

(12) SOLICITUD INTERNACIONAL PUBLICADA EN VIRTUD DEL TRATADO DE COOPERACIÓN
EN MATERIA DE PATENTES (PCT)

(19) Organización Mundial de la Propiedad
Intelectual
Oficina internacional



(43) Fecha de publicación internacional
8 de Febrero de 2007 (08.02.2007)

PCT

(10) Número de Publicación Internacional
WO 2007/014971 A2

(51) Clasificación Internacional de Patentes: Sin clasificar

(21) Número de la solicitud internacional:
PCT/ES2006/000455

(22) Fecha de presentación internacional:
1 de Agosto de 2006 (01.08.2006)

(25) Idioma de presentación: español

(26) Idioma de publicación: español

(30) Datos relativos a la prioridad:
P200501932 2 de Agosto de 2005 (02.08.2005) ES

(71) Solicitante (para todos los Estados designados salvo US):
PHARMA MAR, S.A. [ES/ES]; Polígono Industrial La Mina-Norte, Avda. De Los Reyes, 1, E-28770 Colmenar Viejo - Madrid (ES).

(72) Inventores; e

(75) Inventores/Solicitantes (para US solamente): **RAMOS CASTRO, Angelina** [ES/ES]; Universidad de Oviedo, Área De Microbiología, Departamento De Biología Funcional, Facultad De Medicina, E-33006 Oviedo (ES). **FERNÁNDEZ BRAÑA, Alfredo** [ES/ES]; Universidad de Oviedo, Área De Microbiología, Departamento De Biología Funcional, Facultad De Medicina, E-33006 Oviedo (ES). **LOMBÓ BRUGOS, Felipe** [ES/ES]; Universidad de Oviedo, Área De Microbiología, Departamento De Biología Funcional, Facultad De Medicina, E-33006 Oviedo (ES). **MÉNDEZ FERNÁNDEZ, Carmen** [ES/ES]; Universidad de Oviedo, Área De Microbiología, Departamento De Biología Funcional, Facultad De Medicina, E-33006 Oviedo (ES). **SALAS FERNÁNDEZ, José Antonio** [ES/ES]; Universidad de Oviedo, Área De Microbiología, Departamento De Biología Funcional, Facultad De Medicina, E-33006 Oviedo (ES). **VELASCO IGLESIAS,**

Ana [ES/ES]; Pharma Mar, S.A., Polígono Industrial La Mina-norte, Avda. De Los Reyes, 1, E-28770 Colmenar Viejo - Madrid (ES). **SCHLEISSNER SÁNCHEZ, Carmen** [ES/ES]; Pharma Mar, S.A., Polígono Industrial La Mina-norte, Avda. De Los Reyes, 1, E-28770 Colmenar Viejo - Madrid (ES).

(74) Mandatario: **ARIAS SANZ, Juan**; ABG Patentes, S.L., Orense, 68 7º, E-28020 Madrid (ES).

(81) Estados designados (a menos que se indique otra cosa, para toda clase de protección nacional admisible): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LV, LY, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Estados designados (a menos que se indique otra cosa, para toda clase de protección regional admisible): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), euroasiática (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), europea (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Publicada:

— sin informe de búsqueda internacional, será publicada nuevamente cuando se reciba dicho informe

Para códigos de dos letras y otras abreviaturas, véase la sección "Guidance Notes on Codes and Abbreviations" que aparece al principio de cada número regular de la Gaceta del PCT.

(54) Title: GENES INVOLVED IN THE BIOSYNTHESIS OF THIOCORALINE AND HETEROLOGOUS PRODUCTION OF SAME

(54) Título: GENES INVOLUCRADOS EN LA BIOSÍNTESIS DE TIOCORALINA Y PRODUCCIÓN HETERÓLOGA DE LA MISMA

(57) Abstract: The invention relates to genes involved in the biosynthesis of thiocoraline and to the heterologous production of same. According to the invention, the cluster of genes responsible for the biosynthesis of thiocoraline was identified and cloned. Said gene cluster can be used in the heterologous production of thiocoraline which has an antitumour and antibacterial activity.

(57) Resumen: Se ha identificado y clonado el grupo de genes (cluster) responsables de la biosíntesis de tiocoralina. Dicho grupo de genes puede ser utilizado en la producción heteróloga de tiocoralina. La tiocoralina tiene actividad antitumoral y antibacteriana.



WO 2007/014971 A2

GENES INVOLUCRADOS EN LA BIOSÍNTESIS DE TIOCORALINA Y PRODUCCIÓN HETERÓLOGA DE LA MISMA

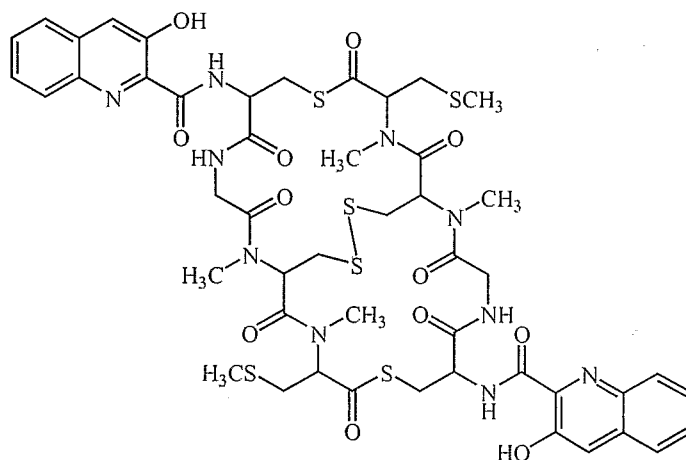
Campo de la invención

5

La presente invención está relacionada con el grupo de genes responsables de la biosíntesis de tiocoralina y su uso en la producción heteróloga de tiocoralina.

10 Antecedentes de la invención

La tiocoralina (I)



15

(I)

es un tiodepsipéptido ciclodimérico aislado de un actinomiceto marino, en concreto de la familia Micromonosporaceae (Pérez Baz et al., J. Antibiotics, 50(9), 738-741, 1997; Romero et al., J. Antibiotics, 50(9), 734-737, 1997). Aunque se ha descrito que la tiocoralina se obtiene de *Micromonospora marina* o *Micromonospora* sp. L-13-ACM-092, posteriores estudios han puesto de manifiesto que el compuesto también puede ser aislado del actinomiceto *Micromonospora* sp. ML1, el cual fue aislado de un molusco marino encontrado en la costa del

25

Océano Índico, en Mozambique (Espliego, F. Tesis Doctoral, 1996, Universidad de León; de la Calle, F. Tesis Doctoral, 1998, Universidad Autónoma de Madrid).

5 Estudios *in vitro* han demostrado la capacidad de la tiocoralina para inhibir el crecimiento de líneas celulares de diferentes tipos de tumores sólidos, tales como melanoma, cáncer de mama, pulmón no microcítico y colon. Asimismo, la tiocoralina ha mostrado poseer una marcada actividad antitumoral en ensayos *in vivo* frente a xenoinjertos
10 (xenografts) de carcinomas humanos (Faircloth et al. Eur. J. Cancer, 33, 175, 1997 (abstract)). Además, la tiocoralina muestra actividad antibacteriana frente a bacterias Gram-positivas.

Aunque la obtención de tiocoralina a partir de dicho actinomiceto marino (*Micromonospora sp.* ML1) es factible a pequeña escala, a gran
15 escala dicha obtención se halla limitada debido a la variabilidad en la producción que se observa con ese microorganismo. Efectivamente, la producción de tiocoralina a partir de dicho organismo es un proceso que consume mucho tiempo debido a la baja tasa de crecimiento de ese
20 organismo y muestra importantes fluctuaciones en los rendimientos de producción en diferentes lotes.

Por tanto, debido a que, por una parte, la obtención de tiocoralina a partir de dicho actinomiceto marino se halla bastante limitada, y, por
25 otra parte, al hecho de que la molécula de tiocoralina posee una estructura compleja y su síntesis puede ser complicada a nivel industrial, es deseable entender las bases genéticas de su biosíntesis con el fin de crear unos medios para que influyan en su obtención de una manera dirigida. Ello podría dar lugar a un incremento en las
30 cantidades de tiocoralina producidas, dado que las cepas productoras naturales generalmente producen el producto en baja concentración y de forma muy desigual. Asimismo, también podría permitir la

producción de tiocoralina en huéspedes que de forma natural no producen este compuesto.

El desarrollo de la tecnología del ADN recombinante ha abierto un
5 interesante campo de investigación para la generación y producción de
compuestos bioactivos mediante la manipulación de genes implicados
en la biosíntesis de tales compuestos bioactivos, principalmente de
bacterias del grupo de los actinomicetos. Estas técnicas pueden ser
usadas para mejorar la producción de compuestos naturales ya
10 conocidos, pues las cepas naturales suelen producir bajas
concentraciones del metabolito de interés.

Asimismo, la expresión heteróloga del agrupamiento (cluster) de
genes involucrados en la biosíntesis de tiocoralina en otros
15 actinomicetos más apropiados para fermentación y manipulación
genética permitiría producir dicho compuesto con rendimientos más
reproducible en tiempos de fermentación más cortos.

