys Phe Asp Thr Arg
                195                 200                 205

aac ttt cgt ctg ggc tgg gac ccg gag acg ggg ctt gtg atg ccg ggg    14113
Asn Phe Arg Leu Gly Trp Asp Pro Glu Thr Gly Leu Val Met Pro Gly
            210                 215                 220

gtg tac acg aac gaa gct ttc cat ccc gac gtg gtc ctc ttg ccg ggc    14161
Val Tyr Thr Asn Glu Ala Phe His Pro Asp Val Val Leu Leu Pro Gly
        225                 230                 235
```

```
tgc ggg gtg gac ttt acc tac agc cgg tta aac aac ctg cta ggc ata    14209
Cys Gly Val Asp Phe Thr Tyr Ser Arg Leu Asn Asn Leu Leu Gly Ile
    240                 245                 250

cgc aag aga atg ccc ttt cag gaa ggg ttt cag atc ctg tac gag gac    14257
Arg Lys Arg Met Pro Phe Gln Glu Gly Phe Gln Ile Leu Tyr Glu Asp
255                 260                 265                 270

ctg gag ggc ggt aac atc ccg gcc ctg ctg gac gtg ccg gcg tac gag    14305
Leu Glu Gly Gly Asn Ile Pro Ala Leu Leu Asp Val Pro Ala Tyr Glu
                275                 280                 285

gag agc atc gcc aac gca agg gag gcg gcg atc agg ggc gat aat ttc    14353
Glu Ser Ile Ala Asn Ala Arg Glu Ala Ala Ile Arg Gly Asp Asn Phe
            290                 295                 300

gcg gcg cag ccc cag gcg gct cca acc ata aaa ccc gtt ttg gaa gac    14401
Ala Ala Gln Pro Gln Ala Ala Pro Thr Ile Lys Pro Val Leu Glu Asp
        305                 310                 315

tcc aaa ggg cgg agc tac aac gta ata gcc aac acc aac aac acg gct    14449
Ser Lys Gly Arg Ser Tyr Asn Val Ile Ala Asn Thr Asn Asn Thr Ala
    320                 325                 330

tac agg agc tgg tat ctg gct tat aac tac ggc gac ccg gag aag ggg    14497
Tyr Arg Ser Trp Tyr Leu Ala Tyr Asn Tyr Gly Asp Pro Glu Lys Gly
335                 340                 345                 350

gtt agg gcc tgg acc ctg ctc acc act ccg gac gtg acg tgc ggt tca    14545
Val Arg Ala Trp Thr Leu Leu Thr Thr Pro Asp Val Thr Cys Gly Ser
                355                 360                 365

gag cag gtc tac tgg tcg ctg cct gac atg tac gtg gac cct gtg acg    14593
Glu Gln Val Tyr Trp Ser Leu Pro Asp Met Tyr Val Asp Pro Val Thr
            370                 375                 380

ttt cgc tcc acg cag caa gtt agc aac tac cca gtg gtg gga gcg gag    14641
Phe Arg Ser Thr Gln Gln Val Ser Asn Tyr Pro Val Val Gly Ala Glu
        385                 390                 395

ctt atg ccg att cac agc aag agc ttt tac aac gag cag gcc gtc tac    14689
Leu Met Pro Ile His Ser Lys Ser Phe Tyr Asn Glu Gln Ala Val Tyr
    400                 405                 410

tca cag ctc att cgt cag acc acc gcc cta acg cac gtt ttc aac cgc    14737
Ser Gln Leu Ile Arg Gln Thr Thr Ala Leu Thr His Val Phe Asn Arg
415                 420                 425                 430

ttc ccc gag aac caa atc cta gtg cga cct cca gcg ccc acc atc acc    14785
Phe Pro Glu Asn Gln Ile Leu Val Arg Pro Pro Ala Pro Thr Ile Thr
                435                 440                 445

acc gtc agc gag aac gtg ccc gct cta acc gat cac ggg acg ctg cct    14833
Thr Val Ser Glu Asn Val Pro Ala Leu Thr Asp His Gly Thr Leu Pro
            450                 455                 460
```

```
ttg cag aac agc atc cgc gga gtt cag cga gtt acc atc acg gac gcc    14881
Leu Gln Asn Ser Ile Arg Gly Val Gln Arg Val Thr Ile Thr Asp Ala
        465                 470                 475

cgt cgt cgg acc tgt ccc tac gtc tac aaa gcc ttg gga atc gtg gcc    14929
Arg Arg Arg Thr Cys Pro Tyr Val Tyr Lys Ala Leu Gly Ile Val Ala
    480                 485                 490

ccg cgc gtc ctg tcg agt cgc act ttc tag atgtccatcc tcatctctcc      14979
Pro Arg Val Leu Ser Ser Arg Thr Phe
495                 500

cagcaacaat accggttggg gtctgggcgt gaccaaaatg tacggaggcg ccaaacgacg  15039

gtccccacaa catcccgtgc gagtgcgcgg gcactttaga gccccatggg ggtcgcacac  15099

gcgcgggcgc accggccgaa ccaccgtcga cgacgtgatc gatagcgtgg tggccgacgc  15159

ccgcaactac cagcccgctc gatccacggt ggacgaagtc atcgacggcg tggtggccga  15219

cgccagggcc tacgcccgca gaaagtctcg tctgcgccgc cgccgttcgc taaagcgccc  15279

cacggccgcc atgaaagccg ctcgctctct gctgcgtcgc gcacgtatcg tgggtcgccg  15339

cgccgccaga cgcgcagccg ccaacgccgc cgccggccga gtgcgccgcc gggccgccca  15399

gcaggccgcc gccgccatct ccagtctatc cgcccccccga cgcgggaatg tgtactgggt  15459

cagggactcg gccaccggcg tgcgagttcc cgtgagaacc cgtcctcctc gtccctgaat  15519

aaaaagttct aagcccaatc ggtgttccgt tgtgtgttca gctcgtcatg accaaacgca  15579

agtttaaaga ggagctgctg caagcgctgg tccccgaaat ctatgcgccg gcgccggacg  15639

tgaaaccgcg tcgcgtgaaa cgcgtgaaga agcaggaaaa gctagagaca aaagaggagg  15699

cggtggcgtt gggagacggg gaggtggagt ttgtgcgctc gttcgcgccg cgtcggcgag  15759

tgaattggaa ggggcgcaag gtgcaacggg tgctgcgtcc cggcacggtg gtgtctttca  15819

ccccgggtga aaaatccgcc tggaagggca taaagcgcgt gtacgatgag gtgtacgggg  15879

acgaagacat tctggagcag gcgctggata gaagcgggga gtttgcttac ggcaagaggg  15939

cgaggacggg cgagatcgcc atcccgctgg acacttccaa ccccacccca agtctgaaac  15999

ccgtgacgct gcaacaggtg ttgccggtga gcgccccctc gcgacgcggc ataaaacgcg  16059

agggcggcga gctgcagccc accatgcagc tcctggttcc caagaggcag aaactagagg  16119

acgtactgga catgataaaa atggagcccg acgtgcagcc cgatattaaa atccgtccca  16179

tcaaagaagt ggcgccggga atgggcgtgc agaccgtgga catccagatt cccatgacca  16239

gcgccgcaca ggcggtagag gccatgcaga ccgacgtggg gatgatgacg gacctgcccg  16299

cagctgctgc cgccgtggcc agcgccgcga cgcaaacgga agccggcatg cagaccgacc  16359
```

```
cgtggacgga ggcgcccgtg cagccggcca gaagacgcgt cagacggacg tacggccccg  16419

tttctggcat aatgccggag tacgcgctgc atccttccat catccccacc cccggctacc  16479

gggggcgcac ctaccgtccg cgacgcagca ccactcgccg ccgtcgccgc acggcacgag  16539

tcgccaccgc cagagtgaga cgcgtaacga cacgtcgcgg ccgccgcttg accctgcccg  16599

tggtgcgcta ccatcccagc attctttaaa aaaccgctcc tacgttgcag atgggcaagc  16659

ttacttgtcg actccgtatg gccgtgcccg gctaccgagg aagatcccgc cgacgacgga  16719

ctttgggagg cagcggtttg cgccgccgtc gggcggttca ccggcgcctc aagggaggca  16779

ttctgccggc cctgatcccc ataatcgccg cagccatcgg ggccattccc ggaatcgcca  16839

gcgtagcggt gcaggctagc cagcgccact gattttacta accctgtcgg tcgcgccgtc  16899

tctttcggca gactcaacgc ccagcatgga agacatcaat ttctcctctc tggccccgcg  16959

gcacggcacg cggccgtata tggggacgtg gagcgagatc ggcacgaacc agatgaacgg  17019

gggcgctttc aattggagcg gtgtgtggag cggcttgaaa aatttcggtt ccactctgaa  17079

aacttacggc aaccgggtgt ggaactccag cacggggcag atgctgaggg acaagctaaa  17139

ggacacgcag tttcagcaaa aggtggtgga cggcatcgct tcgggcctca acggcgccgt  17199

cgacctggcc aaccaggcca ttcaaaagga aattaacagc cgcctggagc cgcggccgca  17259

ggtggaggag aacctgcccc ctctggaggc gctgcccccc aagggagaga agcgcccgcg  17319

gcccgacatg gaggagacgc tagttactaa gagcgaggag ccgccatcat acgaggaggc  17379

ggtgggtagc tcgcagctgc cgtccctcac gctgaagccc accacctatc ccatgaccaa  17439

gcccatcgcc tccatggcgc gccccgtggg agtcgacccg cccatcgacg cggtggccac  17499

tttggacctg ccgcgccccg aacccggcaa ccgcgtgcct cccgtcccca tcgctccgcc  17559

ggtttctcgc cccgccatcc gccccgtcgc cgtggccact ccccgctatc cgagccgcaa  17619

cgccaactgg cagaccaccc tcaacagtat tgtcggactg ggggtgaagt ctctgaagcg  17679

ccgtcgctgt ttttaaagca caatttatta aacgagtagc cctgtcttaa tccatcgttg  17739

tatgtgtgcc tatatcacgc gttcagagcc tgaccgtccg tcaag atg gcc act ccg  17796
                                                Met Ala Thr Pro
                                                    505

tcg atg atg ccg cag tgg tcg tac atg cac atc gcc ggg cag gac gcc    17844
Ser Met Met Pro Gln Trp Ser Tyr Met His Ile Ala Gly Gln Asp Ala
        510                 515                 520

tcg gag tac ctg agc ccg ggt ctg gtg cag ttt gcc cgt gcg acg gaa    17892
Ser Glu Tyr Leu Ser Pro Gly Leu Val Gln Phe Ala Arg Ala Thr Glu
```

```
    525                 530                 535

acc tac ttc tca ctg ggc aac aag ttc agg aac ccc acc gtg gcg ccc    17940
Thr Tyr Phe Ser Leu Gly Asn Lys Phe Arg Asn Pro Thr Val Ala Pro
540                 545                 550                 555

acc cac gac gtc acc acc gat cgg tcc cag cga ctg aca atc cgc ttc    17988
Thr His Asp Val Thr Thr Asp Arg Ser Gln Arg Leu Thr Ile Arg Phe
                560                 565                 570

gtc ccc gtg gac aag gaa gac acc gct tac tcc tac aaa acc cgc ttc    18036
Val Pro Val Asp Lys Glu Asp Thr Ala Tyr Ser Tyr Lys Thr Arg Phe
            575                 580                 585

acg ctg gcc gtg ggc gac aac cgg gtg cta gac atg gcc agt acc tac    18084
Thr Leu Ala Val Gly Asp Asn Arg Val Leu Asp Met Ala Ser Thr Tyr
        590                 595                 600

ttt gac atc cgc ggc gtg atc gac cgc gga cct agc ttc aag cct tac    18132
Phe Asp Ile Arg Gly Val Ile Asp Arg Gly Pro Ser Phe Lys Pro Tyr
    605                 610                 615

tcc ggc acg gct tac aac tca ctg gct ccc aaa ggg gcg ccc aac aac    18180
Ser Gly Thr Ala Tyr Asn Ser Leu Ala Pro Lys Gly Ala Pro Asn Asn
620                 625                 630                 635

agc caa tgg aac gcc aca gat aac ggg aac aag cca gtg tgt ttt gct    18228
Ser Gln Trp Asn Ala Thr Asp Asn Gly Asn Lys Pro Val Cys Phe Ala
                640                 645                 650

cag gca gct ttt ata ggt caa agc att aca aaa gac gga gtg caa ata    18276
Gln Ala Ala Phe Ile Gly Gln Ser Ile Thr Lys Asp Gly Val Gln Ile
            655                 660                 665

cag aac tca gaa aat caa cag gct gct gcc gac aaa act tac caa cca    18324
Gln Asn Ser Glu Asn Gln Gln Ala Ala Ala Asp Lys Thr Tyr Gln Pro
        670                 675                 680

gag cct caa att gga gtt tcc acc tgg gat acc aac gtt acc agt aac    18372
Glu Pro Gln Ile Gly Val Ser Thr Trp Asp Thr Asn Val Thr Ser Asn
    685                 690                 695

gct gcc gga cga gtg tta aaa gcc acc act ccc atg ctg cca tgt tac    18420
Ala Ala Gly Arg Val Leu Lys Ala Thr Thr Pro Met Leu Pro Cys Tyr
700                 705                 710                 715

ggt tca tat gcc aat ccc act aat cca aac ggg ggt cag gca aaa aca    18468
Gly Ser Tyr Ala Asn Pro Thr Asn Pro Asn Gly Gly Gln Ala Lys Thr
                720                 725                 730

gaa gga gac att tcg cta aac ttt ttc aca aca act gcg gca gca gac    18516
Glu Gly Asp Ile Ser Leu Asn Phe Phe Thr Thr Thr Ala Ala Ala Asp
            735                 740                 745

aat aat ccc aaa gtg gtt ctt tac agc gaa gat gta aac ctt caa gcc    18564
Asn Asn Pro Lys Val Val Leu Tyr Ser Glu Asp Val Asn Leu Gln Ala
        750                 755                 760
```

```
ccc gat act cac tta gta tat aag cca acg gtg gga gaa aac gtt atc    18612
Pro Asp Thr His Leu Val Tyr Lys Pro Thr Val Gly Glu Asn Val Ile
    765                 770                 775

gcc gca gaa gcc ctg cta acg cag cag gcg tgt ccc aac aga gca aac    18660
Ala Ala Glu Ala Leu Leu Thr Gln Gln Ala Cys Pro Asn Arg Ala Asn
780                 785                 790                 795

tac ata ggt ttc cga gat aac ttt atc ggt tta atg tat tat aac agc    18708
Tyr Ile Gly Phe Arg Asp Asn Phe Ile Gly Leu Met Tyr Tyr Asn Ser
                800                 805                 810

aca ggg aac atg gga gtt ctg gca ggt cag gcc tcg cag tta aac gca    18756
Thr Gly Asn Met Gly Val Leu Ala Gly Gln Ala Ser Gln Leu Asn Ala
            815                 820                 825

gtt gta gac ctg caa gat cga aac acg gaa ctg tcc tat cag cta atg    18804
Val Val Asp Leu Gln Asp Arg Asn Thr Glu Leu Ser Tyr Gln Leu Met
        830                 835                 840

cta gat gct ctg ggt gac aga act cga tat ttc tca atg tgg aat cag    18852
Leu Asp Ala Leu Gly Asp Arg Thr Arg Tyr Phe Ser Met Trp Asn Gln
    845                 850                 855

gcc gtg gac agc tac gat cca gac gtt agg att atc gag aac cat ggg    18900
Ala Val Asp Ser Tyr Asp Pro Asp Val Arg Ile Ile Glu Asn His Gly
860                 865                 870                 875

gtg gaa gac gag ctg ccc aat tac tgt ttt cca ctc cca ggc atg ggt    18948
Val Glu Asp Glu Leu Pro Asn Tyr Cys Phe Pro Leu Pro Gly Met Gly
                880                 885                 890

att ttt aac tcc tac aag ggg gta aaa cca caa aat ggc ggt aat ggt    18996
Ile Phe Asn Ser Tyr Lys Gly Val Lys Pro Gln Asn Gly Gly Asn Gly
            895                 900                 905

aac tgg gaa gca aac ggg gac cta tca aat gcc aat gag atc gct tta    19044
Asn Trp Glu Ala Asn Gly Asp Leu Ser Asn Ala Asn Glu Ile Ala Leu
        910                 915                 920

gga aac att ttt gcc atg gaa att aac ctc cac gca aac ctg tgg cgc    19092
Gly Asn Ile Phe Ala Met Glu Ile Asn Leu His Ala Asn Leu Trp Arg
    925                 930                 935

agc ttc ttg tac agc aat gtg gcg ctg tac ctg cca gac agc tat aaa    19140
Ser Phe Leu Tyr Ser Asn Val Ala Leu Tyr Leu Pro Asp Ser Tyr Lys
940                 945                 950                 955

ttc act ccc gct aac atc act ctg ccc gcc aac caa aac acc tac gag    19188
Phe Thr Pro Ala Asn Ile Thr Leu Pro Ala Asn Gln Asn Thr Tyr Glu
                960                 965                 970

tat atc aac ggg cgc gtc act tct cca acc ctg gtg gac acc ttt gtt    19236
Tyr Ile Asn Gly Arg Val Thr Ser Pro Thr Leu Val Asp Thr Phe Val
            975                 980                 985
```