Como es conocido, numerosas bacterias y hongos sintetizan una
20 amplia variedad de péptidos biológicamente activos de origen no
ribosomal, incluyendo antitumorales, antibióticos, etc. La biosíntesis de
esta familia de compuestos se lleva a cabo por péptido sintetasas no
ribosomales (NRPSs), que son enzimas multifuncionales con una
organización modular de dominios catalíticos. Cada uno de estos
25 módulos lleva a cabo un ciclo de elongación, es decir, activa e incorpora
un aminoácido concreto en la estructura final del compuesto. Un
módulo mínimo está constituido por tres dominios: (i) un dominio de
adenilación (A, de aproximadamente 550 aminoácidos) que se encarga
de seleccionar un determinado aminoácido y generar la versión
30 aminoacil adenilada del mismo mediante el empleo de ATP; (ii) un
dominio transportador peptidil (P, de aproximadamente 80
aminoácidos) que contiene un grupo prostético de fosfopanteteína (PP)
que actúa como cofactor y se une por enlace covalente al dominio P;

este dominio se encarga de fijar el aminoácido adenilado activado antes de pasar a siguientes centros de reacción; y (iii) un dominio de condensación (C, de aproximadamente 450 aminoácidos) que genera un nuevo enlace peptídico entre dos restos aminoacil adenilados que se localizan en dos dominios P consecutivos. Los dominios C están ausentes en los módulos que activan el primer aminoácido del sistema. Algunas NRPSs poseen unos dominios extra para llevar a cabo actividades específicas, como epimerizaciones dando lugar a D-aminoácidos, metilaciones tipo N- o C-, circularizaciones actuando sobre los aminoácidos L-Cys o L-Ser. Generalmente, un último dominio, situado después del último módulo, se encarga de la liberación del enzima intermediario, generando un péptido lineal o cíclico. Como regla general, la estructura de los diferentes módulos refleja la secuencia final de aminoácidos del péptido producto. Esta regla de colinealidad permite la asignación a cada módulo de una función específica de activación en una NRPS. Información sobre NRPSs puede encontrarse, por ejemplo, en Quing-Tao, S. *et al.*, 2004. Dissecting and Exploiting Nonribosomal Peptide Synthetases. *Acta Biochimica et Biophysica Sinica*, 36 (4): 243-249.

Compendio de la invención

Un objetivo importante de la presente invención consiste en aislar y caracterizar la secuencia de nucleótidos completa que codifica las proteínas responsables de la producción de tiocoralina. A partir de ahí, se puede aislar y determinar la función de las secuencias aminoacídicas que comprenden las proteínas involucradas en la biosíntesis de tiocoralina. Este objetivo puede alcanzarse proporcionando una nueva molécula de ácido nucleico, aislada y, opcionalmente, purificada, que codifica todas las proteínas relacionadas con la ruta biosintética completa de producción de tiocoralina.

Los inventores han sido capaces de identificar y clonar todos los genes responsables de la biosíntesis de tiocoralina, es decir, el cluster de genes implicados en la biosíntesis de tiocoralina, proporcionando las bases genéticas para mejorar y manipular de forma dirigida la producción de este compuesto.

Mediante el empleo de oligonucleótidos iniciadores derivados de secuencias consenso de los dominios de adenilación de péptido sintetasas no ribosomales (NRPSs) se amplificaron, mediante la reacción en cadena de la polimerasa (PCR), 6 fragmentos del cromosoma de *Micromonospora* sp. ML1, conteniendo todos ellos fragmentos de los dominios de adenilación de las NRPSs putativas (hipotéticas) denominados PSV1, PSV2, PSV3, PSV4, PSV5 y PSV6 (Ejemplo 3). La inactivación, por inserción, de dichos dominios de adenilación puso de manifiesto que dos de ellos (PSV2 y PSV5) generaban mutantes no-productores de tiocoralina, lo que indicaba que estaban implicados en la biosíntesis de tiocoralina (Ejemplos 7 y 10).

La secuenciación de una región de ADN de aproximadamente 64,6 kilobases (kb) (SEQ ID NO: 1) reveló la presencia de 36 marcos de lectura abiertos (ORFs) completos y otros 2 ORFs incompletos (Ejemplo 12, Tabla 1). La expresión heteróloga de una región de 53 kb aproximadamente, conteniendo 26 de dichas ORFs, en *Streptomyces coelicolor*, *Streptomyces albus* y *Streptomyces lividans* condujo a la producción de tiocoralina en dichos actinomicetos (Ejemplo 19).

El cluster de genes responsable de la biosíntesis de la tiocoralina se representa esquemáticamente en la Figura 1. Sorprendentemente, el cluster de genes de tiocoralina contiene más genes codificantes de NRPSs que los que cabría esperar en base al número de aminoácidos del esqueleto peptídico. Algunas de las proteínas identificadas están implicadas en la formación de la estructura peptídica de la tiocoralina, como, por ejemplo, varias de las NRPSs identificadas como Tio12, Tio17,

Tio18, Tio19, Tio20, Tio21, Tio22, Tio27 y Tio28. Las proteínas identificadas como Tio20 y Tio21 constituyen probablemente las NRPSs implicadas en la biosíntesis del esqueleto de la tiocoralina y, probablemente, otras dos NRPSs, identificadas como Tio27 y Tio 28
5 podrían ser las responsables de la biosíntesis de un péptido pequeño que podría estar implicado en la regulación de la biosíntesis de tiocoralina en *Micromonospora* sp. ML1. Aparecen también varias proteínas que pueden estar relacionadas con procesos de resistencia, como Tio5, Tio6 y Tio23. Los posibles reguladores de la ruta de
10 tiocoralina identificados en la región secuenciada corresponden a Tio3, Tio4, Tio7, Tio24 y Tio25. Aparecen por último también varias proteínas relacionadas con la generación de la unidad iniciadora 3-hidroxi-quinaldato, Tio8, Tio9, Tio10 y Tio11. Los genes cuya interrupción génica genera un fenotipo no productor de tiocoralina se encuentran
15 señalizados en la Figura 1 mediante un asterisco (*tio20*, *tio27* y *tio28*).

Por tanto, la presente invención se relaciona con la identificación y clonaje del cluster de los genes responsables de la biosíntesis de tiocoralina. Dicho cluster de genes responsables de la biosíntesis de
20 tiocoralina y su expresión en una célula hospedadora apropiada permite la producción eficiente de tiocoralina.

En consecuencia, en un aspecto, la invención se relaciona con una molécula de ácido nucleico aislada que comprende una secuencia
25 de nucleótidos que codifica al menos una proteína de la ruta biosintética de producción de tiocoralina, o un fragmento biológicamente activo de la misma.

En otro aspecto, la invención se relaciona con una composición
30 que comprende, al menos, una molécula de ácido nucleico proporcionada por esta invención.

En otro aspecto, la invención se relaciona con una sonda que comprende una molécula de ácido nucleico proporcionada por esta invención o un fragmento de la misma.

5 En otro aspecto, la invención se relaciona con un vector que comprende una molécula de ácido nucleico proporcionada por esta invención.

10 En otro aspecto, la invención se relaciona con una célula hospedadora trasformada o transfectada con un vector proporcionada por esta invención.

15 En otro aspecto, la invención se relaciona con una proteína codificada por una molécula de ácido nucleico proporcionada por esta invención.

20 En otro aspecto, la invención se relaciona con un método para producir una proteína implicada en la biosíntesis de tiocoralina, que comprende el empleo de un organismo productor de tiocoralina cuyo genoma ha sido manipulado.

25 En otro aspecto, la invención se relaciona con un procedimiento, basado en la utilización de genes responsables de la biosíntesis de tiocoralina de *Micromonospora* sp. ML1, para la producción de tiocoralina en otro actinomiceto.

Descripción de las figuras

30 Figura 1. Representación esquemática del cluster de genes de la tiocoralina y de los genes que los rodean incluyendo la organización génica de la zona del cromosoma de *Micromonospora* sp. ML1 secuenciada. Se muestran los sitios de restricción utilizados para la

construcción de los plásmidos para la expresión heteróloga del cluster de genes de tiocoralina.

Figura 2. Representación esquemática de los cósmidos cosV33-D12 y pCT2c. ori: origen de replicación para *E. coli*. SCP2: origen de replicación para *Streptomyces*. *aac(3)IV*: gen de resistencia a apramicina. *neo*: gen de resistencia a neomicina. *bla*: gen de resistencia a ampicilina. SV40 ori: origen eucariótico para replicación episómica.

Figura 3. Esquema de clonaciones realizadas para la construcción del plásmido pFL1036. ori: origen de replicación para *E. coli*. M13 ori: origen de replicación para el fago M13. oriT: origen de transferencia conjugativa. *lacZ*: gen de la beta-galactosidasa. *kan^R*: gen de resistencia a kanamicina. *aac(3)IV*: gen de resistencia a apramicina. *bla*: gen de resistencia a ampicilina.

Figura 4. Esquema de clonaciones realizadas para la construcción del plásmido pFL1041. ori: origen de replicación para *E. coli*. SCP2: origen de replicación para *Streptomyces*. oriT: origen de transferencia conjugativa. *lacZ*: gen de la beta-galactosidasa. *aac(3)IV*: gen de resistencia a apramicina.

Figura 5. Esquema de clonaciones realizadas para la construcción del plásmido pAR15AT. ori p15A: origen de replicación para *E. coli*. oriT: origen de transferencia conjugativa. *intφC31*: gen de la integrasa del fago φC31. attP: sitio de recombinación sitio-específica. *kan^R*: gen de resistencia a kanamicina. *aac(3)IV*: gen de resistencia a apramicina. K: sitio de corte tratado con el fragmento Klenow de la ADN polimerasa de *E. coli*.

Figura 6. Esquema de clonaciones realizadas para la construcción del plásmido pAPR. ori p15A: origen de replicación para *E. coli*. oriT: origen de transferencia conjugativa. ori M13: origen de replicación del fago M13. ori: origen de replicación para *E. coli*. *lacZ*: gen de la beta-galactosidasa. *lacI*: gen del represor del operón lactosa. *intφC31*: gen de la integrasa del fago φC31. attP: sitio de recombinación sitio-específica. *kan^R*: gen de resistencia a kanamicina. *aac(3)IV*: gen de resistencia a apramicina. *K*: sitio de corte tratado con el fragmento Klenow de la ADN polimerasa de *E. coli*. *P_{ermE}*: promotor del gen *ermE*.

10

Figura 7. Representación de los plásmidos pFL1048, pFL1048r y pFL1049. ori p15A: origen de replicación para *E. coli*. oriT: origen de transferencia conjugativa. *intφC31*: gen de la integrasa del fago φC31. attP: sitio de recombinación sitio-específica. *aac(3)IV*: gen de resistencia a apramicina.

15

Figura 8A. Cromatograma de HPLC de un extracto de un cultivo de *Streptomyces albus* (pFL1049) después de 7 días de crecimiento en medio R5A. Se muestra el pico correspondiente a tiocoralina y su tiempo de retención, 27 minutos.

20

Figura 8B. Espectro de absorción UV del producto (tiocoralina) presente en el pico de 27 minutos mostrado en la Figura 8A.

Figura 8C. Espectro de masas del producto (tiocoralina) presente en el pico de 27 minutos mostrado en la Figura 8A.

25

Descripción detallada de la invención

De acuerdo con la presente invención, se proporciona una nueva molécula de ácido nucleico, aislada y, opcionalmente purificada, que

30

codifica la totalidad o parte de las proteínas involucradas en la ruta biosintética completa de producción de tiocoralina.

Por tanto, en un aspecto, la invención se relaciona con una
5 molécula de ácido nucleico, en adelante, molécula de ácido nucleico de la invención, preferentemente una molécula de ácido nucleico aislada, opcionalmente purificada, que comprende una secuencia de nucleótidos que codifica al menos una proteína de la ruta biosintética de producción de tiocoralina, o un fragmento biológicamente activo de la
10 misma. En general, dicha proteína de la ruta biosintética de producción de tiocoralina es un péptido sintetasa no ribosomal (NRPS). Las NRPSs son las responsables de la biosíntesis de tiocoralina.

Tal como aquí se utiliza, la expresión “fragmento biológicamente
15 activo”, aplicado a una proteína de la ruta biosintética de producción de tiocoralina, se refiere a una parte de la estructura de la proteína que retiene la función activa de la proteína longitud completa. Dichos fragmentos biológicamente activos pueden ser codificados por las regiones correspondientes de la molécula de ácido nucleico de la
20 invención. El tamaño de dichas regiones de la molécula de ácido nucleico de la invención puede variar dentro de un amplio intervalo; no obstante, en una realización particular, dichas regiones pueden tener al menos 10, 15, 20, 25, 50, 100, 1.000, 2.500, 5.000, 10.000, 20.000, 25.000 o más nucleótidos de longitud. Normalmente dichas regiones
25 tienen entre 100 y 10.000 nucleótidos de longitud, preferentemente entre 100 y 7.500, y son biológicamente funcionales, es decir, pueden codificar un fragmento biológicamente activo de una proteína de la ruta biosintética de producción de tiocoralina.

30 La molécula de ácido nucleico de la invención puede ser una molécula de ácido desoxirribonucleico (ADN) o de ácido ribonucleico (ARN). Asimismo, la molécula de ácido nucleico de la invención puede ser una molécula de ácido nucleico monocatenario o bien una molécula

de ácido nucleico bicatenario derivada. Ejemplos ilustrativos, no limitativos, de moléculas de ácido nucleico de la invención moléculas de ADN genómico (ADNg), moléculas de ARN mensajero (ARNm) y moléculas de ADN complementarias (ADNc) a moléculas de ARNm.

5

Quedan incluidos dentro del ámbito de la presente invención los mutantes y variantes de la molécula de ácido nucleico de la invención. Entre dichos mutantes y variantes se incluyen las moléculas de ácido nucleico de la invención en las que, al menos, un nucleótido ha sido
10 alterado, sustituido, eliminado o insertado. A modo ilustrativo, los mutantes y variantes de la molécula de ácido nucleico de la invención pueden tener 1, 2, 3, 4, 5, 10, 15, 25, 50, 100, 200, 500 y más cambios (alteraciones, sustituciones, eliminaciones o inserciones) de nucleótidos. También son posibles variantes degeneradas que codifican la misma
15 proteína, así como variantes no degeneradas que codifican una proteína distinta. La secuencia de nucleótidos de dichos mutantes y variantes codifica una proteína, o un fragmento biológicamente activo de la misma, que conserva, al menos, una de las actividades o funciones biológicas de la proteína correspondiente codificada por cualquier
20 marco de lectura abierto (ORF) del cluster de genes responsables de la biosíntesis de tiocoralina. Las formas alélicas de los genes de dicho cluster así como los polimorfismos también se hallan comprendidos dentro del ámbito de la presente invención.

25 En una realización particular, la molécula de ácido nucleico de la invención es una molécula de ácido nucleico aislada, opcionalmente purificada, que comprende una secuencia de nucleótidos que codifica la totalidad de las proteínas de la ruta biosintética de producción de tiocoralina, o fragmentos biológicamente activos de las mismas. En este
30 caso, la molécula de ácido nucleico de la invención comprende la secuencia de nucleótidos que contiene el cluster completo de genes responsables de la biosíntesis de tiocoralina.

La secuencia de nucleótidos del cluster completo de genes responsables de la biosíntesis de tiocoralina se recoge en la SEQ ID NO: 1, una secuencia de ADN genómico de 64.650 pares de bases (pb) de *Micromonospora* sp. ML1. El ámbito de la invención también incluye la
5 cadena complementaria a la secuencia de nucleótidos mostrada en la SEQ ID NO: 1, es decir, la formada por nucleótidos que son complementarios a los indicados en la SEQ ID NO: 1 (e.g., A sustituido por T, C sustituido por G y viceversa) y/o secuencias de nucleótidos inversas [es decir, las secuencias generadas cambiando la dirección del
10 sentido de la lectura e.g., de (5' → 3') a (3' → 5')].

La presente invención incluye, además, una molécula de ácido nucleico que hibrida con la molécula de ácido nucleico de la invención que tiene la secuencia de nucleótidos mostrada en la SEQ ID NO: 1 o su
15 cadena complementaria; dicha molécula puede ser aislada de un organismo productor de tiocoralina y codifica, al menos, una proteína de la ruta biosintética de producción de tiocoralina. Las técnicas y condiciones típicas de hibridación, conocidas por los expertos en la materia, se mencionan, por ejemplo, en Sambrook *et al.*, Molecular
20 Cloning: A Laboratory Manual, Second Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (1989). En general, se utilizan condiciones convencionales o severas de hibridación para sondas homólogas, mientras que se usan condiciones de hibridación menos severas para sondas parcialmente homólogas que tienen menos del
25 100% de homología con la secuencia de la molécula del ácido nucleico diana. En este último caso (sondas parcialmente homólogas), se pueden realizar una serie de hibridaciones Southern o Northern con diferentes condiciones. A modo ilustrativo, cuando la hibridación se lleva a cabo en un disolvente que contiene formamida, las condiciones preferidas
30 incluyen el empleo de una temperatura y una fuerza iónica constante de, aproximadamente, 42°C con una solución que contiene 6X SSC, 50% de formamida. Condiciones de hibridación menos severas pueden utilizar la misma temperatura y fuerza iónica aunque, en este caso, la

cantidad de formamida en el tampón de anillamiento sería menor (de aproximadamente 45% a 0%). Alternativamente, la hibridación puede llevarse a cabo en soluciones acuosas que no contienen formamida. En general, para la hibridación en medio acuoso, la fuerza iónica de la solución acuosa se mantiene igual, típicamente de aproximadamente Na^+ 1 M, mientras que la temperatura de anillamiento puede reducirse de 68°C a 42°C.

La secuenciación del cluster completo de genes responsables de la biosíntesis de tiocoralina (SEQ ID NO: 1) reveló la presencia de 36 marcos de lectura abiertos (ORFs) completos y de otros 2 ORFs incompletos (ORF1 y ORF38, véase más adelante). En la Tabla 1 (Ejemplo 12) se recoge la posición de los distintos ORFs implicados en la ruta biosintética de producción de tiocoralina, así como las secuencias de aminoácidos codificadas por dichos ORFs.

La molécula completa de ADN cromosómico (genómico) que contiene el cluster de genes responsables de la biosíntesis de tiocoralina, que codifica todas las proteínas biosintéticas esenciales para la producción de tiocoralina, ha sido empaquetada eficientemente en dos plásmidos, concretamente en los cósmidos SuperCos1 y pKC505 (Ejemplos 1 y 2). Esos dos cósmidos, que contienen el cluster de genes responsables de la biosíntesis de tiocoralina, son suficientes para regenerar la ruta biosintética completa para la producción de tiocoralina. Por tanto, en una realización particular, la invención proporciona el cluster completo de genes biosintéticos de tiocoralina en dos cósmidos lo que permite disponer de unos medios sustancialmente más eficientes para producir tiocoralina.