```
aac att gga gcc cga tgg tcg ccg gat ccc atg gac aac  gtc aac ccc   19284
Asn Ile Gly Ala Arg Trp Ser Pro Asp Pro Met Asp Asn  Val Asn Pro
        990                 995                 1000

ttt aac  cat cac cgg aac gcg  ggc ctc cgt tac cgc  tcc atg ctg     19329
Phe Asn  His His Arg Asn Ala  Gly Leu Arg Tyr Arg  Ser Met Leu
    1005                1010                1015

ctg gga  aat gga cgc gtg gtg  cct ttc cac ata caa  gtg ccg caa     19374
Leu Gly  Asn Gly Arg Val Val  Pro Phe His Ile Gln  Val Pro Gln
    1020                1025                1030

aaa ttt  ttc gcg att aag aac  ctc ctg ctt ttg ccc  ggc tcc tac     19419
Lys Phe  Phe Ala Ile Lys Asn  Leu Leu Leu Leu Pro  Gly Ser Tyr
    1035                1040                1045

act tac  gag tgg agc ttc aga  aaa gac gtg aac atg  att ctg cag     19464
Thr Tyr  Glu Trp Ser Phe Arg  Lys Asp Val Asn Met  Ile Leu Gln
    1050                1055                1060

agc acc  ctg ggc aat gat ctt  cga gtg gac ggg gcc  agc gtc cgc     19509
Ser Thr  Leu Gly Asn Asp Leu  Arg Val Asp Gly Ala  Ser Val Arg
    1065                1070                1075

att gac  agc gtc aac ttg tac  gcc aac ttt ttc ccc  atg gcg cac     19554
Ile Asp  Ser Val Asn Leu Tyr  Ala Asn Phe Phe Pro  Met Ala His
    1080                1085                1090

aac acc  gct tct acc ttg gaa  gcc atg ctg cga aac  gac acc aac     19599
Asn Thr  Ala Ser Thr Leu Glu  Ala Met Leu Arg Asn  Asp Thr Asn
    1095                1100                1105

gac cag  tcg ttt aac gac tac  ctc agc gcg gcc aac  atg ctt tat     19644
Asp Gln  Ser Phe Asn Asp Tyr  Leu Ser Ala Ala Asn  Met Leu Tyr
    1110                1115                1120

ccc att  ccg gcc aac gcc acc  aac gtt ccc att tcc  att ccc tcc     19689
Pro Ile  Pro Ala Asn Ala Thr  Asn Val Pro Ile Ser  Ile Pro Ser
    1125                1130                1135

cgc aac  tgg gcg gcc ttc cgg  gga tgg agc ttc acc  cgc ctt aaa     19734
Arg Asn  Trp Ala Ala Phe Arg  Gly Trp Ser Phe Thr  Arg Leu Lys
    1140                1145                1150

gcc aag  gaa acg cct tcc ttg  ggc tcc ggc ttt gac  ccc tac ttt     19779
Ala Lys  Glu Thr Pro Ser Leu  Gly Ser Gly Phe Asp  Pro Tyr Phe
    1155                1160                1165

gtg tac  tca ggc acc att cct  tac ctg gac ggc agc  ttt tac ctc     19824
Val Tyr  Ser Gly Thr Ile Pro  Tyr Leu Asp Gly Ser  Phe Tyr Leu
    1170                1175                1180

aac cac  act ttc aaa cgt ctg  tcc atc atg ttc gat  tct tcc gta     19869
Asn His  Thr Phe Lys Arg Leu  Ser Ile Met Phe Asp  Ser Ser Val
    1185                1190                1195

agt tgg  ccg ggc aac gac cgc  ctc ctg acg ccg aac  gag ttc gaa     19914
```

```
Ser Trp  Pro Gly Asn Asp Arg  Leu Leu Thr Pro Asn  Glu Phe Glu
    1200                 1205                 1210

att aag  cgc att gtg gac ggg  gaa ggc tac aac gtg  gct caa agt     19959
Ile Lys  Arg Ile Val Asp Gly  Glu Gly Tyr Asn Val  Ala Gln Ser
    1215                 1220                 1225

aac atg  acc aaa gac tgg ttt  tta att caa atg ctc  agc cac tac     20004
Asn Met  Thr Lys Asp Trp Phe  Leu Ile Gln Met Leu  Ser His Tyr
    1230                 1235                 1240

aac atc  ggc tac caa ggc ttc  tat gtt ccc gag ggc  tac aag gat     20049
Asn Ile  Gly Tyr Gln Gly Phe  Tyr Val Pro Glu Gly  Tyr Lys Asp
    1245                 1250                 1255

cgg atg  tat tct ttc ttc cga  aac ttt cag ccc atg  agc cgc cag     20094
Arg Met  Tyr Ser Phe Phe Arg  Asn Phe Gln Pro Met  Ser Arg Gln
    1260                 1265                 1270

gtg ccg  gat ccc acc gct gcc  ggc tat caa gcc gtt  ccc ctg ccc     20139
Val Pro  Asp Pro Thr Ala Ala  Gly Tyr Gln Ala Val  Pro Leu Pro
    1275                 1280                 1285

aga caa  cac aac aac tcg ggc  ttt gtg ggg tac atg  ggc ccg acc     20184
Arg Gln  His Asn Asn Ser Gly  Phe Val Gly Tyr Met  Gly Pro Thr
    1290                 1295                 1300

atg cgc  gaa gga cag cca tac  ccg gcc aac tac ccc  tat ccc ctg     20229
Met Arg  Glu Gly Gln Pro Tyr  Pro Ala Asn Tyr Pro  Tyr Pro Leu
    1305                 1310                 1315

atc ggc  gct acc gcc gtc ccc  gcc att acc cag aaa  aag ttt ttg     20274
Ile Gly  Ala Thr Ala Val Pro  Ala Ile Thr Gln Lys  Lys Phe Leu
    1320                 1325                 1330

tgc gac  cgc gtc atg tgg cgc  ata cct ttt tcc agc  aac ttt atg     20319
Cys Asp  Arg Val Met Trp Arg  Ile Pro Phe Ser Ser  Asn Phe Met
    1335                 1340                 1345

tca atg  ggg gcc ctg acc gac  ctc gga cag aac atg  ctt tac gct     20364
Ser Met  Gly Ala Leu Thr Asp  Leu Gly Gln Asn Met  Leu Tyr Ala
    1350                 1355                 1360

aac tcc  gcc cat gcc ctg gat  atg act ttt gag gtg  gac ccc atg     20409
Asn Ser  Ala His Ala Leu Asp  Met Thr Phe Glu Val  Asp Pro Met
    1365                 1370                 1375

aac gag  ccc acg ttg ctg tac  atg ctt ttt gag gtg  ttc gac gtg     20454
Asn Glu  Pro Thr Leu Leu Tyr  Met Leu Phe Glu Val  Phe Asp Val
    1380                 1385                 1390

gtc aga  gtg cac cag ccg cac  cgc ggt att atc gag  gcc gtg tac     20499
Val Arg  Val His Gln Pro His  Arg Gly Ile Ile Glu  Ala Val Tyr
    1395                 1400                 1405

ctg cgc  acc ccc ttc tct gcg  ggc aat gcc acc aca  taa gccgctgaac  20548
Leu Arg  Thr Pro Phe Ser Ala  Gly Asn Ala Thr Thr
```

1410 1415 1420

tagctggttt ttaccccaga tcccatgggc tccacggaag acgaactgcg ggccattgtg 20608

cgagacctgg gctgcggacc ctacttcctg ggcacctttg acaagcggtt tcccgggttc 20668

gtgtctcctc gcaaactcgc gtgcgcgatc gtgaataccg ccggccgaga gaccggagga 20728

gagcattggc tagctctggg ctggaacccc cgctcgtcca cgtttttcct gttcgacccc 20788

tttggctttt cagaccaacg cttgaagcag atctatgcat ttgaatatga gggtctactc 20848

aagcgaagcg cgctggcctc ctccgccgat cactgtctaa ccctggtaaa gagcactcag 20908

acggttcagg gccctcacag cgccgcctgt ggccttttt gttgcatgtt tttgcacgcc 20968

tttgtgaact ggccggacac ccccatggaa aacaacccca ccatggacct cctgactggc 21028

gttcccaact ccatgctcca aagccccagc gtgcagacca ccctcctcca aaaccagaaa 21088

aatctgtacg cctttctgca caagcactct ccctactttc gccgccatcg ggaacaaata 21148

gaaaatgcaa ccgcgtttaa caaaactctg taacgtttaa taaatgaact ttttattgaa 21208

ctggaaaacg ggtttgtgat ttttaaaaat caaaggggtt gagctggaca tccatgtggg 21268

aggccggaag ggtggtgttc ttgtactggt acttgggcag ccacttaaac tctggaatca 21328

caaacttggg cagcggtatt tctgggaagt tgtcgtgcca cagctggcgg gtcagctgaa 21388

gtgcctgcag aacatcgggg gcggagatct tgaagtcgca gtttatctgg ttcacggcac 21448

gcgcgttgcg gtacatggga ttggcacact gaaacaccag caggctggga ttcttgatgc 21508

tagccagggc cacggcgtcg gtcacgtcac cggtgtcttc tatgttggac agcgaaaaag 21568

gcgtgacttt gcaaagctgg cgtcccgcgc gaggcacgca atctcccagg tagttgcact 21628

cacagcggat gggcagaaga agatgcttgt ggccgcgggt catgtaggga taggccgctg 21688

ccataaaagc ttcgatctgc ctgaaagcct gcttggcctt gtgcccttcg gtataaaaaa 21748

caccgcagga cttgttggaa aaggtattac tggcgcaagc ggcatcgtga aagcaagcgc 21808

gtgcgtcttc gtttcgtaac tgcaccacgc tgcggcccca ccggttctga atcaccttgg 21868

ccctgccggg gttttccttg agagcgcgct ggccggcttc gctgcccaca tccatttcca 21928

cgacatgctc cttgttaatc atggccagac cgtggaggca gcgcagctcc tcgtcatcgt 21988

cggtgcagtg atgctcccac acgacgcagc cagtgggctc ccacttgggc ttggaggcct 22048

cggcaatgcc agaatacagg agaacgtagt ggtgcagaaa acgtcccatc atggtgccaa 22108

aggttttctg gctgctgaag gtcatcgggc agtacctcca gtcctcgtta agccaagtgt 22168

tgcagatctt cctgaagacc gtgtactgat cgggcataaa gtggaactca ttgcgctcgg 22228

tcttgtcgat cttatacttt tccatcagac tatgcataat ctccatgccc ttttcccagg 22280

cgcaaacaat cttggtgcta cacgggttag gtatggccaa agtggttggc ctctgaggcg 22340

gcgcttgttc ttcctcttga gccctctccc gactgacggg ggttgaaaga gggtgccccc 22400

tggggaacgg cttgaacacg gtctggcccg aggcgtcccg aagaatctgc atcgggggat 22460

tgctggccgt catggcgatg atctgacccc ggggctcctc cacttcgtcc tcctcgggac 22520

tttcctcgtg cttttcgggg gacggtacgg gagtaggggg aagagcgcgg cgcgccttct 22580

tcttgggcgg cagttccgga gcctgctctt gacgactggc cattgtcttc tcctaggcaa 22640

gaaaaacaag atggaagact ctttctcctc ctcctcgtca acgtcagaaa gcgagtcttc 22700

caccttaagc gccgagaact cccagcgcat agaatccgat gtgggctacg agactccccc 22760

cgcgaacttt tcgccgcccc ccataaacac taacgggtgg acggactacc tggccctagg 22820

agacgtactg ctgaagcaca tcaggcggca gagcgttatc gtgcaagatg ctctcaccga 22880

gcgactcgcg gttccgctgg aagtggcgga acttagcgcc gcctacgagc gaaccctctt 22940

ctccccaaag actcccccca agaggcaggc taacggcacc tgcgagccta accctcgact 23000

caacttctac cctgcctttg ccgtgccaga ggtactggct acgtaccaca ttttttttcca 23060

aaaccacaaa atccctctct cgtgccgcgc caaccgcacc aaagccgatc gcgtgctgcg 23120

actggaggaa ggggctcgca tacctgagat tgcgtgtctg gaggaagtcc caaaaatctt 23180

tgaaggtctg ggccgcgacg aaaagcgagc agcaaacgct ctggaagaga acgcagagag 23240

tcacaacagc gccttggtag aactcgaggg cgacaacgcc agactggccg tcctcaaacg 23300

gtccatagaa gtcacgcact tcgcctaccc cgccgttaac ctccctccaa aagttatgac 23360

agcggtcatg gactcgctgc tcataaagcg cgctcagccc ttagacccag agcacgaaaa 23420

caacagtgac gaaggaaaac cggtggtttc tgatgaggag ttgagcaagt ggctgtcctc 23480

caacgacccc gccacgttgg aggaacgaag aaaaaccatg atggccgtgg tgctagttac 23540

cgtgcaatta gaatgtctgc agaggttctt ttcccaccca gagaccctga gaaaagtgga 23600

ggaaacgctg cactacacat ttaggcacgg ctacgtgaag caagcctgca agatttccaa 23660

cgtagaactt agcaacctca tctcctacct ggggatcttg cacgaaaacc gcctcggaca 23720

aaacgtgctg cacagcacac tgaaaggaga agcccgccga gactatgtgc gagactgcgt 23780

gttcctagcg ctagtgtaca cctggcagag cggaatggga gtctggcagc agtgcctgga 23840

ggacgaaaac ctcaaagagc ttgaaaagct gctggtgcgc tccagaaggg cactgtggac 23900

cagttttgac gagcgcaccg ccgcgcgaga cctagctgat attatttttc ctcccaagct 23968

ggtgcagact ctccgggaag gactgccaga ttttatgagt caaagcatct tgcaaaactt 24028

ccgctctttc atcttggaac gctcgggaat cttgcccgcc actagctgcg ccctacccac 24088

agattttgtg cctctccact accgcgaatg cccaccgccg ctgtggccgt acacttactt 24148

gcttaaactg gccaactttc taatgttcca ctctgacctg gcagaagacg ttagcggcga 24208

ggggctgcta gaatgccact gccgctgcaa cctgtgcacc ccccaccgct ctctagtatg 24268

caacactccc ctgctcaatg agacccagat catcggtacc tttgaaatcc agggaccctc 24328

cgacgcggaa aacggcaagc aggggtctgg gctaaaactc acagccggac tgtggacctc 24388

cgcctacttg cgcaaatttg taccagaaga ctatcacgcc caccaaatta aattttacga 24448

aaaccaatca aaaccaccca aaagcgagtt aacggcttgc gtcattacgc agagcagcat 24508

agttgggcag ttgcaagcca ttaacaaagc gcggcaagag tttctcctaa aaaaaggaaa 24568

aggggtctac ttggaccccc agaccggcga ggaactcaac ggaccctcct cagtcgcagg 24628

ttgtgtgccc catgccgccc aaaaagaaca cctcgcagtg gaacatgcca gagacggagg 24688

aagaggagtg gagcagtgtg agcaacagcg aaacggagga agagccgtgg cccgaggggt 24748

gcaacgggga agaggacacg gagggacggc gaagtcttcg ccgaagaact ctcgccgctg 24808

cccccgaagt cccagccggc cgcctcggcc caagatcccg cacacacccg tagatgggat 24868

agcaagacca aaaagccggg taagagaaac gctcgccccc gccagggcta ccgctcgtgg 24928

agaaagcaca aaaactgcat cttatcgtgc ttgctccagt gcggcggaga cgtttcgttc 24988

acccgtagat acttgctttt taacaaaggg gtggccgtcc cccgtaacgt cctccactac 25048

taccgtcact cttacagctc cgaagcggac ggctaagaaa acgcagcagt tgccggcggg 25108

aggactgcgt ctcagcgccc gagaaccccc agccaccagg gagctccgaa accgcatatt 25168

tcccaccctc tacgctatct ttcagcaaag ccgggggcag cagcaagaac tgaaaataaa 25228

aaaccgcacg ctgaggtcgc ttacccgaag ctgcctctat cacaagagcg aagagcagct 25288

gcagcgaacc ctggaggacg cagaagcgct gttccagaag tactgcgcga ccaccctaaa 25348

taactaaaaa agcccgcgcg cgggacttca aaccgtctga cgtcaccagc cgcgcgccaa 25408

aatgagcaaa gagattccca cgccttacat gtggagttac cagccgcaga tgggattagc 25468

cgccggcgcc gcccaggatt actccacgaa aatgaactgg ctcagcgccg ggccccacat 25528

gatttcccgc gtaaacgaca ttcgcgccca ccgcaatcag ctattgttag aacaggctgc 25588

tctgaccgcc acgccccgta ataacctgaa ccctcccagc tggccagctg ccctggtgta 25648

ccaggaaacg cctccaccca ccagcgtact tttgccccgt gacgcccagg cggaagtcca 25708

gatgactaac gcgggcgcgc aattagcggg cggatcccgg tttcggtaca gagttcacgg 25768

cgccgcaccc tatagcccag gtataaagag gctgatcatt cgaggcagag gtgtccagct 25828

caacgacgag acagtgagct cttcgcttgg tctacgacca gacggagtgt tccagctcgc 25888

gggctcgggc cgctcttcgt tcacgcctcg ccaggcatac ctgactctgc agagctctgc 25948

ctctcagcct cgctcgggag gaatcggacc ccttcagttt gtggaggagt ttgtgccctc 26008

ggtctacttt cagcctttct ccggatcgcc cggccagtac ccggacgagt tcatccccaa 26068

cttcgacgcg gtgagtgact ctgtggacgg ttatgactga tgtcgagccc gcttcagtgc 26128

tagtggaaca agcgcggctc aatcacctgg ttcgttgccg ccgccgctgc tgcgtggctc 26188

gcgacttgag cttagctctc aagtttgtaa aaaaccccgtc cgaaaccggg agcgctgtgc 26248

acgggttgga gctagtgggt cctgagaagg ccaccatcca cgttctcaga aactttgtgg 26308

aaaaacccat tttggttaaa cgagatcagg ggccttttgt aatcagctta ctctgcacct 26368

gtaaccatgt tgaccttcac gactatttta tggatcattt gtgcgctgaa ttcaataagt 26428

aaagcgaatt cttaccaaga ttatgatgtc catgactgtt cctcgccact atacgatgtt 26488

gtgccagtaa actctcttgt cgacatctat ctgaactgtt ccttttggtc cgcacagctt 26548

acttggtact acggtgacac cgtcctttct ggctcactgg gcagctcaca cggaataaca 26608

cttcacctct tttcgccgtt tcgatacgga aactacagct gtcgtgccgg tacctgcctc 26668

cacgttttca atcttcagcc ctgtccaccg accaaacttg tatttgtcga ctctaagcac 26728

ttacagctca actgcagcat tctaggcccc agtatcttgt ggacatacaa taaaatcagg 26788

ttggtggaat ttgtctacta cccacccagc gcccgcggtt ttggggaaat tcctttccag 26848

atctactaca actatcttgc cacacattat gcaagtcaac agcaactaaa cttgcaagca 26908

cccttcacgc caggagagta ctcctgtcac gtaggctcct gcacagaaac ttttattctc 26968

ttcaacagat cttctgccat tgaacgcttc actactaact actttagaaa ccaagttgtg 27028

cttttcactg acgaaacccc taacgtcacc ctggactgtg catgtttttc tcatgacacc 27088

gtaacttgga ctcttaacaa tactctctgg ctcgcgttcg ataaccaaag cttgattgtt 27148

aaaaattttg atttaacctt tactaaaccc tctcctcgcg aaatagttat ctttgctcct 27208

tttaatccaa aaactacctt agcctgtcag gttttgttta agccttgcca aacaaacttt 27268

aagtttgttt atttgcctcc gcaatctgtc aaactcatag aaaaatacaa caaagcgccc 27328

gtcttggctc ctaaaacctt ctaccactgg ctaacctaca cggggctgtt tgcactaatt 27388

gtttttttcc taattaacat ttttatatgt ttcttgcctt cctccttctt ttcgcgaaca 27448

ccgttgccgc agaaagacct ctccttatta ctgtagcgct tgctatacaa aaccaagagt 27508

ggtcaaccgt gctctcaatc tattttcaat ttttcatttt gtccttaata ctttctctta 27568

ttgtcgttaa caatgatctg gagcattggt ctcgcctttt tttggctgct tagtgcaaaa 27628

gccactattt ttcacaggta tgtggaagaa ggaactagca ccctctttac gatacctgaa 27688

acaattaagg cggctgatga agtttcttgg tacaaaggct cgctctcaga cggcaaccac 27748

tcattctcag gacagaccct ttgcatccaa gaaacttatt ttaaatcaga actacaatac 27808

agctgcataa aaaacttttt ccatctctac aacatctcaa aaccctatga gggtatttac 27868

aatgccaagg tttcagacaa ctccagcaca cggaactttt actttaatct gacagttatt 27928

aaagcaattt ccattcctat ctgtgagttt agctcccagt ttctttctga aacctactgt 27988

ttaattacta taaactgcac taaaaatcgc cttcacacca ccataatcta caatcacaca 28048

caatcacctt gggttttaaa cctaaaattt tctccacaca tgccttcgca atttctcacg 28108

caagttaccg tctctaacat aagcaagcag tttggctttt actatccttt ccacgaactg 28168

tgcgaaataa ttgaagccga atatgaacca gactacttta cttacattgc cattggtgta 28228

atcgttgttt gcctttgctt tgttattggg gggtgtgttt atttgtacat tcagagaaaa 28288

atattgctct cgctgtgctc ctgcggttac aaagcagaag aaagaattaa aatctctaca 28348

ctttattaat gttttccaga aatggcaaaa ctaacgctcc tacttttgct tctcacgccg 28408

gtgacgcttt ttaccatcac tttttctgcc gccgccacac tcgaacctca atgtttgcca 28468

ccggttgaag tctactttgt ctacgtgttg ctgtgctgcg ttagcgtttg cagtataaca 28528

tgttttacct ttgtttttct tcagtgcatt gactacttct gggtcagact ctactaccgc 28588

agacacgcgc ctcagtatca aaatcaacaa attgccagac tactcggtct gccatgattg 28648

tcttgtattt taccctgatt ttttttcacc ttacttgcgc ttgtgatttt cacttcactc 28708

aattttggaa aacgcaatgc ttcgacccgc gcctctccaa cgactggatg atggctcttg 28768

caattgccac gcttggggcg tttggacttt ttagtggttt tgctttgcat tacaaattta 28828

agactccatg gacacatggc tttctttcag attttccagt tacacctact ccgccgcctc 28888

ccccggccat cgacgtgcct caggttccct caccttctcc atctgtctgc agctactttc 28948

atctgtaatg gccgacctag aatttgacgg agtgcaatct gagcaaaggg ctatacactt 29008

ccaacgccag tcggaccgcg aacgcaaaaa cagagagctg caaaccatac aaaacaccca 29068