En una realización particular, la molécula de ácido nucleico de la invención es una molécula de ácido nucleico aislada, opcionalmente purificada, que comprende una secuencia de nucleótidos que codifica una proteína de la ruta biosintética de producción de tiocoralina, o un

fragmento biológicamente activo de la misma. En una realización concreta, la molécula de ácido nucleico de la invención se selecciona del grupo de genes formado por:

- la molécula de ácido nucleico que comprende los nucleótidos 2-535 de la SEQ ID NO: 1 (*orf1*);
- la molécula de ácido nucleico que comprende los nucleótidos 993-1130c de la SEQ ID NO: 1 (*orf2*);
- la molécula de ácido nucleico que comprende los nucleótidos 1517-2131 de la SEQ ID NO: 1 (*tio3*);
- la molécula de ácido nucleico que comprende los nucleótidos 2154-2822c de la SEQ ID NO: 1 (*tio4*);
- la molécula de ácido nucleico que comprende los nucleótidos 2970-3791c de la SEQ ID NO: 1 (*tio5*);
- la molécula de ácido nucleico que comprende los nucleótidos 3794-4777c de la SEQ ID NO: 1 (*tio6*);
- la molécula de ácido nucleico que comprende los nucleótidos 4904-5611 de la SEQ ID NO: 1 (*tio7*);
- la molécula de ácido nucleico que comprende los nucleótidos 5701-6426c de la SEQ ID NO: 1 (*tio8*);
- la molécula de ácido nucleico que comprende los nucleótidos 6426-7688c de la SEQ ID NO: 1 (*tio9*);
- la molécula de ácido nucleico que comprende los nucleótidos 7733-8524c de la SEQ ID NO: 1 (*tio10*);
- la molécula de ácido nucleico que comprende los nucleótidos 8791-10002 de la SEQ ID NO: 1 (*tio11*);
- la molécula de ácido nucleico que comprende los nucleótidos 10002-11590c de la SEQ ID NO: 1 (*tio12*);
- la molécula de ácido nucleico que comprende los nucleótidos 11847-13634 de la SEQ ID NO: 1 (*tio13*);
- la molécula de ácido nucleico que comprende los nucleótidos 13734-15005c de la SEQ ID NO: 1 (*tio14*);
- la molécula de ácido nucleico que comprende los nucleótidos 15005-16354c de la SEQ ID NO: 1 (*tio15*);

- la molécula de ácido nucleico que comprende los nucleótidos 16441-18744c de la SEQ ID NO: 1 (*tio16*);
- la molécula de ácido nucleico que comprende los nucleótidos 18774-19055c de la SEQ ID NO: 1 (*tio17*);
- 5 - la molécula de ácido nucleico que comprende los nucleótidos 19260-20036 de la SEQ ID NO: 1 (*tio18*);
- la molécula de ácido nucleico que comprende los nucleótidos 20146-20880c de la SEQ ID NO: 1 (*tio19*);
- la molécula de ácido nucleico que comprende los nucleótidos 21188-28969 de la SEQ ID NO: 1 (*tio20*);
- 10 - la molécula de ácido nucleico que comprende los nucleótidos 28979-38398 de la SEQ ID NO: 1 (*tio21*);
- la molécula de ácido nucleico que comprende los nucleótidos 38449-38661 de la SEQ ID NO: 1 (*tio22*);
- 15 - la molécula de ácido nucleico que comprende los nucleótidos 38642-41263 de la SEQ ID NO: 1 (*tio23*);
- la molécula de ácido nucleico que comprende los nucleótidos 41835-42368 de la SEQ ID NO: 1 (*tio24*);
- la molécula de ácido nucleico que comprende los nucleótidos 42395-43255c de la SEQ ID NO: 1 (*tio25*);
- 20 - la molécula de ácido nucleico que comprende los nucleótidos 43340-43741c de la SEQ ID NO: 1 (*tio26*);
- la molécula de ácido nucleico que comprende los nucleótidos 44152-49563 de la SEQ ID NO: 1 (*tio27*);
- 25 - la molécula de ácido nucleico que comprende los nucleótidos 49635-53669 de la SEQ ID NO: 1 (*tio28*);
- la molécula de ácido nucleico que comprende los nucleótidos 53749-55305c de la SEQ ID NO: 1 (*orf29*);
- la molécula de ácido nucleico que comprende los nucleótidos 55384-57222c de la SEQ ID NO: 1 (*orf30*);
- 30 - la molécula de ácido nucleico que comprende los nucleótidos 57895-58467c de la SEQ ID NO: 1 (*orf31*);

- la molécula de ácido nucleico que comprende los nucleótidos 58535-59206c de la SEQ ID NO: 1 (*orf32*);
- la molécula de ácido nucleico que comprende los nucleótidos 59298-59564c de la SEQ ID NO: 1 (*orf33*);
- 5 - la molécula de ácido nucleico que comprende los nucleótidos 59611-60114c de la SEQ ID NO: 1 (*orf34*);
- la molécula de ácido nucleico que comprende los nucleótidos 60202-60888 de la SEQ ID NO: 1 (*orf35*);
- la molécula de ácido nucleico que comprende los nucleótidos 60960-62240 de la SEQ ID NO: 1 (*orf36*);
- 10 - la molécula de ácido nucleico que comprende los nucleótidos 62300-62833 de la SEQ ID NO: 1 (*orf37*);
- la molécula de ácido nucleico que comprende los nucleótidos 62925-64650 de la SEQ ID NO: 1 (*orf38*); o
- 15 fragmentos de las mismas que codifican fragmentos biológicamente activos de proteínas de la ruta biosintética de producción de tiocoralina.

En otra realización particular, la molécula de ácido nucleico de la invención es una molécula de ácido nucleico aislada, opcionalmente purificada, que comprende una secuencia de nucleótidos que codifica dos o más proteínas de la ruta biosintética de producción de tiocoralina, o fragmentos biológicamente activos de las mismas. En una realización concreta, la molécula de ácido nucleico de la invención comprende una secuencia de nucleótidos que comprende dos o más genes seleccionados entre los genes identificados como *orf1*, *orf2*, *tio3*, *tio4*, *tio5*, *tio6*, *tio7*, *tio8*, *tio9*, *tio10*, *tio11*, *tio12*, *tio13*, *tio14*, *tio15*, *tio16*, *tio17*, *tio18*, *tio19*, *tio20*, *tio21*, *tio22*, *tio23*, *tio24*, *tio25*, *tio26*, *tio27*, *tio28*, *orf29*, *orf30*, *orf31*, *orf32*, *orf33*, *orf34*, *orf35*, *orf36*, *orf37*, *orf38* y fragmentos de los mismos que codifican fragmentos biológicamente activos de proteínas de la ruta biosintética de producción de tiocoralina.

30

En otra realización particular, la molécula de ácido nucleico de la invención es una molécula de ácido nucleico aislada, opcionalmente

purificada, que comprende una secuencia de nucleótidos que codifica al menos una proteína de la ruta biosintética de producción de tiocoralina, o un fragmento biológicamente activo de la misma, o un mutante o variante de la misma, en donde dicha proteína se selecciona del grupo
5 formado por las proteínas identificadas como ORF1 (SEQ ID NO: 2), ORF2 (SEQ ID NO: 3), Tio3 (SEQ ID NO: 4), Tio4 (SEQ ID NO: 5), Tio5 (SEQ ID NO: 6), Tio6 (SEQ ID NO: 7), Tio7 (SEQ ID NO: 8), Tio8 (SEQ ID NO: 9), Tio9 (SEQ ID NO: 10), Tio10 (SEQ ID NO: 11), Tio11 (SEQ ID NO: 12), Tio12 (SEQ ID NO: 13), Tio13 (SEQ ID NO: 14), Tio14 (SEQ ID
10 NO: 15), Tio15 (SEQ ID NO: 16), Tio16 (SEQ ID NO: 17), Tio17 (SEQ ID NO: 18), Tio18 (SEQ ID NO: 19), Tio19 (SEQ ID NO: 20), Tio20 (SEQ ID NO: 21), Tio21 (SEQ ID NO: 22), Tio22 (SEQ ID NO: 23), Tio23 (SEQ ID NO: 24), Tio24 (SEQ ID NO: 25), Tio25 (SEQ ID NO: 26), Tio26 (SEQ ID NO: 27), Tio27 (SEQ ID NO: 28), Tio28 (SEQ ID NO: 29), ORF29 (SEQ ID
15 NO: 30), ORF30 (SEQ ID NO: 31), ORF31 (SEQ ID NO: 32), ORF32 (SEQ ID NO: 33), ORF33 (SEQ ID NO: 34), ORF34 (SEQ ID NO: 35), ORF35 (SEQ ID NO: 36), ORF36 (SEQ ID NO: 37), ORF37 (SEQ ID NO: 38), ORF38 (SEQ ID NO: 39). Dichas proteínas pueden obtenerse a partir de los correspondientes *orfs* previamente mencionados (*orf1*, *orf2*, *tio3*,
20 *tio4*, *tio5*, *tio6*, *tio7*, *tio8*, *tio9*, *tio10*, *tio11*, *tio12*, *tio13*, *tio14*, *tio15*, *tio16*, *tio17*, *tio18*, *tio19*, *tio20*, *tio21*, *tio22*, *tio23*, *tio24*, *tio25*, *tio26*, *tio27*, *tio28*, *orf29*, *orf30*, *orf31*, *orf32*, *orf33*, *orf34*, *orf35*, *orf36*, *orf37*, *orf38*) del cluster de genes responsables de la biosíntesis de tiocoralina (SEQ ID NO: 1), o de las correspondientes regiones, mutantes o variantes de
25 los mismos.

En otra realización particular, la molécula de ácido nucleico de la invención es una molécula de ácido nucleico aislada, opcionalmente purificada, que comprende una secuencia de nucleótidos que codifica al
30 menos una variante de una proteína de la ruta biosintética de producción de tiocoralina, o un fragmento biológicamente activo de la misma, en donde dicha variante es, al menos, un 30%, ventajosamente un 50%, preferentemente un 60%, más preferentemente un 70%, aún

más preferentemente un 80%, particularmente un 90%, más particularmente un 95% o más, idéntica en su secuencia de aminoácidos a la de una proteína seleccionada entre las proteínas cuyas secuencias de aminoácidos se muestran en SEQ ID NO: 2-39, o a
5 fragmentos biológicamente activos de las mismas. Dicha variante conserva, al menos, una de las actividades o funciones biológicas de la correspondiente proteína codificada por cualquiera de los *orfs* del cluster de genes responsables de la biosíntesis de tiocoralina.

10 En otro aspecto, la presente invención se relaciona con una composición que comprende, al menos, una molécula de ácido nucleico de la invención, preferentemente una molécula aislada de ácido nucleico. En una realización particular, dicha composición comprende una molécula de ácido nucleico de la invención. En otra realización
15 particular, dicha composición comprende dos o más moléculas de ácido nucleico de la invención. Dichas moléculas de ácido nucleico pueden ser tanto de ADN como de ARN.