```
ccaatgtaaa cgcgggatat tttgtattgt aaaacaagct aagctccact acgagcttct  29128

atctggcaac gaccacgagc tccaatacgt ggtcgatcag cagcgtcaaa cctgtgtatt  29188

cttaattgga gtttccccca ttaaagttac tcaaaccaag ggtgaaacca agggaaccat  29248

aaggtgctca tgtcacctgt cagaatgcct ttacactcta gttaaaaccc tatgtggctt  29308

acatgattct atccccttta attaaataaa cttactttaa atctgcaatc acttcttcgt  29368

ccttgttttt gtcgccatcc agcagcacca ccttcccctc ttcccaactt tcatagcata  29428

ttttccgaaa agaggcgtac tttcgccaca ccttaaaggg aacgtttact tcgctttcaa  29488

gctctcccac gattttcatt gcagat atg aaa cgc gcc aaa  gtg gaa gaa gga  29541
                           Met Lys Arg Ala Lys  Val Glu Glu Gly
                                               1425

ttt  aac ccc gtt tat ccc  tat gga tat tct act  ccg act gac gtg     29586
Phe  Asn Pro Val Tyr Pro  Tyr Gly Tyr Ser Thr  Pro Thr Asp Val
1430                 1435                 1440

gct  cct ccc ttt gta gcc  tct gac ggt ctt caa  gaa aac cca cct     29631
Ala  Pro Pro Phe Val Ala  Ser Asp Gly Leu Gln  Glu Asn Pro Pro
1445                 1450                 1455

ggg  gtc ttg tcc cta aaa  ata tcc aaa cct tta  act ttt aat gcc     29676
Gly  Val Leu Ser Leu Lys  Ile Ser Lys Pro Leu  Thr Phe Asn Ala
1460                 1465                 1470

tcc  aag gct cta agc ctg  gct att ggt cca gga  tta aaa att caa     29721
Ser  Lys Ala Leu Ser Leu  Ala Ile Gly Pro Gly  Leu Lys Ile Gln
1475                 1480                 1485

gat  ggt aaa cta gtg ggg  gag gga caa gca att  ctt gca aac ctg     29766
Asp  Gly Lys Leu Val Gly  Glu Gly Gln Ala Ile  Leu Ala Asn Leu
1490                 1495                 1500

ccg  ctt caa atc acc aac  aac aca att tca cta  cgt ttt ggg aac     29811
Pro  Leu Gln Ile Thr Asn  Asn Thr Ile Ser Leu  Arg Phe Gly Asn
1505                 1510                 1515

aca  ctt gcc ttg aat gac  aat aat gaa ctc caa  acc aca cta aaa     29856
Thr  Leu Ala Leu Asn Asp  Asn Asn Glu Leu Gln  Thr Thr Leu Lys
1520                 1525                 1530

tct  tca tcg ccc ctt aaa  atc aca gac cag act  ctg tcc ctt aac     29901
Ser  Ser Ser Pro Leu Lys  Ile Thr Asp Gln Thr  Leu Ser Leu Asn
1535                 1540                 1545

ata  ggg gac agc ctt gca  att aaa gat gac aaa  cta gaa agc gct     29946
Ile  Gly Asp Ser Leu Ala  Ile Lys Asp Asp Lys  Leu Glu Ser Ala
1550                 1555                 1560

ctt  caa gcg acc ctc cca  ctc tcc att agc aac  aac acc atc agc     29991
Leu  Gln Ala Thr Leu Pro  Leu Ser Ile Ser Asn  Asn Thr Ile Ser
```

```
1565                1570                1575

ctc  aac gtg ggc acc gga  ctc acc ata aat gga  aac gtt tta caa     30036
Leu  Asn Val Gly Thr Gly  Leu Thr Ile Asn Gly  Asn Val Leu Gln
1580                1585                1590

gct  gtt ccc tta aat gct  cta agt ccc cta act  att tcc aac aat     30081
Ala  Val Pro Leu Asn Ala  Leu Ser Pro Leu Thr  Ile Ser Asn Asn
1595                1600                1605

aac  atc agc ctg cgc tat  ggc agt tcc ctg acg  gtg ctt aac aat     30126
Asn  Ile Ser Leu Arg Tyr  Gly Ser Ser Leu Thr  Val Leu Asn Asn
1610                1615                1620

gaa  ctg caa agc aac ctc  aca gtt cac tcc cct  tta aaa ctc aac     30171
Glu  Leu Gln Ser Asn Leu  Thr Val His Ser Pro  Leu Lys Leu Asn
1625                1630                1635

tcc  aac aac tca att tct  ctc aac act cta tct  ccg ttt aga atc     30216
Ser  Asn Asn Ser Ile Ser  Leu Asn Thr Leu Ser  Pro Phe Arg Ile
1640                1645                1650

gag  aat ggt ttc ctc acg  ctc tat ttg gga aca  aaa tct ggc ttg     30261
Glu  Asn Gly Phe Leu Thr  Leu Tyr Leu Gly Thr  Lys Ser Gly Leu
1655                1660                1665

cta  gtt caa aac agt ggc  tta aaa gtt caa gcg  ggc tac ggc ctg     30306
Leu  Val Gln Asn Ser Gly  Leu Lys Val Gln Ala  Gly Tyr Gly Leu
1670                1675                1680

caa  gta aca gac acc aat  gct ctc aca tta aga  tat ctc gct cca     30351
Gln  Val Thr Asp Thr Asn  Ala Leu Thr Leu Arg  Tyr Leu Ala Pro
1685                1690                1695

ctg  acc att cca gac tcg  ggc tca gaa caa ggc  att ctt aaa gta     30396
Leu  Thr Ile Pro Asp Ser  Gly Ser Glu Gln Gly  Ile Leu Lys Val
1700                1705                1710

aac  act gga cag ggc cta  agt gtg aac caa gct  gga gcg ctt gaa     30441
Asn  Thr Gly Gln Gly Leu  Ser Val Asn Gln Ala  Gly Ala Leu Glu
1715                1720                1725

aca  tcc cta gga ggt gga  tta aaa tat gct gat  aac aaa ata acc     30486
Thr  Ser Leu Gly Gly Gly  Leu Lys Tyr Ala Asp  Asn Lys Ile Thr
1730                1735                1740

ttt  gat aca gga aac gga  ctg aca tta tct gaa  aat aaa ctt gca     30531
Phe  Asp Thr Gly Asn Gly  Leu Thr Leu Ser Glu  Asn Lys Leu Ala
1745                1750                1755

gta  gct gca ggt agt ggt  cta act ttt aga gat  ggt gcc ttg gta     30576
Val  Ala Ala Gly Ser Gly  Leu Thr Phe Arg Asp  Gly Ala Leu Val
1760                1765                1770

gcc  acg gga acc gca ttt  acg caa aca ctg tgg  act acg gct gat     30621
Ala  Thr Gly Thr Ala Phe  Thr Gln Thr Leu Trp  Thr Thr Ala Asp
1775                1780                1785
```

```
ccg  tct ccc aac tgc aca  att ata cag gac cgc  gac aca aaa ttt     30666
Pro  Ser Pro Asn Cys Thr  Ile Ile Gln Asp Arg  Asp Thr Lys Phe
1790                1795                1800

act  ttg gcg ctt acc att  agt ggg agc caa gtg  ctg ggg acg gtt     30711
Thr  Leu Ala Leu Thr Ile  Ser Gly Ser Gln Val  Leu Gly Thr Val
1805                1810                1815

tcc  att att gga gta aaa  ggc ccc ctt tca agt  agc ata ccg tca     30756
Ser  Ile Ile Gly Val Lys  Gly Pro Leu Ser Ser  Ser Ile Pro Ser
1820                1825                1830

gct  acc gtt aca gta caa  ctt aac ttt gat tcc  aac gga gcc cta     30801
Ala  Thr Val Thr Val Gln  Leu Asn Phe Asp Ser  Asn Gly Ala Leu
1835                1840                1845

ttg  agc tcc tct tca ctt  aaa ggt tac tgg ggg  tat cgc caa ggt     30846
Leu  Ser Ser Ser Ser Leu  Lys Gly Tyr Trp Gly  Tyr Arg Gln Gly
1850                1855                1860

ccc  tca att gac cct tac  ccc ata att aat gcc  tta aac ttt atg     30891
Pro  Ser Ile Asp Pro Tyr  Pro Ile Ile Asn Ala  Leu Asn Phe Met
1865                1870                1875

cca  aac tca ctg gct tat  ccc ccg gga caa gaa  atc caa gca aaa     30936
Pro  Asn Ser Leu Ala Tyr  Pro Pro Gly Gln Glu  Ile Gln Ala Lys
1880                1885                1890

tgt  aac atg tac gtt tct  act ttt tta cga gga  aat cca caa aga     30981
Cys  Asn Met Tyr Val Ser  Thr Phe Leu Arg Gly  Asn Pro Gln Arg
1895                1900                1905

cca  ata gtt tta aac atc  act ttt aat aat caa  acc agc ggg ttt     31026
Pro  Ile Val Leu Asn Ile  Thr Phe Asn Asn Gln  Thr Ser Gly Phe
1910                1915                1920

tcc  att aga ttt aca tgg  aca aat tta acc aca  gga gaa gca ttt     31071
Ser  Ile Arg Phe Thr Trp  Thr Asn Leu Thr Thr  Gly Glu Ala Phe
1925                1930                1935

gca  atg ccc cca tgc act  ttt tcc tac att gct  gaa caa caa taa     31116
Ala  Met Pro Pro Cys Thr  Phe Ser Tyr Ile Ala  Glu Gln Gln
1940                1945                1950

actatgtaac cctcaccgtt aacccgcctc cgcccttcca ttttatttta taaaccaccc  31176

gatccacctt ttcagcagta aacaattgca tgtcagtagg ggcagtaaaa cttttgggag  31236

ttaaaatcca cacaggttct tcacaagcta agcgaaaatc agttacactt ataaaaccat  31296

cgctaacatc ggacaaagac aagcatgagt ccaaagcttc cggttctgga tcagattttt  31356

gttcattaac agcgggagaa acagcttctg gaggattttc catctccatc tccttcatca  31416

gttccaccat gtccaccgtg gtcatctggg acgagaacga cagttgtcat acacctcata  31476
```

agtcaccggt cgatgacgaa cgtacagatc tcgaagaatg tcctgtcgcc gcctttcggc 31536

agcactgggc cgaaggcgaa agcgcccatg tttaacaatg gccagcaccg cccgcttcat 31596

caggcgccta gttcttttag cgcaacagcg catgcgcagc tcgctaagac tggcgcaaga 31656

aacacagcac agaaccacca gattgttcat gatcccataa gcgtgctgac accagcccat 31716

actaacaaat tgtttcacta ttctagcatg aatgtcatat ctgatgttca agtaaattaa 31776

atggcgcccc cttatgtaaa cacttcccac gtacaacacc tcctttggca tctgataatt 31836

aaccacctcc cgataccaaa tacatctctg attaatagtc gccccgtaca ctacccgatt 31896

aaaccaagtt gccaacataa tccccccctgc catacactgc aaagaacctg gacggctaca 31956

atgacagtgc aaagtccaca cctcgttgcc atggataact gaggaacgcc ttaagtcaat 32016

agtggcacaa ctaatacaaa catgtaaata gtgtttcaac aagtgccact cgtatgaggt 32076

gagtatcatg tcccagggaa cgggccactc cataaacact gcaaaaccaa cacatcctac 32136

catccccccgc acggcactca catcgtgcat ggtgttcata tcacagtccg gaagctgagg 32196

acaaggaaaa gtctcgggag cattttcata gggcggtagt gggtactcct tgtaggggtt 32256

cagtcggcac cggtatctcc tcaccttctg ggccataaca cacaagttga gatctgattt 32316

caaggtactt tctgaatgaa aaccaagtgc tttcccaaca atgtatccga tgtcttcggt 32376

ccccgcgtcg gtagcgctcc ttgcagtaca cacggaacaa ccactcacgc aggcccagaa 32436

gacagttttc cgcggacggt gacaagttaa tccccctcag tctcagagcc aatatagttt 32496

cttccacagt agcataggcc aaacccaacc aggaaacaca agctggcacg tcccgttcaa 32556

cgggaggaca aggaagcaga ggcagaggca taggcaaagc aacagaattt ttattccaac 32616

tggtcacgta gcacttcaaa caccaggtca cgtaaatggc agcgatcttg ggtttcctga 32676

tggaacataa cagcaagatc aaacatgaga cgattctcaa ggtgattaac cacagctgga 32736

attaaatcct ccacgcgcac atttagaaac accagcaata caaaagcccg gttttctccg 32796

ggatctatca tagcagcaca gtcatcaatt agtcccaagt aattttcccg tttccaatct 32856

gttataattt gcagaataat gccctgtaaa tccaagccgg ccatggcgaa aagctcagat 32916

aatgcacttt ccacgtgcat tcgtaaacac accctcatct tgtcaatcca aaaagtcttc 32976

ttcttgagaa acctgtagta aattaagaat cgccaggtta ggctcgatgc ctacatcccg 33036

gagcttcatt ctcagcatgc actgcaaatg atccagcaga tcagaacagc aattagcagc 33096

cagctcatcc ccggtttcca gttccggagt tcccacggca attatcactc gaaacgtggg 33156

acaaatcgaa ataacatgag ctcccacgtg agcaaaagcc gtagggccag tgcaataatc 33216

acagaaccag cggaaaaaag attgcagctc atgtttcaaa aagctctgca gatcaaaatt 33276

cagctcatgc aaataacaca gtaaagtttg cggtatagta accgaaaacc acacgggtcg 33336

acgttcaaac atctcggctt acctaaaaaa gaagcacatt tttaaaccac agtcgcttcc 33396

tgaacaggag gaaatatggt gcggcgtaaa accagacgcg ccaccggatc tccggcagag 33456

ccctgataat acagccagct gtggttaaac agcaaaacct ttaattcggc aacggttgag 33516

gtctccacat aatcagcgcc cacaaaaatc ccatctcgaa cttgctcgcg tagggagcta 33576

aaatggccag tatagcccca tggcacccga acgctaatct gcaagtatat gagagccacc 33636

ccattcggcg ggatcacaaa atcagtcgga gaaaacaacg tatacacccc ggactgcaaa 33696

agctgttcag gcaaacgccc ctgcggtccc tctcggtaca ccagcaaagc ctcgggtaaa 33756

gcagccatgc caagcgctta ccgtgccaag agcgactcag acgaaaaagt gtactgaggc 33816

gctcagagca gcggctatat actctacctg tgacgtcaag aaccgaaagt caaaagttca 33876

cccggcgcgc ccgaaaaaac ccgcgaaaat ccacccaaaa agcccgcgaa aaacacttcc 33936

gtataaaatt tccgggttac cggcgcgtca ccgccgcgcg acacgcccgc cccgccccgc 33996

gctcctcccc gaaacccgcc gcgcccactt ccgcgttccc aagacaaagg tcgcgtaact 34056

ccgcccacct catttgcatg ttaactcggt cgccatcttg cggtgttata ttgatgatg 34115

<210> 35
<211> 503
<212> PRT
<213> simian adenovirus SV-39

<400> 35

Met Arg Arg Ala Val Ala Val Pro Ser Ala Ala Met Ala Leu Gly Pro
1 5 10 15

Pro Pro Ser Tyr Glu Ser Val Met Ala Ala Ala Thr Leu Gln Ala Pro
20 25 30

Leu Glu Asn Pro Tyr Val Pro Pro Arg Tyr Leu Glu Pro Thr Gly Gly
35 40 45

Arg Asn Ser Ile Arg Tyr Ser Glu Leu Thr Pro Leu Tyr Asp Thr Thr
50 55 60

Arg Leu Tyr Leu Val Asp Asn Lys Ser Ala Asp Ile Ala Thr Leu Asn
65 70 75 80

```
Tyr Gln Asn Asp His Ser Asn Phe Leu Thr Ser Val Val Gln Asn Ser
                85                  90                  95

Asp Tyr Thr Pro Ala Glu Ala Ser Thr Gln Thr Ile Asn Leu Asp Asp
            100                 105                 110