La molécula de ácido nucleico de la invención puede ser aislada a
20 partir de cualquier organismo productor de tiocoralina, bien de forma natural o bien de forma recombinante porque se haya insertado el cluster de genes responsables de la biosíntesis de tiocoralina en una célula hospedadora (host) apropiada; no obstante, en una realización particular, dicha molécula de ácido nucleico de la invención ha sido
25 aislada del actinomiceto marino *Micromonospora* sp. ML1 (véase la parte experimental, Etapa 1, Ejemplos 1-4).

El aislamiento y caracterización del ADN genómico (cromosómico) y del ADN recombinante clonado, a partir de células hospedadoras
30 apropiadas puede realizarse mediante técnicas de hibridación convencionales o severas, utilizando la totalidad o parte de una secuencia de nucleótidos como sonda para rastrear una genoteca apropiada.

Por tanto, en otro aspecto, la invención se relaciona con una sonda que comprende una molécula de ácido nucleico de la invención o un fragmento de la misma. En general, las sondas comprenden
5 convenientemente una secuencia de al menos 5, 10, 15, 20, 25, 30, 40, 50, 60 o más nucleótidos. Las secuencias con una longitud de 20 a 60 nucleótidos son las preferidas. En una realización particular, dicha sonda puede utilizarse para la detección de genes implicados en la biosíntesis de tiocoralina en *Micromonospora sp.* El empleo de dicha
10 sonda para la detección de un ácido nucleico, e.g., ADNg, ADNc o ARNm, relacionado con la biosíntesis de tiocoralina constituye un aspecto adicional de esta invención.

Alternativamente, el aislamiento y caracterización del ADN
15 genómico (cromosómico) y del ADN recombinante clonado, a partir de células hospedadoras apropiadas puede realizarse mediante técnicas basadas en la amplificación enzimática de ácidos nucleicos. A modo ilustrativo, se pueden diseñar oligonucleótidos iniciadores (en base a las secuencias conocidas de ADN y de proteínas implicadas en la
20 biosíntesis de tiocoralina) que pueden ser utilizados en reacciones de amplificación enzimática, por ejemplo, PCR, para amplificar e identificar otras secuencias idénticas o relacionadas.

Las moléculas de ácido nucleico de la invención pueden ser
25 aisladas y, si se desea, purificadas, por métodos convencionales. Aunque, en general, las moléculas de ácido nucleico de la invención se obtendrán por métodos recombinantes o de aislamiento, la invención también contempla la posibilidad de que las moléculas de ácido nucleico de la invención sean obtenidas por síntesis química, las cuales
30 tendrán la misma, o sustancialmente la misma, estructura que las derivadas directamente de los organismos productores de tiocoralina, tanto tipo salvaje (wt) como mutantes.

En otro aspecto, la invención se relaciona con un vector, en adelante vector de la invención, que comprende una molécula de ácido nucleico de la invención que codifica al menos una proteína de la ruta biosintética de producción de tiocoralina, o un fragmento biológicamente activo de la misma. En una realización particular, el vector de la invención es un vector o un plásmido biológicamente funcional, tal como un vector de clonación o un vector de expresión.

En una realización concreta, el vector de la invención es un vector de clonación, preferentemente, un cósmido. Los vectores de clonación preferidos se seleccionan por su capacidad para incorporar grandes secuencias de ADN (e.g., clusters completos de genes implicados en la biosíntesis de productos de interés). En general, dichos vectores son vectores convencionales y están disponibles comúnmente. En la presente invención se contempla, además, que el material genético pueda reducirse para finalmente contenerse en un sólo vector o plásmido de clonación (e.g., cósmido) mediante manipulación genética por técnicas conocidas por el experto en la materia. El ensamblaje puede llevarse a cabo mediante clonaje, PCR o genes sintéticos o combinación de cualquiera de estas técnicas conocidas en el estado de la técnica.

En otra realización particular, el vector de la invención es un vector de expresión apropiado para su inserción en una célula hospedadora adecuada. La inserción de dicho vector en dicha célula hospedadora adecuada puede realizarse por cualquier método convencional de transferencia de material genético (e.g., transformación, transfección, etc.).

Por tanto, en otro aspecto, la invención se relaciona con una célula hospedadora, en adelante célula hospedadora de la invención, trasformada o transfectada con un vector de la invención. Dicha célula hospedadora de la invención contiene una o más moléculas de ácido

nucleico de la invención. En una realización particular, la célula hospedadora de la invención contiene una molécula de ácido nucleico de la invención. En otra realización particular, la célula hospedadora de la invención contiene dos o más moléculas de ácido nucleico de la invención; en este caso, dichas moléculas de ácido nucleico de la invención pueden ser iguales o diferentes entre sí.

Una célula hospedadora de la invención preferida es una célula hospedadora trasformada o transfectada establemente con un vector de la invención que comprende una molécula de ácido nucleico (exógeno) de la invención que comprende una secuencia de nucleótidos que codifica al menos una proteína de la ruta biosintética de producción de tiocoralina, o un fragmento biológicamente activo de la misma, de forma suficiente como para dirigir la biosíntesis y/o ensamblaje de la tiocoralina. Preferentemente, la célula hospedadora es un microorganismo, más preferentemente una bacteria. En una realización particular, dicha célula hospedadora es una bacteria Gram-positiva, tal como un actinomiceto, por ejemplo, un estreptomiceto.

Aunque en los ejemplos de la presente invención se han usado diversas especies de estreptomicetos, tales como *Streptomyces coelicolor*, *Streptomyces lividans*, *Streptomyces albus* y *Streptomyces avermitilis* como hospedadores heterólogos, la expresión heteróloga de los genes implicados en la biosíntesis de la tiocoralina se puede realizar en otros estreptomicetos, actinomicetos, etc., siempre y cuando sean capaces de ser transformados, preferentemente de forma estable, con los vectores de la invención. La expresión *in vitro* de las proteínas puede ser realizada, si se desea, usando métodos convencionales.

En una realización particular, la invención proporciona una célula hospedadora de la invención, tal como, por ejemplo, una bacteria recombinante, en la que, al menos, una región de la molécula de ácido nucleico de la invención ha sido alterada para dar lugar a una célula

hospedadora recombinante, tal como una bacteria recombinante, que produce niveles alterados de tiocoralina con respecto a la correspondiente célula (bacteria) productora de tiocoralina no recombinante, es decir, wt. Para ello, se pueden utilizar técnicas
5 convencionales conocidas por los expertos en la materia que incluyen por ejemplo, aumentar el número de copias de los genes responsables de los dominios más importantes de las NRPSs implicadas en la producción de tiocoralina o aumentar las secuencias reguladoras de la expresión génica de esos genes por técnicas de ingeniería genética
10 conocidas en el estado de la técnica y así aumentar el rendimiento en la producción de tiocoralina.

En otro aspecto, la invención se relaciona con una proteína, en adelante proteína de la invención, codificada por la molécula de ácido
15 nucleico de la invención.

Tal como aquí se utiliza, el término “proteína” significa polipéptidos, enzimas y similares, codificados por la molécula de ácido nucleico de la invención que comprende la ruta biosintética para la
20 producción de tiocoralina. Las proteínas de la invención incluyen cadenas de aminoácidos de longitudes variables, incluyendo cadenas de aminoácidos de longitud completa, en donde los restos de los aminoácidos están unidos por enlaces peptídicos covalentes, así como fragmentos biológicamente activos de dichas proteínas implicadas en la
25 biosíntesis de tiocoralina, así como las variantes biológicamente activas de las mismas. Las proteínas de la invención pueden ser naturales, recombinantes o sintéticas. A modo ilustrativo, dichas proteínas implicadas en la biosíntesis de tiocoralina pueden ser producidas a través de la tecnología del ADN recombinante convencional insertando
30 una secuencia de nucleótidos que codifica la proteína en un vector de expresión apropiado y expresar la proteína en una célula hospedadora apropiada o bien a través de síntesis química convencional de péptidos, por ejemplo, mediante el método de síntesis en fase sólida de Merrifield

(Merrifield, J. Am. Chem. Soc. 85:2149-2154 (1963)) en el que los aminoácidos se unen de forma individual y secuencial a la cadena aminoacídica. Alternativamente, la síntesis de proteínas de la invención puede realizarse utilizando equipos para la síntesis automatizada de proteínas comercializados por distintos fabricantes (e.g., Perkin-Elmer, Inc.).

Las variantes biológicamente activas incluidas dentro del ámbito de la presente invención comprenden, como mínimo, un fragmento biológicamente activo de la secuencia de aminoácidos codificada por la molécula de ácido nucleico de la invención, es decir, una parte de la estructura de la proteína que retiene la función activa de la proteína, por ejemplo, la parte de la tioesterasa codificada por el gen *tio18* que tiene la misma, o sustancialmente la misma, actividad que la proteína Tio18 codificada por dicho gen *tio18*, es decir, tiene, al menos, una similaridad o potencia de, al menos, aproximadamente un 70%, ventajosamente de, al menos, un 80%, preferentemente de, al menos, un 90%, más preferentemente, de alrededor del 95% aproximadamente.

Las variantes biológicamente activas de las proteínas de la invención incluyen estructuras de aminoácidos activas en las que se han eliminado, sustituido o añadido aminoácidos, alelos que ocurren de forma natural, etc. El fragmento biológicamente activo puede ser identificado fácilmente sometiendo la proteína, longitud completa, a digestion química o enzimática para preparar fragmentos y, a continuación, ensayar dichos fragmentos mediante ensayos convencionales para analizar los fragmentos de la estructura aminoacídica que conservan la misma, o sustancialmente la misma, actividad biológica que la proteína longitud completa.

30

En una realización particular, la proteína de la invención es una proteína aislada, opcionalmente purificada, implicada en la biosíntesis de tiocoralina, codificada por un gen seleccionado del grupo formado

por los genes identificados como *orf1*, *orf2*, *tio3*, *tio4*, *tio5*, *tio6*, *tio7*, *tio8*, *tio9*, *tio10*, *tio11*, *tio12*, *tio13*, *tio14*, *tio15*, *tio16*, *tio17*, *tio18*, *tio19*, *tio20*, *tio21*, *tio22*, *tio23*, *tio24*, *tio25*, *tio26*, *tio27*, *tio28*, *orf29*, *orf30*, *orf31*, *orf32*, *orf33*, *orf34*, *orf35*, *orf36*, *orf37* y *orf38*.

5

En otra realización particular, la proteína de la invención es una proteína aislada, opcionalmente purificada, implicada en la biosíntesis de tiocoralina, seleccionada del grupo formado por las proteínas identificadas como ORF1 (SEQ ID NO: 2), ORF2 (SEQ ID NO: 3), Tio3 (SEQ ID NO: 4), Tio4 (SEQ ID NO: 5), Tio5 (SEQ ID NO: 6), Tio6 (SEQ ID NO: 7), Tio7 (SEQ ID NO: 8), Tio8 (SEQ ID NO: 9), Tio9 (SEQ ID NO: 10), Tio10 (SEQ ID NO: 11), Tio11 (SEQ ID NO: 12), Tio12 (SEQ ID NO: 13), Tio13 (SEQ ID NO: 14), Tio14 (SEQ ID NO: 15), Tio15 (SEQ ID NO: 16), Tio16 (SEQ ID NO: 17), Tio17 (SEQ ID NO: 18), Tio18 (SEQ ID NO: 19), Tio19 (SEQ ID NO: 20), Tio20 (SEQ ID NO: 21), Tio21 (SEQ ID NO: 22), Tio22 (SEQ ID NO: 23), Tio23 (SEQ ID NO: 24), Tio24 (SEQ ID NO: 25), Tio25 (SEQ ID NO: 26), Tio26 (SEQ ID NO: 27), Tio27 (SEQ ID NO: 28), Tio28 (SEQ ID NO: 29), ORF29 (SEQ ID NO: 30), ORF30 (SEQ ID NO: 31), ORF31 (SEQ ID NO: 32), ORF32 (SEQ ID NO: 33), ORF33 (SEQ ID NO: 34), ORF34 (SEQ ID NO: 35), ORF35 (SEQ ID NO: 36), ORF36 (SEQ ID NO: 37), ORF37 (SEQ ID NO: 38), ORF38 (SEQ ID NO: 39), y combinaciones de las mismas, o fragmentos biológicamente activos de las mismas. Las funciones hipotéticas de dichas proteínas se recogen en la Tabla 1.

25

Los *orfs* del cluster de genes responsable de la biosíntesis de tiocoralina, que codifican las proteínas implicadas en la biosíntesis de dicho compuesto, pueden ser identificados utilizando técnicas convencionales. Ejemplos ilustrativos, no limitativos, de dichas técnicas incluyen el análisis computacional para localizar los codones de inicio y parada, las localizaciones putativas de los marcos de lectura basadas en las frecuencias de los codones, alineamientos por similitud a genes expresados en otros actinomicetos y similares. De esta manera, las

30

proteínas de la invención pueden ser identificadas utilizando la secuencia de nucleótidos de la presente invención y los *orfs* o las proteínas codificadas por los mismos pueden ser aisladas y, si se desea, purificadas, o, alternativamente, sintetizadas por métodos químicos. Se
5 pueden diseñar construcciones génicas para la expresión de dichos productos en base a los *orfs* e incluir los elementos de regulación de expresión (promotores, terminadores, etc.) apropiados y dichas construcciones génicas pueden ser introducidas en células hospedadoras apropiadas para expresar la proteína o proteínas
10 codificadas por una o más *orfs*.

Las proteínas de la invención pueden ser aisladas y, si se desea, purificadas, por métodos convencionales. Preferentemente, las proteínas se obtienen en una forma sustancialmente pura, aunque también
15 puede ser aceptable un grado inferior de pureza, típicamente del 80% al 90% aproximadamente. La invención también contempla la posibilidad de que las proteínas de la invención sean obtenidas por síntesis química, las cuales tendrán la misma, o sustancialmente la misma, estructura que las derivadas directamente de los organismos
20 productores de tiocoralina, tanto tipo salvaje (wt) como mutantes.

En otro aspecto, la invención se relaciona con un procedimiento para producir una proteína de la invención implicada en la biosíntesis de tiocoralina que comprende crecer, bajo condiciones (nutrientes y
25 ambientales) apropiadas, un organismo productor de tiocoralina, y, si se desea, aislar una o más de dichas proteínas implicadas en la biosíntesis de tiocoralina. Si se desea, dicha proteína de la invención puede ser aislada y purificada por métodos convencionales, tales como los descritos previamente.

30

En otro aspecto, la invención se relaciona con un método para producir tiocoralina que comprende crecer, bajo condiciones apropiadas para producir dicho compuesto, un organismo productor de tiocoralina

en el que se ha aumentado el número de copias de genes que codifican proteínas implicadas en la biosíntesis de tiocoralina, y, si se desea, aislar la tiocoralina.

5 En una realización particular, el organismo productor de tiocoralina es un actinomiceto, tal como, por ejemplo, *Micromonospora* *sp.*, en el que se ha aumentado el número de copias de genes que codifican proteínas implicadas en la biosíntesis de tiocoralina. El aumento en el número de copias de genes que codifican proteínas
10 implicadas en la biosíntesis de tiocoralina puede realizarse por métodos convencionales conocidos por los expertos en la materia. En este caso, el método descrito previamente comprende fermentar dicho organismo bajo condiciones nutrientes y ambientales adecuadas para la expresión de los genes implicados en la producción de tiocoralina. Si se desea, la
15 tiocoralina producida puede ser aislada y purificada del medio de cultivo por métodos convencionales.

 En otro aspecto, la invención se relaciona con un método para producir tiocoralina que comprende crecer, bajo condiciones apropiadas
20 para producir dicho compuesto, un organismo productor de tiocoralina en el que la expresión de los genes que codifican las proteínas responsables de la biosíntesis de tiocoralina ha sido modulada mediante manipulación o reemplazo de uno o más genes que codifican proteínas implicadas en la biosíntesis de tiocoralina o bien mediante la
25 manipulación de las secuencias responsables de regular la expresión de dichos genes, y, si se desea, aislar la tiocoralina. Preferentemente, la expresión de los genes que codifican para dichas proteínas responsables de la biosíntesis de tiocoralina ha sido mejorada. Para ello, se pueden eliminar las secuencias génicas no esenciales en el proceso de
30 biosíntesis de tiocoralina, o bien se puede aumentar la eficiencia de las secuencias reguladoras de la expresión génica de dichos genes por técnicas de ingeniería genética conocidas por los expertos en la materia. De este modo se puede aumentar el rendimiento en la producción de

tiocoralina. La manipulación genética para eliminar las secuencias
génicas no esenciales en el proceso de biosíntesis de tiocoralina, o bien
para aumentar la eficiencia de las secuencias reguladoras de la
expresión génica de dichos genes puede llevarse a cabo por técnicas de
5 ingeniería genética conocidas por los expertos en la materia.

En una realización particular, el organismo productor de
tiocoralina es un actinomiceto, tal como, por ejemplo, *Micromonospora*
sp, en el que la expresión de los genes que codifican las proteínas
10 responsables de la biosíntesis de tiocoralina ha sido modulada
mediante manipulación o reemplazo de uno o más genes que codifican
proteínas implicadas en la biosíntesis de tiocoralina o bien mediante la
manipulación de las secuencias responsables de regular la expresión de
dichos genes, lo que puede llevarse a cabo por métodos convencionales
15 conocidos por los expertos en la materia. En este caso, el método
descrito previamente comprende fermentar dicho organismo bajo
condiciones nutrientes y ambientales adecuadas para la expresión de
los genes implicados en la producción de tiocoralina. Si se desea, la
tiocoralina producida puede ser aislada y purificada del medio de
20 cultivo por métodos convencionales.

En otro aspecto, la invención se relaciona con un método para
producir tiocoralina que comprende crecer, bajo condiciones apropiadas
para producir dicho compuesto, una célula hospedadora de la invención
25 transformada o transfectada con un vector de la invención que
comprende el cluster de genes responsable de la biosíntesis de
tiocoralina, y, si se desea, aislar la tiocoralina. Las condiciones
(nutrientes, ambientales, etc.) se seleccionarán en función de la
naturaleza de la células hospedadora.

30

En una realización particular, la célula hospedadora de la
invención se selecciona entre un organismo productor de tiocoralina de
forma nativa, un organismo no-productor de tiocoralina de forma nativa

y un organismo manipulado genéticamente para producir tiocoralina. En una realización particular, dicha célula hospedadora de la invención es un actinomiceto o un estreptomiceto.

5 En otro aspecto, la invención se relaciona con un procedimiento, basado en la utilización de genes responsables de la biosíntesis de tiocoralina de *Micromonospora sp.* ML1, para la producción de dicho compuesto en otro actinomiceto, que comprende:

- 10 (1) la obtención de mutantes afectados en genes específicos de la ruta de biosíntesis de tiocoralina;
- (2) el aislamiento de la región del cromosoma de *Micromonospora sp.* ML1 que contiene el cluster de genes
15 responsables de la biosíntesis de tiocoralina;
- (3) la obtención y análisis de la secuencia de nucleótidos del cluster de genes responsables de la biosíntesis de tiocoralina; y
- 20 (4) la producción heteróloga de tiocoralina en otros actinomicetos.

 La identificación y el aislamiento de la región del cromosoma de
25 *Micromonospora sp.* ML1 que contiene el cluster de genes responsables de la biosíntesis de tiocoralina, así como el análisis de la secuencia de nucleótidos de dicho cluster puede llevarse a cabo en base a las enseñanzas proporcionadas por esta invención, ilustrada, de forma no limitativa, en los Ejemplos que acompañan a esta descripción.