Arg Ser Arg Trp Gly Gly Asp Leu Lys Thr Ile Leu His Thr Asn Met
        115                 120                 125

Pro Asn Val Asn Glu Phe Met Phe Thr Asn Ser Phe Arg Ala Lys Leu
    130                 135                 140

Met Val Ala His Glu Ala Asp Lys Asp Pro Val Tyr Glu Trp Val Gln
145                 150                 155                 160

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

Met Asn Asn Ala Ile Ile Asp His Tyr Leu Ala Val Ala Arg Gln Gln
            180                 185                 190

Gly Val Lys Glu Ser Glu Ile Gly Val Lys Phe Asp Thr Arg Asn Phe
        195                 200                 205

Arg Leu Gly Trp Asp Pro Glu Thr Gly Leu Val Met Pro Gly Val Tyr
    210                 215                 220

Thr Asn Glu Ala Phe His Pro Asp Val Val Leu Leu Pro Gly Cys Gly
225                 230                 235                 240

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

Arg Met Pro Phe Gln Glu Gly Phe Gln Ile Leu Tyr Glu Asp Leu Glu
            260                 265                 270

Gly Gly Asn Ile Pro Ala Leu Leu Asp Val Pro Ala Tyr Glu Glu Ser
        275                 280                 285

Ile Ala Asn Ala Arg Glu Ala Ala Ile Arg Gly Asp Asn Phe Ala Ala
    290                 295                 300
```

Gln Pro Gln Ala Ala Pro Thr Ile Lys Pro Val Leu Glu Asp Ser Lys
305 310 315 320

Gly Arg Ser Tyr Asn Val Ile Ala Asn Thr Asn Asn Thr Ala Tyr Arg
325 330 335

Ser Trp Tyr Leu Ala Tyr Asn Tyr Gly Asp Pro Glu Lys Gly Val Arg
340 345 350

Ala Trp Thr Leu Leu Thr Thr Pro Asp Val Thr Cys Gly Ser Glu Gln
355 360 365

Val Tyr Trp Ser Leu Pro Asp Met Tyr Val Asp Pro Val Thr Phe Arg
370 375 380

Ser Thr Gln Gln Val Ser Asn Tyr Pro Val Val Gly Ala Glu Leu Met
385 390 395 400

Pro Ile His Ser Lys Ser Phe Tyr Asn Glu Gln Ala Val Tyr Ser Gln
405 410 415

Leu Ile Arg Gln Thr Thr Ala Leu Thr His Val Phe Asn Arg Phe Pro
420 425 430

Glu Asn Gln Ile Leu Val Arg Pro Pro Ala Pro Thr Ile Thr Thr Val
435 440 445

Ser Glu Asn Val Pro Ala Leu Thr Asp His Gly Thr Leu Pro Leu Gln
450 455 460

Asn Ser Ile Arg Gly Val Gln Arg Val Thr Ile Thr Asp Ala Arg Arg
465 470 475 480

Arg Thr Cys Pro Tyr Val Tyr Lys Ala Leu Gly Ile Val Ala Pro Arg
485 490 495

Val Leu Ser Ser Arg Thr Phe
500

<210> 36
<211> 917
<212> PRT
<213> simian adenovirus SV-39

<400> 36

```
Met Ala Thr Pro Ser Met Met Pro Gln Trp Ser Tyr Met His Ile Ala
1               5                  10                  15

Gly Gln Asp Ala Ser Glu Tyr Leu Ser Pro Gly Leu Val Gln Phe Ala
            20                  25                  30

Arg Ala Thr Glu Thr Tyr Phe Ser Leu Gly Asn Lys Phe Arg Asn Pro
        35                  40                  45

Thr Val Ala Pro Thr His Asp Val Thr Thr Asp Arg Ser Gln Arg Leu
    50                  55                  60

Thr Ile Arg Phe Val Pro Val Asp Lys Glu Asp Thr Ala Tyr Ser Tyr
65                  70                  75                  80

Lys Thr Arg Phe Thr Leu Ala Val Gly Asp Asn Arg Val Leu Asp Met
                85                  90                  95

Ala Ser Thr Tyr Phe Asp Ile Arg Gly Val Ile Asp Arg Gly Pro Ser
            100                 105                 110

Phe Lys Pro Tyr Ser Gly Thr Ala Tyr Asn Ser Leu Ala Pro Lys Gly
        115                 120                 125

Ala Pro Asn Asn Ser Gln Trp Asn Ala Thr Asp Asn Gly Asn Lys Pro
    130                 135                 140

Val Cys Phe Ala Gln Ala Ala Phe Ile Gly Gln Ser Ile Thr Lys Asp
145                 150                 155                 160

Gly Val Gln Ile Gln Asn Ser Glu Asn Gln Gln Ala Ala Ala Asp Lys
                165                 170                 175

Thr Tyr Gln Pro Glu Pro Gln Ile Gly Val Ser Thr Trp Asp Thr Asn
            180                 185                 190

Val Thr Ser Asn Ala Ala Gly Arg Val Leu Lys Ala Thr Thr Pro Met
        195                 200                 205

Leu Pro Cys Tyr Gly Ser Tyr Ala Asn Pro Thr Asn Pro Asn Gly Gly
    210                 215                 220
```

```
Gln Ala Lys Thr Glu Gly Asp Ile Ser Leu Asn Phe Phe Thr Thr Thr
225                 230                 235                 240

Ala Ala Ala Asp Asn Asn Pro Lys Val Val Leu Tyr Ser Glu Asp Val
                245                 250                 255

Asn Leu Gln Ala Pro Asp Thr His Leu Val Tyr Lys Pro Thr Val Gly
            260                 265                 270

Glu Asn Val Ile Ala Ala Glu Ala Leu Leu Thr Gln Gln Ala Cys Pro
        275                 280                 285

Asn Arg Ala Asn Tyr Ile Gly Phe Arg Asp Asn Phe Ile Gly Leu Met
    290                 295                 300

Tyr Tyr Asn Ser Thr Gly Asn Met Gly Val Leu Ala Gly Gln Ala Ser
305                 310                 315                 320

Gln Leu Asn Ala Val Val Asp Leu Gln Asp Arg Asn Thr Glu Leu Ser
                325                 330                 335

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

Met Trp Asn Gln Ala Val Asp Ser Tyr Asp Pro Asp Val Arg Ile Ile
        355                 360                 365

Glu Asn His Gly Val Glu Asp Glu Leu Pro Asn Tyr Cys Phe Pro Leu
    370                 375                 380

Pro Gly Met Gly Ile Phe Asn Ser Tyr Lys Gly Val Lys Pro Gln Asn
385                 390                 395                 400

Gly Gly Asn Gly Asn Trp Glu Ala Asn Gly Asp Leu Ser Asn Ala Asn
                405                 410                 415

Glu Ile Ala Leu Gly Asn Ile Phe Ala Met Glu Ile Asn Leu His Ala
            420                 425                 430

Asn Leu Trp Arg Ser Phe Leu Tyr Ser Asn Val Ala Leu Tyr Leu Pro
        435                 440                 445
```

```
Asp Ser Tyr Lys Phe Thr Pro Ala Asn Ile Thr Leu Pro Ala Asn Gln
    450                 455                 460

Asn Thr Tyr Glu Tyr Ile Asn Gly Arg Val Thr Ser Pro Thr Leu Val
465                 470                 475                 480

Asp Thr Phe Val Asn Ile Gly Ala Arg Trp Ser Pro Asp Pro Met Asp
                485                 490                 495

Asn Val Asn Pro Phe Asn His His Arg Asn Ala Gly Leu Arg Tyr Arg
            500                 505                 510

Ser Met Leu Leu Gly Asn Gly Arg Val Val Pro Phe His Ile Gln Val
        515                 520                 525

Pro Gln Lys Phe Phe Ala Ile Lys Asn Leu Leu Leu Leu Pro Gly Ser
    530                 535                 540

Tyr Thr Tyr Glu Trp Ser Phe Arg Lys Asp Val Asn Met Ile Leu Gln
545                 550                 555                 560

Ser Thr Leu Gly Asn Asp Leu Arg Val Asp Gly Ala Ser Val Arg Ile
                565                 570                 575

Asp Ser Val Asn Leu Tyr Ala Asn Phe Phe Pro Met Ala His Asn Thr
            580                 585                 590

Ala Ser Thr Leu Glu Ala Met Leu Arg Asn Asp Thr Asn Asp Gln Ser
        595                 600                 605

Phe Asn Asp Tyr Leu Ser Ala Ala Asn Met Leu Tyr Pro Ile Pro Ala
    610                 615                 620

Asn Ala Thr Asn Val Pro Ile Ser Ile Pro Ser Arg Asn Trp Ala Ala
625                 630                 635                 640

Phe Arg Gly Trp Ser Phe Thr Arg Leu Lys Ala Lys Glu Thr Pro Ser
                645                 650                 655

Leu Gly Ser Gly Phe Asp Pro Tyr Phe Val Tyr Ser Gly Thr Ile Pro
            660                 665                 670
```

```
Tyr Leu Asp Gly Ser Phe Tyr Leu Asn His Thr Phe Lys Arg Leu Ser
        675                 680                 685

Ile Met Phe Asp Ser Ser Val Ser Trp Pro Gly Asn Asp Arg Leu Leu
    690                 695                 700

Thr Pro Asn Glu Phe Glu Ile Lys Arg Ile Val Asp Gly Glu Gly Tyr
705                 710                 715                 720

Asn Val Ala Gln Ser Asn Met Thr Lys Asp Trp Phe Leu Ile Gln Met
                725                 730                 735

Leu Ser His Tyr Asn Ile Gly Tyr Gln Gly Phe Tyr Val Pro Glu Gly
            740                 745                 750

Tyr Lys Asp Arg Met Tyr Ser Phe Phe Arg Asn Phe Gln Pro Met Ser
        755                 760                 765

Arg Gln Val Pro Asp Pro Thr Ala Ala Gly Tyr Gln Ala Val Pro Leu
    770                 775                 780

Pro Arg Gln His Asn Asn Ser Gly Phe Val Gly Tyr Met Gly Pro Thr
785                 790                 795                 800

Met Arg Glu Gly Gln Pro Tyr Pro Ala Asn Tyr Pro Tyr Pro Leu Ile
                805                 810                 815

Gly Ala Thr Ala Val Pro Ala Ile Thr Gln Lys Lys Phe Leu Cys Asp
            820                 825                 830

Arg Val Met Trp Arg Ile Pro Phe Ser Ser Asn Phe Met Ser Met Gly
        835                 840                 845

Ala Leu Thr Asp Leu Gly Gln Asn Met Leu Tyr Ala Asn Ser Ala His
    850                 855                 860

Ala Leu Asp Met Thr Phe Glu Val Asp Pro Met Asn Glu Pro Thr Leu
865                 870                 875                 880

Leu Tyr Met Leu Phe Glu Val Phe Asp Val Val Arg Val His Gln Pro
                885                 890                 895

His Arg Gly Ile Ile Glu Ala Val Tyr Leu Arg Thr Pro Phe Ser Ala
```

900 905 910

Gly Asn Ala Thr Thr
915

<210> 37
<211> 533
<212> PRT
<213> simian adenovirus SV-39

<400> 37

Met Lys Arg Ala Lys Val Glu Glu Gly Phe Asn Pro Val Tyr Pro Tyr
1 5 10 15

Gly Tyr Ser Thr Pro Thr Asp Val Ala Pro Pro Phe Val Ala Ser Asp
20 25 30

Gly Leu Gln Glu Asn Pro Pro Gly Val Leu Ser Leu Lys Ile Ser Lys
35 40 45

Pro Leu Thr Phe Asn Ala Ser Lys Ala Leu Ser Leu Ala Ile Gly Pro
50 55 60

Gly Leu Lys Ile Gln Asp Gly Lys Leu Val Gly Glu Gly Gln Ala Ile
65 70 75 80

Leu Ala Asn Leu Pro Leu Gln Ile Thr Asn Asn Thr Ile Ser Leu Arg
85 90 95

Phe Gly Asn Thr Leu Ala Leu Asn Asp Asn Asn Glu Leu Gln Thr Thr
100 105 110

Leu Lys Ser Ser Ser Pro Leu Lys Ile Thr Asp Gln Thr Leu Ser Leu
115 120 125

Asn Ile Gly Asp Ser Leu Ala Ile Lys Asp Asp Lys Leu Glu Ser Ala
130 135 140

Leu Gln Ala Thr Leu Pro Leu Ser Ile Ser Asn Asn Thr Ile Ser Leu
145 150 155 160

Asn Val Gly Thr Gly Leu Thr Ile Asn Gly Asn Val Leu Gln Ala Val
165 170 175

```
Pro Leu Asn Ala Leu Ser Pro Leu Thr Ile Ser Asn Asn Asn Ile Ser
            180                 185                 190

Leu Arg Tyr Gly Ser Ser Leu Thr Val Leu Asn Asn Glu Leu Gln Ser
        195                 200                 205

Asn Leu Thr Val His Ser Pro Leu Lys Leu Asn Ser Asn Asn Ser Ile
    210                 215                 220

Ser Leu Asn Thr Leu Ser Pro Phe Arg Ile Glu Asn Gly Phe Leu Thr
225                 230                 235                 240

Leu Tyr Leu Gly Thr Lys Ser Gly Leu Leu Val Gln Asn Ser Gly Leu
                245                 250                 255

Lys Val Gln Ala Gly Tyr Gly Leu Gln Val Thr Asp Thr Asn Ala Leu
            260                 265                 270

Thr Leu Arg Tyr Leu Ala Pro Leu Thr Ile Pro Asp Ser Gly Ser Glu
        275                 280                 285

Gln Gly Ile Leu Lys Val Asn Thr Gly Gln Gly Leu Ser Val Asn Gln
    290                 295                 300

Ala Gly Ala Leu Glu Thr Ser Leu Gly Gly Gly Leu Lys Tyr Ala Asp
305                 310                 315                 320

Asn Lys Ile Thr Phe Asp Thr Gly Asn Gly Leu Thr Leu Ser Glu Asn
                325                 330                 335

Lys Leu Ala Val Ala Ala Gly Ser Gly Leu Thr Phe Arg Asp Gly Ala
            340                 345                 350

Leu Val Ala Thr Gly Thr Ala Phe Thr Gln Thr Leu Trp Thr Thr Ala
        355                 360                 365

Asp Pro Ser Pro Asn Cys Thr Ile Ile Gln Asp Arg Asp Thr Lys Phe
    370                 375                 380