30

 Los mutantes afectados en genes específicos de la ruta de biosíntesis de tiocoralina pueden ser identificados por métodos convencionales. En una realización particular, dichos mutantes pueden

ser identificados mediante cultivo y medida de la producción de tiocoralina por métodos convencionales, por ejemplo, por HPLC-MS, tal como se menciona en el Ejemplo 5.

- 5 La totalidad o una parte del cluster de genes responsables de la biosíntesis de tiocoralina puede ser introducida en un actinomiceto por métodos convencionales, e.g., por transformación o transfección, para la producción heteróloga de tiocoralina por fermentación de un medio nutriente apropiado, bajo las condiciones apropiadas para la
- 10 producción de tiocoralina y, si se desea, la tiocoralina así obtenida puede ser aislada y/o purificada por métodos convencionales.

 La determinación del cluster de genes responsable de la biosíntesis de la tiocoralina es de gran importancia comercial. El

15 aislamiento y descripción completa del cluster de genes responsables de la biosíntesis de tiocoralina proporcionado por esta invención permite aumentar la producción de la tiocoralina y manipular organismos para que produzcan tiocoralina. En este sentido, se puede aumentar el

20 número de copias de los genes responsables de los dominios más importantes de las NRPSs implicadas en la producción de tiocoralina o aumentar la eficiencia de las secuencias reguladoras de la expresión génica de esos genes por técnicas de ingeniería genética conocidas en el

 estado de la técnica y así aumentar el rendimiento en su producción.

25 Otra ventaja asociada con la identificación y clonaje del cluster completo de genes de la tiocoralina proporcionado por la presente invención se relaciona con la producción eficiente de tiocoralina. De hecho, permite obtener un compuesto de gran interés en un menor número de pasos. La eliminación de secuencias no esenciales en el

30 proceso de biosíntesis en mutantes del cluster reduce considerablemente el tiempo necesario para producir el compuesto de interés. Las secuencias restantes son suficientes y mantienen su funcionalidad para producir tiocoralina.

PARTE EXPERIMENTAL

Los procedimientos experimentales de la presente invención incluyen métodos de Biología Molecular convencionales en el actual estado de la técnica. Pueden encontrarse descripciones detalladas de las técnicas no explicadas aquí en los manuales de Kieser et al. (Practical *Streptomyces* genetics. The John Innes Foundation, Norwich, Gran Bretaña, 2000) y Sambrook et al. (Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, Nueva York, EE.UU, 2001). Las siguientes etapas describen en detalle, sin limitación, la presente invención.

Etapas 1. Aislamiento de la región del cromosoma de *Micromonospora* sp. ML1 que contiene los genes de la ruta de biosíntesis de tiocoralina

Ejemplo 1. Construcción de una genoteca en SuperCos1 a partir de ADN cromosómico de *Micromonospora* sp. ML1

A partir de un cultivo de *Micromonospora* sp. ML1 (Espliego, F. Tesis Doctoral, 1996, Universidad de León; de la Calle, F. Tesis Doctoral, 1998, Universidad Autónoma de Madrid), disponible en la colección de cultivos de Pharma Mar, S.A., en medio MIAM2 (5 g/L de extracto de levadura, 3 g/L de extracto de carne, 5 g/L de triptona, 5 g/L de glucosa, 20 g/L de dextrina, 4 g/L de CaCO₃, 10 g/L de sales marinas. pH 6,8) se obtuvo ADN cromosómico utilizando el protocolo "salting out" (Kieser et al. 2000). Este ADN cromosómico fue sometido a digestión parcial con la endonucleasa *Bam*HI y los fragmentos obtenidos fueron utilizados para generar una genoteca en el cósmido SuperCos 1 (Stratagene), digerido con *Bam*HI. La generación de esta genoteca en *E. coli* XL-1 Blue MR (Stratagene) se realizó siguiendo procedimientos ya descritos (Sambrook et al. 2001) y el kit de

empaquetamiento *in vitro* Gigapack III Gold Packaging Extract Kit (Stratagene).

1.000 colonias transductantes de *E. coli* fueron depositadas sobre
5 membranas de nylon para realizar un análisis de hibridación de colonia
in situ mediante los protocolos habituales (Sambrook et al. 2001).

Ejemplo 2. Construcción de una genoteca en pKC505 a partir de ADN cromosómico de *Micromonospora sp.* ML1

10

A partir de un cultivo de *Micromonospora sp.* ML1 en medio MIAM2 se obtuvo ADN cromosómico utilizando el protocolo “salting out” (Kieser et al. 2000). Este ADN cromosómico fue sometido a digestión parcial con la endonucleasa *Sau3AI* y los fragmentos obtenidos fueron
15 utilizados para generar una genoteca en el cósmido bifuncional *Escherichia coli*/ *Streptomyces* pKC505 (Richardson et al. 1987, Gene 61, 231-241), digerido con *Bam*HI. La generación de esta genoteca en *E. coli* ED8767 se realizó siguiendo procedimientos ya descritos (Sambrook et al. 2001) y el kit de empaquetamiento *in vitro* Gigapack III Gold
20 Packaging Extract Kit (Stratagene).

3.300 colonias transductantes de *E. coli* fueron depositadas sobre placas de microtitulación de 96 pocillos conteniendo medio TSB (Merck) con 25 µg/ml de apramicina e incubadas a 30°C durante 24 horas.
25 Estos clones fueron replicados a placas de TSA (Tryptic Soy Agar) con 25 µg/ml de apramicina, y tras una noche a 30°C, las colonias fueron transferidas a membranas de nylon para realizar un análisis de hibridación de colonia *in situ* mediante los protocolos habituales (Sambrook et al. 2001).

30

Ejemplo 3. Diseño de oligonucleótidos específicos para dominios de adenilación en NRPS y su amplificación por PCR a partir de ADN cromosómico de *Micromonospora sp.* ML1

En base a la estructura de la tiocoralina se esperaba que las NRPSs encargadas de su biosíntesis poseyeran de uno a tres dominios de adenilación que activasen L-cisteína y un dominio que activase glicina. En base a esto se diseñaron oligonucleótidos degenerados, basados en regiones conservadas dentro de los dominios de adenilación de NRPSs, capaces de amplificar específicamente fragmentos de ADN codificantes de dominios de adenilación de NRPSs que se combinaron con oligonucleótidos descritos en la bibliografía para la amplificación de dominios de adenilación de NRPSs.

La amplificación por PCR con los oligonucleótidos iniciadores:

MTF2 (5'- GCNGGYGGYGCNTAYGTNCC-3'; Neilan et al. 1999. J. Bacteriol. 181(13):4089-4097) y

PSV-4 (5'-SAGSAGGSWGTGGCCGCSAGCTCGAAGAA-3')

dio como resultado una banda de 1,3 kb que fue clonada en un vector pGEM-T Easy (Promega). El programa de PCR utilizado fue de un ciclo inicial de 95°C-2 min; 60°C-15 min; 72°C-6 min seguido de 20 ciclos de 95°C-1 min; 60°C-2 min; 72°C-2 min. Como molde se utilizó ADN cromosómico de *Micromonospora sp.* ML1.

El análisis de los clones por la longitud de los fragmentos de restricción (RFLP) demostró que había tres tipos de clones diferentes correspondientes a péptido sintetasas, pGPSV1, pGPSV2 y pGPSV3, que contenían los fragmentos de los dominios de adenilación denominados PSV1, PSV2 y PSV3, respectivamente.

Posteriormente, el inserto de los clones fue liberado con una digestión EcoRI y el fragmento fue clonado en pBBR1-MCS2 (Kovach, M.E. et al. 1995. Gene. 166:175-176), para construir los plásmidos pBPSV1, pBPSV2 y pBPSV3, respectivamente, que contenían los fragmentos de los dominios de adenilación denominados PSV1, PSV2 y PSV3, respectivamente.

A partir de la banda de PCR obtenida con los oligonucleótidos iniciadores MTF2 y PS4 se realizó una PCR anidada o “nested-PCR” (30 ciclos de 95°C-1 min; 60°C-1 min; 72°C-1 min) con los oligonucleótidos iniciadores PS2-TG: 5'-ACNGGNMRNCCNAARGG-3' y MTR: 5'-CCNCGDATYTTNACY-3' (Neilan et al. 1999. J. Bacteriol. 181(13):4089-4097) para obtener una banda de 750 pb que fue clonado en un vector pGEM-T Easy (Promega). El análisis de los clones por RFLP demostró que había dos nuevos tipos de clones diferentes correspondientes a péptido sintetasas, pGPSV4 y pGPSV5 respectivamente, que contenían los fragmentos de los dominios de adenilación denominados PSV4 y PSV5, respectivamente.

La amplificación por PCR con los oligonucleótidos iniciadores PS2M: 5'-TACACSGGCWSSACSGG-3' y PSV-4 dio como resultado una banda de 1,3 kb que fue clonada en un vector pGEM-T Easy (Promega). El programa utilizado fue un Touch-down empezando con 5 ciclos a la temperatura de anillamiento de 72°C, seguido de 10 ciclos a 70°C de anillamiento para terminar con 20 ciclos a 68°C (96°C-1 min; 72°C-68°C-2 min; 72°C-3 min). El análisis de los clones por RFLP demostró que había un nuevo tipo de clon correspondiente a una péptido sintetasa, pGPSV6 que contenía el fragmento del dominio de adenilación denominado PSV6.

Ejemplo 4. Análisis de las genotecas por hibridación en colonia

Las genotecas construidas en SuperCos1 y en pKC505 (Ejemplos 1 y 2) fueron sometidas a sendos análisis de hibridación de colonias *in situ* (Sambrook et al. 2001) utilizando el sistema DIG DNA Labeling and Detection Kit (Roche). Como sondas se utilizaron los 6 fragmentos de los dominios de adenilación denominados PSV1- PSV6.