Thr Leu Ala Leu Thr Ile Ser Gly Ser Gln Val Leu Gly Thr Val Ser
385                 390                 395                 400
```

Ile Ile Gly Val Lys Gly Pro Leu Ser Ser Ser Ile Pro Ser Ala Thr
405 410 415

Val Thr Val Gln Leu Asn Phe Asp Ser Asn Gly Ala Leu Leu Ser Ser
420 425 430

Ser Ser Leu Lys Gly Tyr Trp Gly Tyr Arg Gln Gly Pro Ser Ile Asp
435 440 445

Pro Tyr Pro Ile Ile Asn Ala Leu Asn Phe Met Pro Asn Ser Leu Ala
450 455 460

Tyr Pro Pro Gly Gln Glu Ile Gln Ala Lys Cys Asn Met Tyr Val Ser
465 470 475 480

Thr Phe Leu Arg Gly Asn Pro Gln Arg Pro Ile Val Leu Asn Ile Thr
485 490 495

Phe Asn Asn Gln Thr Ser Gly Phe Ser Ile Arg Phe Thr Trp Thr Asn
500 505 510

Leu Thr Thr Gly Glu Ala Phe Ala Met Pro Pro Cys Thr Phe Ser Tyr
515 520 525

Ile Ala Glu Gln Gln
530

<210> 38
<211> 50
<212> DNA
<213> Artificial sequence

<220>
<223> oligomer SV25T

<400> 38
aatttaaata cgtagcgcac tagtcgcgct aagcgcggat atcatttaaa 50

<210> 39
<211> 49
<212> DNA
<213> Artificial sequence

<220>
<223> oligomer SV25B

<400> 39

tatttaaatg atatccgcgc ttaagcgcga ctagtgcgct acgtattta 49

379

380

380
381

PCT

REQUEST

The undersigned requests that the present international application be processed according to the Patent Cooperation Treaty.

For receiving Office use only

International Application No.

International Filing Date

Name of receiving Office and "PCT International Application"

Applicant's or agent's file reference *(if desired) (12 characters maximum)* UPN-02677PCT

Box No. I TITLE OF INVENTION

SIMIAN ADENOVIRUS NUCLEIC ACID AND AMINO ACID SEQUENCES, VECTORS CONTAINING SAME, AND METHODS OF USE

Box No. II APPLICANT ☐ This person is also inventor

Name and address: *(Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country. The country of the address indicated in this Box is the applicant's State (that is, country) of residence if no State of residence is indicated below.)*

The Trustees of the University of Pennsylvania
3160 Chestnut Street
Suite 200
Philadelphia, Pennsylvania 19104 US

Telephone No. 215-573-4508

Facsimile No. 215-898-9519

Teleprinter No.

Applicant's registration No. with the Office

State *(that is, country)* of nationality: US

State *(that is, country)* of residence: US

This person is applicant for the purposes of: ☐ all designated States ☒ all designated States except the United States of America ☐ the United States of America only ☐ the States indicated in the Supplemental Box

Box No. III FURTHER APPLICANT(S) AND/OR (FURTHER) INVENTOR(S)

Name and address: *(Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country. The country of the address indicated in this Box is the applicant's State (that is, country) of residence if no State of residence is indicated below.)*

Wilson, James M.
1350 N. Avignon Drive
Gladwyne, Pennsylvania 19035 US

This person is:
☐ applicant only
☒ applicant and inventor
☐ inventor only *(If this check-box is marked, do not fill in below.)*

Applicant's registration No. with the Office

State *(that is, country)* of nationality: US

State *(that is, country)* of residence: US

This person is applicant for the purposes of: ☐ all designated States ☐ all designated States except the United States of America ☒ the United States of America only ☐ the States indicated in the Supplemental Box

☒ Further applicants and/or (further) inventors are indicated on a continuation sheet.

Box No. IV AGENT OR COMMON REPRESENTATIVE; OR ADDRESS FOR CORRESPONDENCE

The person identified below is hereby/has been appointed to act on behalf of the applicant(s) before the competent International Authorities as: ☒ agent ☐ common representative

Name and address: *(Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country.)*

Kodroff, Cathy A.
Howson and Howson
Spring House Corporate Center
P.O. Box 457
Spring House, Pennsylvania 19477 US

Telephone No. 215-540-9200

Facsimile No. 215-540-5818

Teleprinter No.

Agent's registration No. with the Office 33,980

☐ Address for correspondence: Mark this check-box where no agent or common representative is/has been appointed and the space above is used instead to indicate a special address to which correspondence should be sent.

Continuation of Box No. III FURTHER APPLICANT(S) AND/OR (FURTHER) INVENTOR(S)

If none of the following sub-boxes is used, this sheet should not be included in the request.

Name and address: *(Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country. The country of the address indicated in this Box is the applicant's State (that is, country) of residence if no State of residence is indicated below.)*

Gao, Guangping
408 Yorkshire Road
Rosemont, Pennsylvania 19010 US

This person is:
- ☐ applicant only
- ☒ applicant and inventor
- ☐ inventor only *(If this check-box is marked, do not fill in below.)*

Applicant's registration No. with the Office:

State *(that is, country)* of nationality: US

State *(that is, country)* of residence: US

This person is applicant for the purposes of:
- ☐ all designated States
- ☐ all designated States except the United States of America
- ☒ the United States of America only
- ☐ the States indicated in the Supplemental Box

Name and address: *(Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country. The country of the address indicated in this Box is the applicant's State (that is, country) of residence if no State of residence is indicated below.)*

Roy, Soumitra
240 Pugh Road
Wayne, Pennsylvania 19087 US

This person is:
- ☐ applicant only
- ☒ applicant and inventor
- ☐ inventor only *(If this check-box is marked, do not fill in below.)*

Applicant's registration No. with the Office:

State *(that is, country)* of nationality: US

State *(that is, country)* of residence: US

This person is applicant for the purposes of:
- ☐ all designated States
- ☐ all designated States except the United States of America
- ☒ the United States of America only
- ☐ the States indicated in the Supplemental Box

Name and address: *(Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country. The country of the address indicated in this Box is the applicant's State (that is, country) of residence if no State of residence is indicated below.)*

This person is:
- ☐ applicant only
- ☐ applicant and inventor
- ☐ inventor only *(If this check-box is marked, do not fill in below.)*

Applicant's registration No. with the Office:

State *(that is, country)* of nationality:

State *(that is, country)* of residence:

This person is applicant for the purposes of:
- ☐ all designated States
- ☐ all designated States except the United States of America
- ☐ the United States of America only
- ☐ the States indicated in the Supplemental Box

Name and address: *(Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country. The country of the address indicated in this Box is the applicant's State (that is, country) of residence if no State of residence is indicated below.)*

This person is:
- ☐ applicant only
- ☐ applicant and inventor
- ☐ inventor only *(If this check-box is marked, do not fill in below.)*

Applicant's registration No. with the Office:

State *(that is, country)* of nationality:

State *(that is, country)* of residence:

This person is applicant for the purposes of:
- ☐ all designated States
- ☐ all designated States except the United States of America
- ☐ the United States of America only
- ☐ the States indicated in the Supplemental Box

☐ Further applicants and/or (further) inventors are indicated on another continuation sheet.

Box No. V DESIGNATION OF STATES *Mark the applicable check-boxes below; at least one must be marked.*

The following designations are hereby made under Rule 4.9(a):

Regional Patent

☒ **AP** **ARIPO Patent:** GH Ghana, GM Gambia, KE Kenya, LS Lesotho, MW Malawi, MZ Mozambique, SD Sudan, SL Sierra Leone, SZ Swaziland, TZ United Republic of Tanzania, UG Uganda, ZM Zambia, ZW Zimbabwe, and any other State which is a Contracting State of the Harare Protocol and of the PCT *(If other kind of protection or treatment desired, specify on dotted line)* .....

☒ **EA** **Eurasian Patent:** AM Armenia, AZ Azerbaijan, BY Belarus, KG Kyrgyzstan, KZ Kazakhstan, MD Republic of Moldova, RU Russian Federation, TJ Tajikistan, TM Turkmenistan, and any other State which is a Contracting State of the Eurasian Patent Convention and of the PCT

☒ **EP** **European Patent:** AT Austria, BE Belgium, BG Bulgaria, CH & LI Switzerland and Liechtenstein, CY Cyprus, CZ Czech Republic, DE Germany, DK Denmark, EE Estonia, ES Spain, FI Finland, FR France, GB United Kingdom, GR Greece, IE Ireland, IT Italy, LU Luxembourg, MC Monaco, NL Netherlands, PT Portugal, SE Sweden, SK Slovakia, TR Turkey, and any other State which is a Contracting State of the European Patent Convention and of the PCT

☒ **OA** **OAPI Patent:** BF Burkina Faso, BJ Benin, CF Central African Republic, CG Congo, CI Côte d'Ivoire, CM Cameroon, GA Gabon, GN Guinea, GQ Equatorial Guinea, GW Guinea-Bissau, ML Mali, MR Mauritania, NE Niger, SN Senegal, TD Chad, TG Togo, and any other State which is a member State of OAPI and a Contracting State of the PCT *(If other kind of protection or treatment desired, specify on dotted line)* .....

National Patent *(If other kind of protection or treatment desired, specify on dotted line)*:

☒ AE United Arab Emirates	☒ GM Gambia	☒ NZ New Zealand
☒ AG Antigua and Barbuda	☒ HR Croatia	☒ OM Oman
☒ AL Albania	☒ HU Hungary	☒ PH Philippines
☒ AM Armenia	☒ ID Indonesia	☒ PL Poland
☒ AT Austria	☒ IL Israel	☒ PT Portugal
☒ AU Australia	☒ IN India	☒ RO Romania
☒ AZ Azerbaijan	☒ IS Iceland	☒ RU Russian Federation
☒ BA Bosnia and Herzegovina	☒ JP Japan	
☒ BB Barbados	☒ KE Kenya	☒ SD Sudan
☒ BG Bulgaria	☒ KG Kyrgyzstan	☒ SE Sweden
☒ BR Brazil	☒ KP Democratic People's Republic of Korea	☒ SG Singapore
☒ BY Belarus		☒ SI Slovenia
☒ BZ Belize	☒ KR Republic of Korea	☒ SK Slovakia
☒ CA Canada	☒ KZ Kazakhstan	☒ SL Sierra Leone
☒ CH & LI Switzerland and Liechtenstein	☒ LC Saint Lucia	☒ TJ Tajikistan
☒ CN China	☒ LK Sri Lanka	☒ TM Turkmenistan
☒ CO Colombia	☒ LR Liberia	☒ TN Tunisia
☒ CR Costa Rica	☒ LS Lesotho	☒ TR Turkey
☒ CU Cuba	☒ LT Lithuania	☒ TT Trinidad and Tobago
☒ CZ Czech Republic	☒ LU Luxembourg	
☒ DE Germany	☒ LV Latvia	☒ TZ United Republic of Tanzania
☒ DK Denmark	☒ MA Morocco	☒ UA Ukraine
☒ DM Dominica	☒ MD Republic of Moldova	☒ UG Uganda
☒ DZ Algeria		☒ US United States of America
☒ EC Ecuador	☒ MG Madagascar	
☒ EE Estonia	☒ MK The former Yugoslav Republic of Macedonia	☒ UZ Uzbekistan
☒ ES Spain		☒ VN Viet Nam
☒ FI Finland	☒ MN Mongolia	☒ YU Yugoslavia
☒ GB United Kingdom	☒ MW Malawi	☒ ZA South Africa
☒ GD Grenada	☒ MX Mexico	☒ ZM Zambia
☒ GE Georgia	☒ MZ Mozambique	☒ ZW Zimbabwe
☒ GH Ghana	☒ NO Norway	

Check-boxes below reserved for designating States which have become party to the PCT after issuance of this sheet:

☒ VC Saint Vincent & the Grenadines	☐	☐
☐	☐	☐

Precautionary Designation Statement: In addition to the designations made above, the applicant also makes under Rule 4.9(b) all other designations which would be permitted under the PCT except any designation(s) indicated in the Supplemental Box as being excluded from the scope of this statement. The applicant declares that those additional designations are subject to confirmation and that any designation which is not confirmed before the expiration of 15 months from the priority date is to be regarded as withdrawn by the applicant at the expiration of that time limit. *(Confirmation (including fees) must reach the receiving Office within the 15-month time limit.)*

Supplemental Box *If the Supplemental Box is not used, this sheet should not be included in the request.*

1. If, in any of the Boxes, except Boxes Nos. VIII(i) to (v) for which a special continuation box is provided, the space is insufficient to furnish all the information: in such case, write "Continuation of Box No...." (indicate the number of the Box) and furnish the information in the same manner as required according to the captions of the Box in which the space was insufficient, in particular:

(i) if more than two persons are to be indicated as applicants and/or inventors and no "continuation sheet" is available: in such case, write "Continuation of Box No. III" and indicate for each additional person the same type of information as required in Box No. III. The country of the address indicated in this Box is the applicant's State (that is, country) of residence if no State of residence is indicated below;

(ii) if, in Box No. II or in any of the sub-boxes of Box No. III, the indication "the States indicated in the Supplemental Box" is checked: in such case, write "Continuation of Box No. II" or "Continuation of Box No. III" or "Continuation of Boxes No. II and No. III" (as the case may be), indicate the name of the applicant(s) involved and, next to (each) such name, the State(s) (and/or, where applicable, ARIPO, Eurasian, European or OAPI patent) for the purposes of which the named person is applicant;

(iii) if, in Box No. II or in any of the sub-boxes of Box No. III, the inventor or the inventor/applicant is not inventor for the purposes of all designated States or for the purposes of the United States of America: in such case, write "Continuation of Box No. II" or "Continuation of Box No. III" or "Continuation of Boxes No. II and No. III" (as the case may be), indicate the name of the inventor(s) and, next to (each) such name, the State(s) (and/or, where applicable, ARIPO, Eurasian, European or OAPI patent) for the purposes of which the named person is inventor;

(iv) if, in addition to the agent(s) indicated in Box No. IV, there are further agents: in such case, write "Continuation of Box No. IV" and indicate for each further agent the same type of information as required in Box No. IV;

(v) if, in Box No. V, the name of any State (or OAPI) is accompanied by the indication "patent of addition," or "certificate of addition," or if, in Box No. V, the name of the United States of America is accompanied by an indication "continuation" or "continuation-in-part": in such case, write "Continuation of Box No. V" and the name of each State involved (or OAPI), and after the name of each such State (or OAPI), the number of the parent title or parent application and the date of grant of the parent title or filing of the parent application;

(vi) if, in Box No. VI, there are more than five earlier applications whose priority is claimed: in such case, write "Continuation of Box No. VI" and indicate for each additional earlier application the same type of information as required in Box No. VI.

2. If, with regard to the precautionary designation statement contained in Box No. V, the applicant wishes to exclude any State(s) from the scope of that statement: in such case, write "Designation(s) excluded from precautionary designation statement" and indicate the name or two-letter code of each State so excluded.

"Continuation of Box No. IV"

Kita, Stanley B.
Smith, George A., Jr.
Bak, Mary E.
Bak, William
Palovich, Tracy *

* registered USPTO patent agent only

All above attorneys are members of the firm of Howson and Howson.

Address of all is indicated in Box No. IV.

Box No. VI PRIORITY CLAIM

The priority of the following earlier application(s) is hereby claimed:

Filing date of earlier application *(day/month/year)*	Number of earlier application	Where earlier application is:		
		national application: country or Member of WTO	regional application:* regional Office	international application: receiving Office
item (1) 21 November 2001 (21/11/01)	60/331,951	US		
item (2) 22 March 2002 (22/03/02)	60/366,798	US		
item (3)				
item (4)				
item (5)				

☐ Further priority claims are indicated in the Supplemental Box.

The receiving Office is requested to prepare and transmit to the International Bureau a certified copy of the earlier application(s) *(only if the earlier application was filed with the Office which for the purposes of this international application is the receiving Office)* identified above as:

☒ all items ☐ item (1) ☐ item (2) ☐ item (3) ☐ item (4) ☐ item (5) ☐ other, see Supplemental Box

* *Where the earlier application is an ARIPO application, indicate at least one country party to the Paris Convention for the Protection of Industrial Property or one Member of the World Trade Organization for which that earlier application was filed (Rule 4.10(b)(ii)):* ....

.....

Box No. VII INTERNATIONAL SEARCHING AUTHORITY

Choice of International Searching Authority (ISA) *(if two or more International Searching Authorities are competent to carry out the international search, indicate the Authority chosen; the two-letter code may be used)*:

ISA / US .....

Request to use results of earlier search; reference to that search *(if an earlier search has been carried out by or requested from the International Searching Authority)*:

Date *(day/month/year)* Number Country *(or regional Office)*

Box No. VIII DECLARATIONS

The following declarations are contained in Boxes Nos. VIII (i) to (v) *(mark the applicable check-boxes below and indicate in the right column the number of each type of declaration)*:		Number of declarations
☐ Box No. VIII (i)	Declaration as to the identity of the inventor	:
☐ Box No. VIII (ii)	Declaration as to the applicant's entitlement, as at the international filing date, to apply for and be granted a patent	:
☐ Box No. VIII (iii)	Declaration as to the applicant's entitlement, as at the international filing date, to claim the priority of the earlier application	:
☐ Box No. VIII (iv)	Declaration of inventorship (only for the purposes of the designation of the United States of America)	:
☐ Box No. VIII (v)	Declaration as to non-prejudicial disclosures or exceptions to lack of novelty	:

Box No. IX CHECK LIST; LANGUAGE OF FILING

This international application contains:

(a) the following number of sheets in paper form:

request (including declaration sheets)	:	6
description (excluding sequence listing part)	:	80
claims	:	6
abstract	:	1
drawings	:	2
Sub-total number of sheets	:	95
sequence listing part of description *(actual number of sheets if filed in paper form, whether or not also filed in computer readable form; see (b) below)*	:	217
Total number of sheets	:	312

(b) sequence listing part of description filed in computer readable form

(i) ☐ only (under Section 801(a)(i))

(ii) ☒ in addition to being filed in paper form (under Section 801(a)(ii))

Type and number of carriers (diskette, CD-ROM, CD-R or other) on which the sequence listing part is contained *(additional copies to be indicated under item 9(ii), in right column)*: ...1 diskette...

This international application is accompanied by the following item(s) *(mark the applicable check-boxes below and indicate in right column the number of each item)*: | Number of items

1. ☒ fee calculation sheet	:	1
2. ☐ original separate power of attorney	:	
3. ☐ original general power of attorney	:	
4. ☒ copy of general power of attorney; reference number, if any: ...	:	1
5. ☐ statement explaining lack of signature	:	
6. ☐ priority document(s) identified in Box No. VI as item(s): ...	:	
7. ☐ translation of international application into *(language)*: ...	:	
8. ☐ separate indications concerning deposited microorganism or other biological material	:	
9. ☒ sequence listing in computer readable form (indicate also type and number of carriers (diskette, CD-ROM, CD-R or other))		
(i) ☐ copy submitted for the purposes of international search under Rule 13*ter* only (and not as part of the international application)	:	
(ii) ☐ *(only where check-box (b)(i) or (b)(ii) is marked in left column)* additional copies including, where applicable, the copy for the purposes of international search under Rule 13*ter*	:	
(iii) ☒ together with relevant statement as to the identity of the copy or copies with the sequence listing part mentioned in left column	:	1
10. ☒ other *(specify)*: transmittal letter to the USRO ...	:	1

Figure of the drawings which should accompany the abstract:

Language of filing of the international application: English

Box No. X SIGNATURE OF APPLICANT, AGENT OR COMMON REPRESENTATIVE

Next to each signature, indicate the name of the person signing and the capacity in which the person signs (if such capacity is not obvious from reading the request).

By: Cathy A Kodroff
Cathy A. Kodroff
Attorney for Applicants

For receiving Office use only

1. Date of actual receipt of the purported international application:

2. Drawings: ☐ received: ☐ not received:

3. Corrected date of actual receipt due to later but timely received papers or drawings completing the purported international application:

4. Date of timely receipt of the required corrections under PCT Article 11(2):

5. International Searching Authority (if two or more are competent): ISA /

6. ☐ Transmittal of search copy delayed until search fee is paid

For International Bureau use only

Date of receipt of the record copy by the International Bureau:

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
5 June 2003 (05.06.2003)

PCT

(10) International Publication Number
WO 03/046124 A2

(51) International Patent Classification[7]: C12N

(21) International Application Number: PCT/US02/33645

(22) International Filing Date:
20 November 2002 (20.11.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/331,951 21 November 2001 (21.11.2001) US
60/366,798 22 March 2002 (22.03.2002) US

(71) Applicant *(for all designated States except US)*: THE TRUSTEES OF THE UNIVERSITY OF PENNSYLVANIA [US/US]; 3160 Chestnut Street, Suite 200, Philadelphia, PA 19104 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: WILSON, James, M. [US/US]; 1350 N. Avignon Drive, Gladwyne, PA 19035 (US). GAO, Guangping [US/US]; 408 Yorkshire Road, Rosemont, PA 19010 (US). ROY, Soumitra [US/US]; 240 Pugh Road, Wayne, PA 19087 (US).

(74) Agents: Kodroff, Cathy, A. et al.; Howson and Howson, Spring House Corporate Center, P.O. Box 457, Spring House, PA 19477 (US).

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

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

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

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

WO 03/046124 A2

(54) Title: SIMIAN ADENOVIRUS NUCLEIC ACID AND AMINO ACID SEQUENCES, VECTORS CONTAINING SAME, AND METHODS OF USE

(57) Abstract: A recombinant vector comprises simian adenovirus sequences and a heterologous gene under the control of regulatory sequences. A cell line which expresses simian adenovirus gene(s) is also disclosed. Methods of using the vectors and cell lines are provided.

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

388

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
5 June 2003 (05.06.2003)

PCT

(10) International Publication Number
WO 03/046124 A3

(51) International Patent Classification⁷: A61K 47/42, C07H 21/00, C07K 14/075, C12N 7/00, 15/33, 15/861

(21) International Application Number: PCT/US02/33645

(22) International Filing Date:
20 November 2002 (20.11.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/331,951 21 November 2001 (21.11.2001) US
60/366,798 22 March 2002 (22.03.2002) US

(71) Applicant *(for all designated States except US)*: THE TRUSTEES OF THE UNIVERSITY OF PENNSYLVANIA [US/US]; 3160 Chestnut Street, Suite 200, Philadelphia, PA 19104 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: WILSON, James, M. [US/US]; 1350 N. Avignon Drive, Gladwyne, PA 19035 (US). GAO, Guangping [US/US]; 408 Yorkshire Road, Rosemont, PA 19010 (US). ROY, Soumitra [US/US]; 240 Pugh Road, Wayne, PA 19087 (US).

(74) Agents: Kodroff, Cathy, A. et al.; Howson and Howson, Spring House Corporate Center, P.O. Box 457, Spring House, PA 19477 (US).

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

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

Published:
— *with international search report*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

(88) Date of publication of the international search report:
20 November 2003

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

WO 03/046124 A3

(54) Title: SIMIAN ADENOVIRUS NUCLEIC ACID AND AMINO ACID SEQUENCES, VECTORS CONTAINING SAME, AND METHODS OF USE

(57) Abstract: A recombinant vector comprises simian adenovirus sequences and a heterologous gene under the control of regulatory sequences. A cell line which expresses simian adenovirus gene(s) is also disclosed. Methods of using the vectors and cell lines are provided.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US02/33645

Box I Observations where certain claims were found unsearchable (Continuation of Item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claim Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claim Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful International search can be carried out, specifically:

3. ☒ Claim Nos.: 12-21, 23-27
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of Item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:
Please See Continuation Sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☒ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.: Please See Continuation Sheet
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest
☐ The additional search fees were accompanied by the applicant's protest.
☒ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

PCT/US02/33645

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING

This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fees must be paid.

Groups 1-5, claim(s) 1, 2, 4-7, and 22, each group drawn to the nucleic acid sequences of Ad Pan-5, Ad Pan-6, Ad Pan-7, Ad SV-1, Ad SV-25 and Ad SV-39, respectively.

Groups 6-17, 19-22, 24-37, 39-55, 57-60, 62-75, 77-93, 95-98, 100-113; 115-129, 131-134, 136-145, 148-162, 164-167, 169-173, 176-189, 191-194, 196-203, 205-210, claim(s) 3, each drawn to in order:

an Ad Pan-5 E1a protein, the Ad Pan-5 Small T protein, the Ad Pan-5 Large T protein, the Ad Pan-5 protein IX, the Ad Pan-5 PTP, the Ad Pan-5 Polymerase, the Ad Pan-5 IVa2 protein, the Ad Pan-5 28.1 kD protein, the Ad Pan-5 agnoprotein, the Ad Pan-5 52/55D protein, the Ad Pan-5 IIIa protein, the Ad Pan-5 protein VII , the Ad Pan-5 protein V, the Ad Pan-5 Mu protein, the Ad Pan-5 protein VI, the Ad Pan-5 endoprotease, the Ad Pan-5 DBP, the Ad Pan-5 100kD protein, the Ad Pan-5 33kD homolog, the Ad Pan-5 protein VIII, the Ad Pan-5 E3 ORF1 protein, the Ad Pan-5 E3 ORF2 protein, the Ad Pan-5 E3 ORF3 protein, the Ad Pan-5 E3 ORF4 protein, the Ad Pan-5 E3 ORF5 protein, the Ad Pan-5 E3 ORF6 protein, the Ad Pan-5 E3 ORF7 protein, the Ad Pan-5 E3 ORF8 protein, the Ad Pan-5 E3 ORF9 protein, the Ad Pan-5 E4 ORF7 protein, the Ad Pan-5 E4 ORF6 protein, the Ad Pan-5 E4 ORF4 protein, the Ad Pan-5 E4 ORF3 protein, the Ad Pan-5 E4 ORF2 protein, the Ad Pan-5 E4 ORF1 protein,

an Ad Pan-6 E1a protein, the Ad Pan-6 Small T protein, the Ad Pan-6 Large T protein, the Ad Pan-6 protein IX, the Ad Pan-6 PTP, the Ad Pan-6 Polymerase, the Ad Pan-6 IVa2 protein, the Ad Pan-6 28.1 kD protein, the Ad Pan-6 agnoprotein, the Ad Pan-6 52/55D protein, the Ad Pan-6 IIIa protein, the Ad Pan-6 protein VII , the Ad Pan-6 protein V, the Ad Pan-6 Mu protein, the Ad Pan-6 protein VI, the Ad Pan-6 endoprotease, the Ad Pan-6 DBP, the Ad Pan-6 100kD protein, the Ad Pan-6 33kD homolog, the Ad Pan-6 protein VIII, the Ad Pan-6 E3 ORF1 protein, the Ad Pan-6 E3 ORF2 protein, the Ad Pan-6 E3 ORF3 protein, the Ad Pan-6 E3 ORF4 protein, the Ad Pan-6 E3 ORF5 protein, the Ad Pan-6 E3 ORF6 protein, the Ad Pan-6 E3 ORF7 protein, the Ad Pan-6 E3 ORF8 protein, the Ad Pan-6 E3 ORF9 protein, the Ad Pan-6 E4 ORF7 protein, the Ad Pan-6 E4 ORF6 protein, the Ad Pan-6 E4 ORF4 protein, the Ad Pan-6 E4 ORF3 protein, the Ad Pan-6 E4 ORF2 protein, the Ad Pan-6 E4 ORF1 protein,

an Ad Pan-7 E1a protein, the Ad Pan-7 Small T protein, the Ad Pan-7 Large T protein, the Ad Pan-7 protein IX, the Ad Pan-7 PTP, the Ad Pan-7 Polymerase, the Ad Pan-7 IVa2 protein, the Ad Pan-7 28.1 kD protein, the Ad Pan-7 agnoprotein, the Ad Pan-7 52/55D protein, the Ad Pan-7 IIIa protein, the Ad Pan-7 protein VII , the Ad Pan-7 protein V, the Ad Pan-7 Mu protein, the Ad Pan-7 protein VI, the Ad Pan-7 endoprotease, the Ad Pan-7 DBP, the Ad Pan-7 100kD protein, the Ad Pan-7 33kD homolog, the Ad Pan-7 protein VIII, the Ad Pan-7 E3 ORF1 protein, the Ad Pan-7 E3 ORF2 protein, the Ad Pan-7 E3 ORF3 protein, the Ad Pan-7 E3 ORF4 protein, the Ad Pan-7 E3 ORF5 protein, the Ad Pan-7 E3 ORF6 protein, the Ad Pan-7 E3 ORF7 protein, the Ad Pan-7 E3 ORF8 protein, the Ad Pan-7 E3 ORF9 protein, the Ad Pan-7 E4 ORF7 protein, the Ad Pan-7 E4 ORF6 protein, the Ad Pan-7 E4 ORF4 protein, the Ad Pan-7 E4 ORF3 protein, the Ad Pan-7 E4 ORF2 protein, the Ad Pan-7 E4 ORF1 protein,

an Ad SV-1 E1a protein, the Ad SV-1 Large T protein, the Ad SV-1 protein IX, the Ad SV-1 PTP, the Ad SV-1 Polymerase, the Ad SV-1 IVa2 protein, the Ad SV-1 agnoprotein, the Ad SV-1 52/55D protein, the Ad SV-1 IIIa protein, the Ad SV-1 protein VII , the Ad SV-1 protein V, the Ad SV-1 Mu protein, the Ad SV-1 protein VI, the Ad SV-1 endoprotease, the Ad SV-1 DBP, the Ad SV-1 100kD protein, the Ad SV-1 protein VIII, the Ad SV-1 E3 ORF1 protein, the Ad SV-1 E3 ORF2 protein, the Ad SV-1 E3 ORF3 protein, the Ad SV-1 E3 ORF4 protein, the Ad SV-1 E3 ORF5 protein, the Ad SV-1 E3 ORF6 protein, the Ad SV-1 E4 ORF7 protein, the Ad SV-1 E4 ORF6 protein, the Ad SV-1 E4 ORF4 protein, the Ad SV-1 E4 ORF3 protein, the Ad SV-1 E4 ORF2 protein, the Ad SV-1 E4 ORF1 protein,

the Ad SV-25 Small T protein, the Ad SV-25 Large T protein, the Ad SV-25 protein IX, the Ad SV-25 PTP, the Ad SV-25 Polymerase, the Ad SV-25 IVa2 protein, the Ad SV-25 agnoprotein, the Ad SV-25 52/55D protein, the Ad SV-25 IIIa protein, the Ad SV-25 protein VII , the Ad SV-25 protein V, the Ad SV-25 Mu protein, the Ad SV-25 protein VI, the Ad SV-25 endoprotease, the Ad SV-25 DBP, the Ad SV-25 100kD protein, the Ad SV-25 protein VIII, the Ad SV-25 E3 ORF1 protein, the Ad SV-25 E4 ORF6 protein, the Ad SV-25 E4 ORF4 protein, the Ad SV-25 E4 ORF3 protein, the Ad SV-25 E4 ORF2 protein, the Ad SV-25 E4 ORF1 protein,

an Ad SV-39 E1a protein, the Ad SV-39 Small T protein, the Ad SV-39 Large T protein, the Ad SV-39 protein IX, the Ad SV-39 PTP, the Ad SV-39 Polymerase, the Ad SV-39 IVa2 protein, the Ad SV-39 52/55D protein, the Ad SV-39 IIIa protein, the Ad SV-39 protein VII , the Ad SV-39 protein V, the Ad SV-39 Mu protein, the Ad SV-39 protein VI, the Ad SV-39 endoprotease, the Ad SV-39 DBP, the Ad SV-39 100kD protein, the Ad SV-39 protein VIII, the Ad SV-39 E3 ORF1 protein, the Ad SV-39 E3 ORF2 protein, the Ad SV-39 E3 ORF3 protein, the Ad SV-39 E3 ORF4 protein, the Ad SV-39 E4 ORF7 protein, the Ad SV-39 E4 ORF6 protein, the Ad SV-39 E4 ORF4 protein, the Ad SV-39 E4 ORF3 protein, the Ad SV-39 E4 ORF2 protein, and the Ad SV-39 E4 ORF1 protein.

Groups 18, 56, 94, 130, 163, and 190, claim(s) 3, 8, 9, 28 and 29, each drawn to the penton protein of Ad Pan-5, Ad Pan-6, Ad Pan-7, Ad SV-1, Ad SV-25 and Ad SV-39, respectively.

INTERNATIONAL SEARCH REPORT

PCT/US02/33645

Groups 23, 61, 99, 135, 168, and 195, claim(s) 3, 8-10, 28 and 29, each drawn to the hexon protein of Ad Pan-5, Ad Pan-6, Ad Pan-7, Ad SV-1, Ad SV-25 and Ad SV-39, respectively.

Groups 38, 76, 114, 146, 147, 174, 175, and 204, claim(s) 3, 8, 9, 11, 28 and 29, each drawn to the Ad Pan-5 fiber protein, the Ad Pan-6 fiber protein, the Ad Pan-7 fiber protein, the Ad SV-1 fiber1 protein, the Ad SV-1 fiber2 protein, the Ad SV-25 fiber1 protein, the Ad SV-25 fiber2 protein and the Ad SV-39 fiber protein, respectively.
The inventions listed as Groups 1-210 do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons:

Each of groups 1-5 is directed to the genomic nucleic acid (or part thereof) of a different adenovirus, with no special technical feature linking them since adenoviruses, including simian adenoviruses, were known in the art as indicated in the disclosure (page 1, lines 22-24). Each of groups 6-210 is directed to the different proteins produced by one of five different adenoviruses. For proteins of one adenovirus, these proteins are materially different proteins with different functions, and thus share no special technical feature with one another. The proteins of groups 6-210 are materially different compounds from the nucleic acids which encode them. Inventions directed to different products are not recognized as sharing a special technical feature under 37 CFR 1.475(b).

Continuation of Box II Item 3:
1, 2, 4-7, 22 (directed to Ad-Pan5); 3, 8 (part a), 9, 10, 28, 29 (directed to hexon protein)

Continuation of B. FIELDS SEARCHED Item 3:
USPT, PGPB, DERWENT, GENBANK, PIR, SPTREMBL, SWISSPROT, GENESEQ
search terms: SEQ ID NO: 1 (at least 100 nucleotides identical, GenBank only); SEQ ID NOs: 3, 7, 11, 15, 16, 17, 26, 31, 36; adenovir?, chimpanzee, simian, Pan

392

PATENT COOPERATION TREATY

From the INTERNATIONAL SEARCHING AUTHORITY

OCT - 2 2003

To:
CATHY A. KODROFF
HOWSON AND HOWSON
SPRING HOUSE CORPORATE CENTER
P.O. BOX 457
SPRING HOUSE, PA 19477

PCT

NOTIFICATION OF TRANSMITTAL OF THE INTERNATIONAL SEARCH REPORT OR THE DECLARATION

(PCT Rule 44.1)

ENTERED
DUE 11/30/0

	Date of Mailing (*day/month/year*) 30 SEP 2003
Applicant's or agent's file reference UPN-02677/PCT	FOR FURTHER ACTION See paragraphs 1 and 4 below
International application No. PCT/US02/33645	International filing date (*day/month/year*) 20 November 2002 (20.11.2002)
Applicant THE TRUSTEES OF THE UNIVERSITY OF PENNSYLVANIA	

1. ☒ The applicant is hereby notified that the international search report has been established and is transmitted herewith.

   Filing of amendments and statement under Article 19:
   The applicant is entitled, if he so wishes, to amend the claims of the international application (see Rule 46):

   When? The time limit for filing such amendments is normally two months from the date of transmittal of the international search report.

   Where? Directly to the International Bureau of WIPO, 34, chemin des Colombettes 1211 Geneva 20, Switzerland, Facsimile No.: (41-22) 740.14.35

   For more detailed instructions, see the notes on the accompanying sheet.

2. ☐ The applicant is hereby notified that no international search report will be established and that the declaration under Article 17(2)(a) to that effect is transmitted herewith.

3. ☐ With regard to the protest against payment of (an) additional fee(s) under Rule 40.2, the applicant is notified that:

   ☐ the protest together with the decision thereon has been transmitted to the International Bureau together with the applicant's request to forward the texts of both the protest and the decision thereon to the designated Offices.

   ☐ no decision has been made yet on the protest; the applicant will be notified as soon as a decision is made.

4. Reminders

   Shortly after 18 months from the priority date, the international application will be published by the International Bureau. If the applicant wishes to avoid or postpone publication, a notice of withdrawal of the international application, or of the priority claim, must reach the International Bureau as provided in Rules 90 *bis*.1 and 90 *bis*.3, respectively, before the completion of the technical preparations for international publication.

   Within 19 months from the priority date, but only in respect of some designated Offices, a demand for international preliminary examination must be filed if the applicant wishes to postpone the entry into the national phase until 30 months from the priority date (in some Offices even later); otherwise the applicant must, within 20 months from the priority date, perform the prescribed acts for entry into the national phase before those designated Offices.

   In respect of other designated Offices, the time limit of 30 months (or later) will apply even if no demand is filed within 19 months.

   See the Annex to Form PCT/IB/301 and, for details about the applicable time limits, Office by Office, see the *PCT Applicant's Guide*, Volume II, National Chapters and the WIPO Internet site.

Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US Commissioner for Patents P.O. Box 1450 Alexandria, Virginia 22313-1450 Facsimile No. (703)305-3230	Authorized officer Scott D. Priebe Telephone No. (703) 308-0196

Form PCT/ISA/220 (April 2002) (*See notes on accompanying sheet*)

PATENT COOPERATION TREATY

PCT

INTERNATIONAL SEARCH REPORT

(PCT Article 18 and Rules 43 and 44)

Applicant's or agent's file reference UPN-02677PCT	FOR FURTHER ACTION	see Notification of Transmittal of International Search Report (Form PCT/ISA/220) as well as, where applicable, item 5 below.
International application No. PCT/US02/33645	International filing date (day/month/year) 20 November 2002 (20.11.2002)	(Earliest) Priority Date (day/month/year) 21 November 2001 (21.11.2001)
Applicant THE TRUSTEES OF THE UNIVERSITY OF PENNSYLVANIA		

This international search report has been prepared by this International Searching Authority and is transmitted to the applicant according to Article 18. A copy is being transmitted to the International Bureau.

This international search report consists of a total of 5 sheets.

☒ It is also accompanied by a copy of each prior art document cited in this report.

1. **Basis of the Report**
   a. With regard to the language, the international search was carried out on the basis of the international application in the language in which it was filed, unless otherwise indicated under this item.
   ☐ the international search was carried out on the basis of a translation of the international application furnished to this Authority (Rule 23.1(b)).
   b. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of the sequence listing:
   ☒ contained in the international application in written form.
   ☒ filed together with the international application in computer readable form.
   ☐ furnished subsequently to this Authority in written form.
   ☐ furnished subsequently to this Authority in computer readable form.
   ☐ the statement that the subsequently furnished written sequence listing does not go beyond the disclosure in the international application as filed has been furnished.
   ☐ the statement that the information recorded in computer readable form is identical to the written sequence listing has been furnished.
2. ☒ Certain claims were found unsearchable (See Box I).
3. ☒ Unity of invention is lacking (See Box II).
4. With regard to the title,
   ☒ the text is approved as submitted by the applicant.
   ☐ the text has been established by this Authority to read as follows:
5. With regard to the abstract,
   ☒ the text is approved as submitted by the applicant.
   ☐ the text has been established, according to Rule 38.2(b), by this Authority as it appears in Box III. The applicant may, within one month from the date of mailing of this international search report, submit comments to this Authority.
6. The figure of the drawings to be published with the abstract is Figure No. ____
   ☐ as suggested by the applicant. ☒ None of the figures
   ☐ because the applicant failed to suggest a figure.
   ☐ because this figure better characterizes the invention.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US02/33645

Box I Observations where certain claims were found unsearchable (Continuation of Item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claim Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claim Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☒ Claim Nos.: 12-21, 23-27
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of Item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:
Please See Continuation Sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☒ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.: Please See Continuation Sheet

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest
☐ The additional search fees were accompanied by the applicant's protest.
☒ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US02/33645

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : A61K 47/42; C07H 21/00; C07K 14/075; C12N 7/00, 15/33, 15/861
US CL : 435/235.1, 320.1; 514/773; 530/350; 536/23.1, 23.7, 23.72

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
U.S. : 435/235.1, 320.1; 514/773; 530/350; 536/23.1, 23.7, 23.72

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
Please See Continuation Sheet

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 6,083,716 A (WILSON et al.) 04 July 2000 (04.07.2000).	1, 2, 4-7, 22
P, A	FARINA et al. Replication-defective vector based on a chimpanzee adenovirus. Journal of Virology. December 2001, Vol. 75, No. 23, pages 11603-11613.	1-10, 22
X	US 6,127,525 A (CRYSTAL et al.) 03 October 2000 (03.10.2000), see entire document,	9
—	especially col. 3-11.	—
A		3, 8-10, 22
A	CRAWFORD-MIKSZA et al. Analysis of 15 adenovirus hexon proteins reveals the location and structure of seven hypervariable regions containing serotype-specific residues. Journal of Virology. March 1996, Vol. 70, No. 3, pages 1836-1844.	3, 8

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
22 August 2003 (22.08.2003)	30 SEP 2003
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US Commissioner for Patents P.O. Box 1450 Alexandria, Virginia 22313-1450 Facsimile No. (703)305-3230	Authorized officer Sean D. Priebe Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet) (July 1998)

INTERNATIONAL SEARCH REPORT

PCT/US02/33645

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING

This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fees must be paid.

Groups 1-5, claim(s) 1, 2, 4-7, and 22, each group drawn to the nucleic acid sequences of Ad Pan-5, Ad Pan-6, Ad Pan-7, Ad SV-1, Ad SV-25 and Ad SV-39, respectively.

Groups 6-17, 19-22, 24-37, 39-55, 57-60, 62-75, 77-93, 95-98, 100-113, 115-129, 131-134, 136-145, 148-162, 164-167, 169-173, 176-189, 191-194, 196-203, 205-210, claim(s) 3, each drawn to in order:

an Ad Pan-5 E1a protein, the Ad Pan-5 Small T protein, the Ad Pan-5 Large T protein, the Ad Pan-5 protein IX, the Ad Pan-5 PTP, the Ad Pan-5 Polymerase, the Ad Pan-5 IVa2 protein, the Ad Pan-5 28.1 kD protein, the Ad Pan-5 agnoprotein, the Ad Pan-5 52/55D protein, the Ad Pan-5 IIIa protein, the Ad Pan-5 protein VII , the Ad Pan-5 protein V, the Ad Pan-5 Mu protein, the Ad Pan-5 protein VI, the Ad Pan-5 endoprotease, the Ad Pan-5 DBP, the Ad Pan-5 100kD protein, the Ad Pan-5 33kD homolog, the Ad Pan-5 protein VIII, the Ad Pan-5 E3 ORF1 protein, the Ad Pan-5 E3 ORF2 protein, the Ad Pan-5 E3 ORF3 protein, the Ad Pan-5 E3 ORF4 protein, the Ad Pan-5 E3 ORF5 protein, the Ad Pan-5 E3 ORF6 protein, the Ad Pan-5 E3 ORF7 protein, the Ad Pan-5 E3 ORF8 protein, the Ad Pan-5 E3 ORF9 protein, the Ad Pan-5 E4 ORF7 protein, the Ad Pan-5 E4 ORF6 protein, the Ad Pan-5 E4 ORF4 protein, the Ad Pan-5 E4 ORF3 protein, the Ad Pan-5 E4 ORF2 protein, the Ad Pan-5 E4 ORF1 protein,

an Ad Pan-6 E1a protein, the Ad Pan-6 Small T protein, the Ad Pan-6 Large T protein, the Ad Pan-6 protein IX, the Ad Pan-6 PTP, the Ad Pan-6 Polymerase, the Ad Pan-6 IVa2 protein, the Ad Pan-6 28.1 kD protein, the Ad Pan-6 agnoprotein, the Ad Pan-6 52/55D protein, the Ad Pan-6 IIIa protein, the Ad Pan-6 protein VII , the Ad Pan-6 protein V, the Ad Pan-6 Mu protein, the Ad Pan-6 protein VI, the Ad Pan-6 endoprotease, the Ad Pan-6 DBP, the Ad Pan-6 100kD protein, the Ad Pan-6 33kD homolog, the Ad Pan-6 protein VIII, the Ad Pan-6 E3 ORF1 protein, the Ad Pan-6 E3 ORF2 protein, the Ad Pan-6 E3 ORF3 protein, the Ad Pan-6 E3 ORF4 protein, the Ad Pan-6 E3 ORF5 protein, the Ad Pan-6 E3 ORF6 protein, the Ad Pan-6 E3 ORF7 protein, the Ad Pan-6 E3 ORF8 protein, the Ad Pan-6 E3 ORF9 protein, the Ad Pan-6 E4 ORF7 protein, the Ad Pan-6 E4 ORF6 protein, the Ad Pan-6 E4 ORF4 protein, the Ad Pan-6 E4 ORF3 protein, the Ad Pan-6 E4 ORF2 protein, the Ad Pan-6 E4 ORF1 protein,

an Ad Pan-7 E1a protein, the Ad Pan-7 Small T protein, the Ad Pan-7 Large T protein, the Ad Pan-7 protein IX, the Ad Pan-7 PTP, the Ad Pan-7 Polymerase, the Ad Pan-7 IVa2 protein, the Ad Pan-7 28.1 kD protein, the Ad Pan-7 agnoprotein, the Ad Pan-7 52/55D protein, the Ad Pan-7 IIIa protein, the Ad Pan-7 protein VII , the Ad Pan-7 protein V, the Ad Pan-7 Mu protein, the Ad Pan-7 protein VI, the Ad Pan-7 endoprotease, the Ad Pan-7 DBP, the Ad Pan-7 100kD protein, the Ad Pan-7 33kD homolog, the Ad Pan-7 protein VIII, the Ad Pan-7 E3 ORF1 protein, the Ad Pan-7 E3 ORF2 protein, the Ad Pan-7 E3 ORF3 protein, the Ad Pan-7 E3 ORF4 protein, the Ad Pan-7 E3 ORF5 protein, the Ad Pan-7 E3 ORF6 protein, the Ad Pan-7 E3 ORF7 protein, the Ad Pan-7 E3 ORF8 protein, the Ad Pan-7 E3 ORF9 protein, the Ad Pan-7 E4 ORF7 protein, the Ad Pan-7 E4 ORF6 protein, the Ad Pan-7 E4 ORF4 protein, the Ad Pan-7 E4 ORF3 protein, the Ad Pan-7 E4 ORF2 protein, the Ad Pan-7 E4 ORF1 protein,

an Ad SV-1 E1a protein, the Ad SV-1 Large T protein, the Ad SV-1 protein IX, the Ad SV-1 PTP, the Ad SV-1 Polymerase, the Ad SV-1 IVa2 protein, the Ad SV-1 agnoprotein, the Ad SV-1 52/55D protein, the Ad SV-1 IIIa protein, the Ad SV-1 protein VII , the Ad SV-1 protein V, the Ad SV-1 Mu protein, the Ad SV-1 protein VI, the Ad SV-1 endoprotease, the Ad SV-1 DBP, the Ad SV-1 100kD protein, the Ad SV-1 protein VIII, the Ad SV-1 E3 ORF1 protein, the Ad SV-1 E3 ORF2 protein, the Ad SV-1 E3 ORF3 protein, the Ad SV-1 E3 ORF4 protein, the Ad SV-1 E3 ORF5 protein, the Ad SV-1 E3 ORF6 protein, the Ad SV-1 E4 ORF7 protein, the Ad SV-1 E4 ORF6 protein, the Ad SV-1 E4 ORF4 protein, the Ad SV-1 E4 ORF3 protein, the Ad SV-1 E4 ORF2 protein, the Ad SV-1 E4 ORF1 protein,

the Ad SV-25 Small T protein, the Ad SV-25 Large T protein, the Ad SV-25 protein IX, the Ad SV-25 PTP, the Ad SV-25 Polymerase, the Ad SV-25 IVa2 protein, the Ad SV-25 agnoprotein, the Ad SV-25 52/55D protein, the Ad SV-25 IIIa protein, the Ad SV-25 protein VII , the Ad SV-25 protein V, the Ad SV-25 Mu protein, the Ad SV-25 protein VI, the Ad SV-25 endoprotease, the Ad SV-25 DBP, the Ad SV-25 100kD protein, the Ad SV-25 protein VIII, the Ad SV-25 E3 ORF1 protein, the Ad SV-25 E4 ORF6 protein, the Ad SV-25 E4 ORF4 protein, the Ad SV-25 E4 ORF3 protein, the Ad SV-25 E4 ORF2 protein, the Ad SV-25 E4 ORF1 protein,

an Ad SV-39 E1a protein, the Ad SV-39 Small T protein, the Ad SV-39 Large T protein, the Ad SV-39 protein IX, the Ad SV-39 PTP, the Ad SV-39 Polymerase, the Ad SV-39 IVa2 protein, the Ad SV-39 52/55D protein, the Ad SV-39 IIIa protein, the Ad SV-39 protein VII , the Ad SV-39 protein V, the Ad SV-39 Mu protein, the Ad SV-39 protein VI, the Ad SV-39 endoprotease, the Ad SV-39 DBP, the Ad SV-39 100kD protein, the Ad SV-39 protein VIII, the Ad SV-39 E3 ORF1 protein, the Ad SV-39 E3 ORF2 protein, the Ad SV-39 E3 ORF3 protein, the Ad SV-39 E3 ORF4 protein, the Ad SV-39 E4 ORF7 protein, the Ad SV-39 E4 ORF6 protein, the Ad SV-39 E4 ORF4 protein, the Ad SV-39 E4 ORF3 protein, the Ad SV-39 E4 ORF2 protein, and the Ad SV-39 E4 ORF1 protein.

Groups 18, 56, 94, 130, 163, and 190, claim(s) 3, 8, 9, 28 and 29, each drawn to the penton protein of Ad Pan-5, Ad Pan-6, Ad Pan-7, Ad SV-1, Ad SV-25 and Ad SV-39, respectively.

INTERNATIONAL SEARCH REPORT

PCT/US02/33645

Groups 23, 61, 99, 135, 168, and 195, claim(s) 3, 8-10, 28 and 29, each drawn to the hexon protein of Ad Pan-5, Ad Pan-6, Ad Pan-7, Ad SV-1, Ad SV-25 and Ad SV-39, respectively.

Groups 38, 76, 114, 146, 147, 174, 175, and 204, claim(s) 3, 8, 9, 11, 28 and 29, each drawn to the Ad Pan-5 fiber protein, the Ad Pan-6 fiber protein, the Ad Pan-7 fiber protein, the Ad SV-1 fiber1 protein, the Ad SV-1 fiber2 protein, the Ad SV-25 fiber1 protein, the Ad SV-25 fiber2 protein and the Ad SV-39 fiber protein, respectively.
The inventions listed as Groups 1-210 do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons:

Each of groups 1-5 is directed to the genomic nucleic acid (or part thereof) of a different adenovirus, with no special technical feature linking them since adenoviruses, including simian adenoviruses, were known in the art as indicated in the disclosure (page 1, lines 22-24). Each of groups 6-210 is directed to the different proteins produced by one of five different adenoviruses. For proteins of one adenovirus, these proteins are materially different proteins with different functions, and thus share no special technical feature with one another. The proteins of groups 6-210 are materially different compounds from the nucleic acids which encode them. Inventions directed to different products are not recognized as sharing a special technical feature under 37 CFR 1.475(b).

Continuation of Box II Item 3:
1, 2, 4-7, 22 (directed to Ad-Pan5); 3, 8 (part a), 9, 10, 28, 29 (directed to hexon protein)

Continuation of B. FIELDS SEARCHED Item 3:
USPT, PGPB, DERWENT, GENBANK, PIR, SPTREMBL, SWISSPROT, GENESEQ
search terms: SEQ ID NO: 1 (at least 100 nucleotides identical, GenBank only); SEQ ID NOs: 3, 7, 11, 15, 16, 17, 26, 31, 36; adenovir?, chimpanzee, simian, Pan

The demand must be filed directly with the competent International Preliminary Examining Authority or, if two or more Authorities are competent, with the one chosen by the applicant. The full name or two-letter code of that Authority may be indicated by the applicant on the line below:

IPEA/ US

PCT

CHAPTER II

DEMAND

under Article 31 of the Patent Cooperation Treaty:
The undersigned requests that the international application specified below be the subject of international preliminary examination according to the Patent Cooperation Treaty and hereby elects all eligible States (except where otherwise indicated).

For International Preliminary Examining Authority use only	
Identification of IPEA	Date of receipt of DEMAND

Box No. I IDENTIFICATION OF THE INTERNATIONAL APPLICATION		Applicant's or agent's file reference
		UPN-02677PCT
International application No.	International filing date *(day/month/year)*	(Earliest) Priority date *(day/month/year)*
PCT/US02/33645	20 November 2002	21 November 2001

Title of invention
SIMIAN ADENOVIRUS NUCLEIC ACID AND AMINO ACID SEQUENCES, VECTORS CONTAINING SAME, AND METHODS OF USE

Box No. II APPLICANT(S)

Name and address: *(Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country.)*	Telephone No.
The Trustees of the University of Pennsylvania 3160 Chestnut Street Suite 200 Philadelphia, Pennsylvania 19104 US	215-573-4508
	Facsimile No. 215-898-9519
	Teleprinter No.
	Applicant's registration No. with the Office

State *(that is, country)* of nationality:	State *(that is, country)* of residence:
US	US

Name and address: *(Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country.)*

Wilson, James M.
1350 N. Avignon Drive
Gladwyne, Pennsylvania 19035 US

State *(that is, country)* of nationality:	State *(that is, country)* of residence:
US	US

Name and address: *(Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country.)*

Gao, Guangping
408 Yorkshire Road
Rosemont, Pennsylvania 19010 US

State *(that is, country)* of nationality:	State *(that is, country)* of residence:
US	US

[X] Further applicants are indicated on a continuation sheet.

Sheet No. 2.

International application No.
PCT/US02/33645

Continuation of Box No. II APPLICANT(S)

If none of the following sub-boxes is used, this sheet should not be included in the demand.

Name and address: *(Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country.)*

Roy, Soumitra
240 Pugh Road
Wayne, Pennsylvania 19087 US

State *(that is, country)* of nationality:	State *(that is, country)* of residence:
US	US

Name and address: *(Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country.)*

State *(that is, country)* of nationality:	State *(that is, country)* of residence:

Name and address: *(Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country.)*

State *(that is, country)* of nationality:	State *(that is, country)* of residence:

Name and address: *(Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country.)*

State *(that is, country)* of nationality:	State *(that is, country)* of residence:

☐ Further applicants are indicated on another continuation sheet.

International application No.
PCT/US02/33645

Box No. III AGENT OR COMMON REPRESENTATIVE; OR ADDRESS FOR CORRESPONDENCE

The following person is ☒ agent ☐ common representative

and ☒ has been appointed earlier and represents the applicant(s) also for international preliminary examination.

☐ is hereby appointed and any earlier appointment of (an) agent(s)/common representative is hereby revoked.

☐ is hereby appointed, specifically for the procedure before the International Preliminary Examining Authority, in addition to the agent(s)/common representative appointed earlier.

Name and address: *(Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country.)*

Kodroff, Cathy A.
Howson and Howson
Spring House Corporate Center
P.O. Box 457
Spring House, Pennsylvania 19477 US

Telephone No.
215-540-9200

Facsimile No.
215-540-5818

Teleprinter No.

Agent's registration No. with the Office
33,980

☐ Address for correspondence: Mark this check-box where no agent or common representative is/has been appointed and the space above is used instead to indicate a special address to which correspondence should be sent.

Box No. IV BASIS FOR INTERNATIONAL PRELIMINARY EXAMINATION

Statement concerning amendments:*

1. The applicant wishes the international preliminary examination to start on the basis of:

☒ the international application as originally filed

the description ☐ as originally filed
☐ as amended under Article 34

the claims ☐ as originally filed
☐ as amended under Article 19 (together with any accompanying statement)
☐ as amended under Article 34

the drawings ☐ as originally filed
☐ as amended under Article 34

2. ☐ The applicant wishes any amendment to the claims under Article 19 to be considered as reversed.

3. ☐ The applicant wishes the start of the international preliminary examination to be postponed until the expiration of 20 months from the priority date unless the International Preliminary Examining Authority receives a copy of any amendments made under Article 19 or a notice from the applicant that he does not wish to make such amendments (Rule 69.1(d)). *(This check-box may be marked only where the time limit under Article 19 has not yet expired.)*

* Where no check-box is marked, international preliminary examination will start on the basis of the international application as originally filed or, where a copy of amendments to the claims under Article 19 and/or amendments of the international application under Article 34 are received by the International Preliminary Examining Authority before it has begun to draw up a written opinion or the international preliminary examination report, as so amended.

Language for the purposes of international preliminary examination: English

☒ which is the language in which the international application was filed.

☐ which is the language of a translation furnished for the purposes of international search.

☐ which is the language of publication of the international application.

☐ which is the language of the translation (to be) furnished for the purposes of international preliminary examination.

Box No. V ELECTION OF STATES

The applicant hereby elects all eligible States *(that is, all States which have been designated and which are bound by Chapter II of the PCT)*

excluding the following States which the applicant wishes not to elect:

International application No.
PCT/US02/33645

Box No. VI CHECK LIST

The demand is accompanied by the following elements, in the language referred to in Box No. IV, for the purposes of international preliminary examination:			For International Preliminary Examining Authority use only: received	not received
1. translation of international application	:	sheets	☐	☐
2. amendments under Article 34	:	sheets	☐	☐
3. copy (or, where required, translation) of amendments under Article 19	:	sheets	☐	☐
4. copy (or, where required, translation) of statement under Article 19	:	sheets	☐	☐
5. letter	:	sheets	☐	☐
6. other *(specify)*	:	sheets	☐	☐

The demand is also accompanied by the item(s) marked below:

1. ☒ fee calculation sheet
2. ☐ original separate power of attorney
3. ☐ original general power of attorney
4. ☐ copy of general power of attorney; reference number, if any:
5. ☐ statement explaining lack of signature
6. ☐ sequence listings in computer readable form
7. ☐ tables in computer readable form related to sequence listings
8. ☐ other *(specify)*:

Box No. VII SIGNATURE OF APPLICANT, AGENT OR COMMON REPRESENTATIVE

Next to each signature, indicate the name of the person signing and the capacity in which the person signs (if such capacity is not obvious from reading the demand).

By: Cathy A Kodroff
Cathy A. Kodroff
Attorney for Applicants

For International Preliminary Examining Authority use only

1. *Date of actual receipt of DEMAND:*

2. Adjusted date of receipt of demand due to CORRECTIONS under Rule 60.1(b):

3. ☐ The date of receipt of the demand is AFTER the expiration of 19 months from the priority date and item 4 or 5, below, does not apply. ☐ The applicant has been informed accordingly.

4. ☐ The date of receipt of the demand is WITHIN the period of 19 months from the priority date as extended by virtue of Rule 80.5.

5. ☐ Although the date of receipt of the demand is after the expiration of 19 months from the priority date, the delay in arrival is EXCUSED pursuant to Rule 82.

For International Bureau use only

Demand received from IPEA on:

PATENT COOPERATION TREATY

From the INTERNATIONAL BUREAU

PCT

NOTIFICATION CONCERNING SUBMISSION OR TRANSMITTAL OF PRIORITY DOCUMENT

(PCT Administrative Instructions, Section 411)

To:

Kodroff, Cathy, A.
Howson and Howson
Spring House Corporate Center
P.O. Box 457
Spring House, PA 19477
United States of America

RECEIVED
APR 22 2003

Date of mailing (day/month/year) 12 April 2003 (12.04.03)	
Applicant's or agent's file reference UPN-02677PCT	IMPORTANT NOTIFICATION
International application No. PCT/US02/33645	International filing date (day/month/year) 20 November 2002 (20.11.02)
International publication date (day/month/year) Not yet published	Priority date (day/month/year) 21 November 2001 (21.11.01)
Applicant THE TRUSTEES OF THE UNIVERSITY OF PENNSYLVANIA et al	

1. The applicant is hereby notified of the date of receipt (except where the letters "NR" appear in the right-hand column) by the International Bureau of the priority document(s) relating to the earlier application(s) indicated below. Unless otherwise indicated by an asterisk appearing next to a date of receipt, or by the letters "NR", in the right-hand column, the priority document concerned was submitted or transmitted to the International Bureau in compliance with Rule 17.1(a) or (b).

2. This updates and replaces any previously issued notification concerning submission or transmittal of priority documents.

3. An asterisk(*) appearing next to a date of receipt, in the right-hand column, denotes a priority document submitted or transmitted to the International Bureau but not in compliance with Rule 17.1(a) or (b). In such a case, the attention of the applicant is directed to Rule 17.1(c) which provides that no designated Office may disregard the priority claim concerned before giving the applicant an opportunity, upon entry into the national phase, to furnish the priority document within a time limit which is reasonable under the circumstances.

4. The letters "NR" appearing in the right-hand column denote a priority document which was not received by the International Bureau or which the applicant did not request the receiving Office to prepare and transmit to the International Bureau, as provided by Rule 17.1(a) or (b), respectively. In such a case, the attention of the applicant is directed to Rule 17.1(c) which provides that no designated Office may disregard the priority claim concerned before giving the applicant an opportunity, upon entry into the national phase, to furnish the priority document within a time limit which is reasonable under the circumstances.

Priority date	Priority application No.	Country or regional Office or PCT receiving Office	Date of receipt of priority document
21 Nove 2001 (21.11.01)	60/331,951	US	18 Febr 2003 (18.02.03)
22 Marc 2002 (22.03.02)	60/366,798	US	25 Febr 2003 (25.02.03)

The International Bureau of WIPO 34, chemin des Colombettes 1211 Geneva 20, Switzerland	Authorized officer Sean TAYLOR (Fax 338 9090)
Facsimile No. (41-22) 338.90.90	Telephone No. (41-22) 338 9811

00270
PATENT TRADEMARK OFFICE

ASSIGNMENT

WHEREAS, James M. Wilson, Guangping Gao, and Soumitra Roy, hereinafter referred to as ASSIGNORS, have invented and are the owners of an invention entitled CHIMPANZEE ADENOVIRUS NUCLEIC ACID AND AMINO ACID SEQUENCES, VECTORS CONTAINING SAME, AND METHODS OF USE said invention being fully set forth and described in the application for Letters Patent of the United States which has been prepared in connection with said invention, and which has been filed on November 21, 2001 and given Application Number 60/331,951.

WHEREAS, The Trustees of the University of Pennsylvania, a corporation organized and existing under and by virtue of the laws of the Commonwealth of Pennsylvania and engaged in business at 3160 Chestnut Street, Suite 200, Philadelphia, PA 19104-6283, United States of America, hereinafter referred to as ASSIGNEE, is desirous of acquiring the said invention and any and all applications for Letters Patent and any and all Letters Patent of the United States and foreign countries which may be issued therefor;

NOW, THEREFORE, THIS IS TO WITNESS that in consideration of the sum of One Dollar ($1.00) paid to ASSIGNORS by ASSIGNEE, the receipt of which is hereby acknowledged, and for other good and valuable consideration, ASSIGNORS have sold, assigned and set over, and hereby do sell, assign and set over, unto ASSIGNEE, its successors and assigns, the entire right, title and interest for, to and within the United States and all foreign countries, in and to the aforesaid invention and in and to any application for Letters Patent which ASSIGNEE at its expense may cause to be filed in the United States, under an international treaty, or in any foreign country, and in any United States or foreign patent which may be granted on said invention, including any conversions, non-provisional applications, extensions, divisions, reissues, continuations or continuations-in-part.

And ASSIGNORS hereby authorize and request the Commissioner of Patents and Trademarks to issue any and all Letters Patent of the United States to ASSIGNEE, its successors and assigns, as the owner of all right, title and interest therein.

And for said consideration, it is hereby covenanted and agreed that at the request and expense of ASSIGNEE, its successors and assigns, ASSIGNORS will execute any further papers and do such other acts and things as may be necessary and proper to procure Letters Patent for said invention in the United States and in such foreign countries as ASSIGNEE, its successors and assigns may elect, and vest the full title thereto in ASSIGNEE, its successors and assigns.

IN WITNESS WHEREOF, this assignment has been executed on the dates indicated below.

Date 5-24-02 _____(SEAL)
James M. Wilson

Date 5-28-02 _____(SEAL)
Guangping Gao

Date 5-24-02 _____(SEAL)
Soumitra Roy

NOTARIAL CERTIFICATE

State of :
: ss.
County of :

Before me, the undersigned, a notary public, in and for the county aforesaid, on this 24th day of May 2002, personally appeared James M. Wilson, to me known to be the person described in and who executed the foregoing instrument and who acknowledged that she executed the same as her free act and deed and desired the same to be recorded as such.

NOTARIAL SEAL
JANET C. DEENEY Notary Public
City of Philadelphia, Phila. County
My Commission Expires April 8, 2004

_____
Notary Public

NOTARIAL CERTIFICATE

State of PA :
: ss.
County of Phila :

Before me, the undersigned, a notary public, in and for the county aforesaid, on this 28th day of May 2002, personally appeared Guangping Gao, to me known to be the person described in and who executed the foregoing instrument and who acknowledged that he executed the same as his free act and deed and desired the same to be recorded as such.

_____
Notary Public

NOTARIAL SEAL
JANET C. DEENEY, Notary Public
City of Philadelphia, Phila. County
My Commission Expires April 8, 2004

105

NOTARIAL CERTIFICATE

State of Pa :
: ss.
County of Phila :

Before me, the undersigned, a notary public, in and for the county aforesaid, on this 24th day of May 2002, personally appeared Soumitra Roy, to me known to be the person described in and who executed the foregoing instrument and who acknowledged that she executed the same as her free act and deed and desired the same to be recorded as such.

Janet C. Deeney
Notary Public

NOTARIAL SEAL
JANET C. DEENEY, Notary Public
City of Philadelphia, Phila. County
My Commission Expires April 8, 2004

UNITED STATES
PATENT AND
TRADEMARK OFFICE

AUGUST 12, 2002

PTAS

Chief Information Officer
Washington, DC 20231
www.uspto.gov

HOWSON AND HOWSON
CATHY A. KODROFF, ESQUIRE
SPRING HOUSE CORPORATE CENTER
BOX 457
SPRING HOUSE, PA 19477

RECEIVED
AUG 16 2002
TECHNOLOGY CENTER 2100

102120259A

UNITED STATES PATENT AND TRADEMARK OFFICE
NOTICE OF RECORDATION OF ASSIGNMENT DOCUMENT

THE ENCLOSED DOCUMENT HAS BEEN RECORDED BY THE ASSIGNMENT DIVISION OF THE U.S. PATENT AND TRADEMARK OFFICE. A COMPLETE MICROFILM COPY IS AVAILABLE AT THE ASSIGNMENT SEARCH ROOM ON THE REEL AND FRAME NUMBER REFERENCED BELOW.

PLEASE REVIEW ALL INFORMATION CONTAINED ON THIS NOTICE. THE INFORMATION CONTAINED ON THIS RECORDATION NOTICE REFLECTS THE DATA PRESENT IN THE PATENT AND TRADEMARK ASSIGNMENT SYSTEM. IF YOU SHOULD FIND ANY ERRORS OR HAVE QUESTIONS CONCERNING THIS NOTICE, YOU MAY CONTACT THE EMPLOYEE WHOSE NAME APPEARS ON THIS NOTICE AT 703-308-9723. PLEASE SEND REQUEST FOR CORRECTION TO: U.S. PATENT AND TRADEMARK OFFICE, ASSIGNMENT DIVISION, BOX ASSIGNMENTS, CG-4, 1213 JEFFERSON DAVIS HWY, SUITE 320, WASHINGTON, D.C. 20231.

RECORDATION DATE: 06/10/2002

REEL/FRAME: 012970/0955
NUMBER OF PAGES: 4

BRIEF: ASSIGNMENT OF ASSIGNOR'S INTEREST (SEE DOCUMENT FOR DETAILS).

ASSIGNOR:
WILSON, JAMES M.

DOC DATE: 05/24/2002

ASSIGNOR:
GAO, GUANGPING

DOC DATE: 05/28/2002

ASSIGNOR:
ROY, SOUMITRA

DOC DATE: 05/24/2002

ASSIGNEE:
TRUSTEE OF THE UNIVERSITY OF
PENNSYLVANIA, THE
3160 CHESTNUT STREET
SUITE 200
PHILADELPHIA, PENNSYLVANIA
19104-
6283

SERIAL NUMBER: 60331951
PATENT NUMBER:

FILING DATE: 11/21/2001
ISSUE DATE:

00270
PATENT TRADEMARK OFFICE

ASSIGNMENT

WHEREAS, James M. Wilson, Guangping Gao, and Soumitra Roy, hereinafter referred to as ASSIGNORS, have invented and are the owners of an invention entitled CHIMPANZEE ADENOVIRUS NUCLEIC ACID AND AMINO ACID SEQUENCES, VECTORS CONTAINING SAME, AND METHODS OF USE said invention being fully set forth and described in the application for Letters Patent of the United States which has been prepared in connection with said invention, and which has been filed on March 22, 2002 and given Application Number 60/366,798.

WHEREAS, The Trustees of the University of Pennsylvania, a corporation organized and existing under and by virtue of the laws of the Commonwealth of Pennsylvania and engaged in business at 3160 Chestnut Street, Suite 200, Philadelphia, PA 19104-6283, United States of America, hereinafter referred to as ASSIGNEE, is desirous of acquiring the said invention and any and all applications for Letters Patent and any and all Letters Patent of the United States and foreign countries which may be issued therefor;

NOW, THEREFORE, THIS IS TO WITNESS that in consideration of the sum of One Dollar ($1.00) paid to ASSIGNORS by ASSIGNEE, the receipt of which is hereby acknowledged, and for other good and valuable consideration, ASSIGNORS have sold, assigned and set over, and hereby do sell, assign and set over, unto ASSIGNEE, its successors and assigns, the entire right, title and interest for, to and within the United States and all foreign countries, in and to the aforesaid invention and in and to any application for Letters Patent which ASSIGNEE at its expense may cause to be filed in the United States, under an international treaty, or in any foreign country, and in any United States or foreign patent which may be granted on said invention, including any conversions, non-provisional applications, extensions, divisions, reissues, continuations or continuations-in-part.

And ASSIGNORS hereby authorize and request the Commissioner of Patents and Trademarks to issue any and all Letters Patent of the United States to ASSIGNEE, its successors and assigns, as the owner of all right, title and interest therein.

And for said consideration, it is hereby covenanted and agreed that at the request and expense of ASSIGNEE, its successors and assigns, ASSIGNORS will execute any further papers and do such other acts and things as may be necessary and proper to procure Letters Patent for said invention in the United States and in such foreign countries as ASSIGNEE, its successors and assigns may elect, and vest the full title thereto in ASSIGNEE, its successors and assigns.

IN WITNESS WHEREOF, this assignment has been executed on the dates indicated below.

Date 5-24-02 _____ (SEAL)
James M. Wilson

Date 5/28-02 _____ (SEAL)
Guangping Gao

Date 5.24.02 _____ (SEAL)
Soumitra Roy

NOTARIAL CERTIFICATE

State of Pa :
: ss.
County of Phila :

Before me, the undersigned, a notary public, in and for the county aforesaid, on this 23rd day of May 2002, personally appeared James M. Wilson, to me known to be the person described in and who executed the foregoing instrument and who acknowledged that she executed the same as her free act and deed and desired the same to be recorded as such.

NOTARIAL SEAL
JANET C. DEENEY, Notary Public
City of Philadelphia, Phila. County
My Commission Expires April 8, 2004

Janet C Deeney
Notary Public

NOTARIAL CERTIFICATE

State of Pa :
: ss.
County of Phila :

Before me, the undersigned, a notary public, in and for the county aforesaid, on this 28th day of May 2002, personally appeared Guangping Gao, to me known to be the person described in and who executed the foregoing instrument and who acknowledged that he executed the same as his free act and deed and desired the same to be recorded as such.

Janet C. Deeney
Notary Public

NOTARIAL SEAL
JANET C. DEENEY, Notary Public
City of Philadelphia, Phila. County
My Commission Expires April 8, 2004

409

NOTARIAL CERTIFICATE

State of Pa :
: ss.
County of Phila :

Before me, the undersigned, a notary public, in and for the county aforesaid, on this 24th day of May 2002, personally appeared **Soumitra Roy**, to me known to be the person described in and who executed the foregoing instrument and who acknowledged that she executed the same as her free act and deed and desired the same to be recorded as such.

Janet C Deeney
Notary Public

NOTARIAL SEAL
JANET C. DEENEY, Notary Public
City of Philadelphia, Phila. County
My Commission Expires April 8, 2004

UNITED STATES PATENT AND TRADEMARK OFFICE

AUGUST 12, 2002

PTO
AUG 16 2002

Chief Information Officer
Washington, DC 20231
www.uspto.gov

HOWSON AND HOWSON
CATHY A. KODROFF, ESQ.
SPRINGHOUSE CORPORATE CENTER, BX 457
SPRING HOUSE, PA 19477

102120258A

UNITED STATES PATENT AND TRADEMARK OFFICE
NOTICE OF RECORDATION OF ASSIGNMENT DOCUMENT

THE ENCLOSED DOCUMENT HAS BEEN RECORDED BY THE ASSIGNMENT DIVISION OF THE U.S. PATENT AND TRADEMARK OFFICE. A COMPLETE MICROFILM COPY IS AVAILABLE AT THE ASSIGNMENT SEARCH ROOM ON THE REEL AND FRAME NUMBER REFERENCED BELOW.

PLEASE REVIEW ALL INFORMATION CONTAINED ON THIS NOTICE. THE INFORMATION CONTAINED ON THIS RECORDATION NOTICE REFLECTS THE DATA PRESENT IN THE PATENT AND TRADEMARK ASSIGNMENT SYSTEM. IF YOU SHOULD FIND ANY ERRORS OR HAVE QUESTIONS CONCERNING THIS NOTICE, YOU MAY CONTACT THE EMPLOYEE WHOSE NAME APPEARS ON THIS NOTICE AT 703-308-9723. PLEASE SEND REQUEST FOR CORRECTION TO: U.S. PATENT AND TRADEMARK OFFICE, ASSIGNMENT DIVISION, BOX ASSIGNMENTS, CG-4, 1213 JEFFERSON DAVIS HWY, SUITE 320, WASHINGTON, D.C. 20231.

RECORDATION DATE: 06/10/2002
REEL/FRAME: 012970/0963
NUMBER OF PAGES: 4

BRIEF: ASSIGNMENT OF ASSIGNOR'S INTEREST (SEE DOCUMENT FOR DETAILS).

ASSIGNOR:
WILSON, JAMES M.
DOC DATE: 05/24/2002

ASSIGNOR:
GAO, GUANGPING
DOC DATE: 05/28/2002

ASSIGNOR:
ROY, SOUMITRA
DOC DATE: 05/24/2002

ASSIGNEE:
TRUSTEES OF THE UNIVERSITY OF
PENNSYLVANIA, THE
3160 CHESTNUT STREET, SUITE 200
PHILADELPHIA, PENNSYLVANIA

19104-6283

SERIAL NUMBER: 60366798
FILING DATE: 03/22/2002
PATENT NUMBER:
ISSUE DATE:

411

SUPERINDUSTRIA Y COMERCIO
Radicacion : 04056841 00000000 Folios: 418
Fecha (MMD): 2004-06-17 16:14:34
Tramite : 611 PCT-PATENT 379 FASE NACI 411 PRESENTAC
Dependencia: 2020 DIVISION DE NUEVAS CREACIONES

REQUISITOS NECESARIOS PARA PRESENTACION Y AGILIZACION DEL TRAMITE

	USUARIO	RADICADOR
PATENTE DE INVENCIÓN / MODELO DE UTILIDAD	( )	( )
- Indicación de que se solicita la concesión de una patente.	( )	( )
- Datos de identificación del solicitante o de la persona que se presenta la solicitud.	( )	( )
- Descripción de la invención.	( )	( )
- Dibujos de ser estos pertinentes.	( )	( )
- Comprobante de Pago de las tasas establecidas.		
COMPLETA ( )	INCOMPLETA ( )	( )
DISEÑO INDUSTRIAL ( )		
- Indicación de que se solicita el registro de Diseño Industrial	( )	( )
- Datos de identificación del solicitante o de la persona que presenta	( )	( )
- la solicitud.		
- Representación gráfica o fotográfica del Diseño Industrial o muestra del Material que incorpora el diseño.	( )	( )
Comprobante de pago de las tasas establecidas	( )	( )
COMPLETA ( )	INCOMPLETA ( )	( )
ESQUEMA DE TRAZADO ( )		
- Indicación de que solicita el registro de un Esquema de trazado.	( )	( )
- Datos de identificación del solicitante o de la persona que presenta La solicitud.	( )	( )
Representación gráfica del esquema de trazado	( )	( )
Comprobante de pago de las tasas establecidas.	( )	( )
COMPLETA ( )	INCOMPLETA ( )	( )

Firma Responsable:

Fecha:

412

2020
Bogotá D C

Doctor(a)
JIMENA ESCOBAR URIBE.
Apoderado(a)

Oficio N° 06630

ASUNTO:

Radicación N°	04-56841
Solicitud Internacional	PCT/US02/33645
Fecha inicio fase nacional	21/06/2004
Trámite	011
Evento	379
Actuación	416
Oficio N°	
Folios	001

Como resultado del examen de forma, realizado con fundamento en el artículo 38 de la Decisión 486 de la Comisión de la Comunidad Andina, artículo 27 del Tratado de Cooperación en Materia de Patentes (PCT), regla 51 bis del Reglamento y Circular Única de la Superintendencia de Industria y Comercio, se le comunica que:

- La solicitud no contiene el poder debidamente otorgado, con el lleno de requisitos exigidos por los artículos 63 y subsiguientes del Código de Procedimiento Civil. Adicionalmente deberá presentar el documento que acredita la existencia y representación legal de la sociedad solicitante, cuando éste fuere una persona jurídica.

En cumplimiento de lo dispuesto por el artículo 39 de la Decisión 486, se le requiere para que dentro del término de dos meses siguientes a la fecha de notificación del presente oficio, complete los requisitos anteriormente enunciados. En caso de no dar cumplimiento al presente requerimiento, la solicitud se considerará abandonada y perderá su prelación.

Notifíquese el presente oficio conforme a lo dispuesto en el numeral 5.2, literal e), del capítulo quinto del título primero de la Circular Única N° 10 de 2001.

ALIX CARMENZA CESPEDES DE VERGEL
Jefe de la División de Nuevas Creaciones.

El anterior requerimiento fue notificado por fijación en lista de fecha, 16 JUL. 2004
202051