A partir de la genoteca construida en SuperCos1 se obtuvieron:

- 3 cósmidos (clones) positivos que hibridaban con el fragmento PSV1, denominados pCT1a, pCT1b y pCT1c;

- 3 cósmidos (clones) positivos que hibridaban con el fragmento PSV2, denominados pCT2a, pCT2b y pCT2c; de éstos, pCT2c hibridaba
5 también con el fragmento PSV5;

- 2 cósmidos (clones) positivos que hibridaban con el fragmento PSV3, denominados pCT3a y pCT3b; además, ambos hibridaban también con el fragmento PSV6; y

- 1 cósmido (clon) positivo que hibridaba con PSV4, denominado
10 pCT4a.

A partir de la genoteca construida en pKC505 se obtuvieron 55 cósmidos positivos:

- 10 cósmidos (clones) positivos que hibridaban con el fragmento
15 PSV2, denominados cosV1-F8, cosV7-D2, cosV7-D12, cosV14-H4, cosV19-B4, cosV29-B9, cosV31-B11, cosV31-H10, cosV33-D12, cosV33-F7;

- 7 cósmidos positivos que hibridaban con el fragmento PSV5, denominados cosV1-B6, cosV6-H8, cosV11-F10, cosV20-F8, cosV22-F7,
20 cosV25-B3, cosV32-B4; y

- 38 cósmidos positivos que hibridaban con los fragmentos PSV1, PSV3, PSV4 o PSV6, denominados cosV1-B7, cosV1-F5, cosV2-E5, cosV2-F11, cosV3-D9, cosV4-D2, cosV5-D7, cosV5-G6, cosV6-A7, cosV6-A12, cosV7-E7, cosV8-F8, cosV9-H7, cosV10-A3, cosV11-B4,
25 cosV11-G2, cosV12-B12, cosV13-B2, cosV16-H11, cosV17-A3, cosV19-F4, cosV20-B3, cosV20-H5, cosV21-H6, cosV22-B11, cosV23-F8, cosV26-H11, cosV28-G1, cosV29-E1, cosV29-G6, cosV30-G5, cosV31-A12, cosV31-E10, cosV32-A7, cosV32-D10, cosV33-A8, cosV33-D10, cosV33-F10.

30

Etapas 2. Generación de mutantes en seis regiones de adenilación aisladas

Los seis fragmentos de los dominios de adenilación amplificados previamente a partir de ADN cromosómico de *Micromonospora sp.* ML1 (PSV1, PSV2, PSV3, PSV4, PSV5 y PSV6) fueron utilizados para experimentos independientes de interrupción génica con el fin de evaluar las regiones implicadas en la biosíntesis de tiocoralina (Ejemplos 6-11).

El plásmido conjugativo *E. coli*/*Streptomyces* pOJ260 (Bierman et al. 1992, Gene 116, 43-49) fue utilizado para generar las construcciones pFL903, pFL904, pFL905, pFL906, pFL940 y pFL941 que contenían las regiones PSV1 a PSV6, respectivamente. Estas construcciones fueron introducidas en la cepa conjugativa *E. coli* ET12567 (pUB307) (Kieser et al. 2000) y desde aquí, por conjugación, en la cepa *Micromonospora sp.* ML1, utilizando procedimientos descritos (Kieser et al. 2000). Los clones transconjugantes se seleccionaron con apramicina y la integración en la región cromosómica apropiada fue comprobada mediante hibridación Southern utilizando como sondas las regiones de los fragmentos de los dominios de adenilación PSV1 a PSV6 correspondientes. Los transconjugantes seleccionados de cada región de los dominios de adenilación PSV (PSV1-PSV6) se crecieron en medio MT4 de producción de tiocoralina y posteriormente su micelio extraído con acetonitrilo y analizado por HPLC-MS (Ejemplo 5). Solo los mutantes afectados en los dominios de adenilación PSV2 y PSV5 poseían un fenotipo no-productor de tiocoralina (Ejemplos 7 y 10). La producción de tiocoralina en mutantes con deleciones en PSV1, PSV3, PSV4 y PSV6 fue similar a la de la cepa wt (Ejemplos 6, 8, 9 y 11). Estos experimentos pusieron de manifiesto que los dominios de adenilación PSV2 y PSV5 estaban implicados en la biosíntesis de tiocoralina.

Ejemplo 5. Detección por HPLC de la producción de tiocoralina

Los extractos con acetonitrilo de las diferentes cepas analizadas fueron concentrados en rotavapor y resuspendidos en DMSO antes de ser utilizados para análisis por HPLC-MS.

5 Las muestras (10 μ l) se analizaron por HPLC, utilizando una columna de fase reversa (Symmetry C₁₈, 2.1 X 150 mm, Waters), usando como disolventes acetonitrilo y una mezcla de 0,1% de ácido trifluoroacético en agua. Durante los primeros 4 minutos se mantiene un concentración de la fase móvil con 10% de acetonitrilo de forma
10 isocrática. Después, hasta los 30 minutos, se inicia un gradiente lineal desde el 10% hasta el 100% de acetonitrilo. El flujo utilizado es de 0,25 mL/min. La detección y caracterización espectral de los picos se llevaron a cabo utilizando un detector de fotodiodos y mediante el empleo del software informático Millennium (Waters). Los
15 cromatogramas se extrajeron a una absorbancia de 230 nm.

Ejemplo 6. Interrupción génica en la región PSV1

A partir del plásmido pBPSV1 se obtuvo la región PSV1 como una
20 banda *Eco*RI de 1,3 kb y se clonó en el sitio *Eco*RI del plásmido conjugativo *E. coli*/*Streptomyces* pOJ260, generándose pFL903. pOJ260 contiene un gen que confiere resistencia a apramicina en *Streptomyces* y en estas células es un plásmido suicida.

25 La construcción pFL903 fue introducida en la cepa conjugativa *E. coli* ET12567 (pUB307) y desde ahí, por conjugación, en la cepa *Micromonospora* sp. ML1, utilizando procedimientos descritos (Kieser et al. 2000). Los clones transconjugantes se seleccionaron con 25 μ g/ml de apramicina y a partir de ADN cromosómico de los mismos se
30 comprobó mediante hibridación Southern que efectivamente la región PSV1 había sido interrumpida. La sonda utilizada en este caso fue la banda PSV1.

El mutante *Micromonospora sp.* ΔPSV1 fue crecido en medio MT4 de producción de tiocoralina y posteriormente su micelio fue extraído con acetonitrilo y analizado por HPLC-MS (véase el Ejemplo 5), resultando ser productor de tiocoralina. La composición del medio de cultivo MT4 por litro es la siguiente: 6 g de harina de soja, 2,5 g de extracto de malta, 2,5 g de peptona, 5 g de dextrosa, 20 g de dextrina, 4 g de CaCO₃, 10 g de sales marinas, ajustar el pH hasta 6,8.

Ejemplo 7. Interrupción génica en la región PSV2

A partir del plásmido pBPSV2 se obtuvo la región PSV2 como una banda *EcoRI* de 1,3 kb y se clonó en el sitio *EcoRI* del plásmido pOJ260, generándose pFL904.

La construcción pFL904 fue introducida en la cepa conjugativa *E. coli* ET12567 (pUB307) y desde ahí, por conjugación, en la cepa *Micromonospora sp.* ML1. Los clones transconjugantes se seleccionaron con 25 µg/ml de apramicina y a partir de ADN cromosómico de los mismos se comprobó mediante hibridación Southern que efectivamente la región PSV2 había sido interrumpida. La sonda utilizada en este caso fue la banda PSV2.

El mutante *Micromonospora sp.* ΔPSV2 fue crecido en medio MT4 de producción de tiocoralina y posteriormente su micelio fue extraído con acetonitrilo y analizado por HPLC-MS (Ejemplo 5), dando como resultado que esta cepa no producía tiocoralina.

Ejemplo 8. Interrupción génica en la región PSV3

A partir del plásmido pBPSV3 se obtuvo la región PSV3 como una banda *EcoRI* de 1,4 kb y se clonó en el sitio *EcoRI* del plásmido pOJ260, generándose pFL905.

La construcción pFL905 fue introducida en la cepa conjugativa *E. coli* ET12567 (pUB307) y desde ahí, por conjugación, en la cepa *Micromonospora sp.* ML1. Los clones transconjugantes se seleccionaron con 25 µg/ml de apramicina y a partir de ADN cromosómico de los mismos se comprobó mediante hibridación Southern que efectivamente la región PSV3 había sido interrumpida. La sonda utilizada en este caso fue la banda PSV3.

El mutante *Micromonospora sp.* ΔPSV3 fue crecido en medio MT4 de producción de tiocoralina y posteriormente su micelio fue extraído con acetonitrilo y analizado por HPLC-MS (Ejemplo 5), resultando ser productor de tiocoralina.

Ejemplo 9. Interrupción génica en región PSV4

A partir del plásmido pGPSV4 se obtuvo la región PSV4 como una banda *EcoRI* de 1,2 kb y se clonó en el sitio *EcoRI* del plásmido pOJ260, generándose pFL906.

La construcción pFL906 fue introducida en la cepa conjugativa *E. coli* ET12567 (pUB307) y desde ahí, por conjugación, en la cepa *Micromonospora sp.* ML1. Los clones transconjugantes se seleccionaron con 25 µg/ml de apramicina y a partir de ADN cromosómico de los mismos se comprobó mediante hibridación Southern que efectivamente la región PSV4 había sido interrumpida. La sonda utilizada en este caso fue la banda PSV4.

El mutante *Micromonospora sp.* ΔPSV4 fue crecido en medio MT4 de producción de tiocoralina y posteriormente su micelio fue extraído con acetonitrilo y analizado por HPLC-MS (Ejemplo 5), resultando ser productor de tiocoralina.

Ejemplo 10. Interrupción génica en región PSV5

A partir del plásmido pGPSV5 se obtuvo la región PSV5 como una banda *EcoRI* de 1,1 kb y se clonó en el sitio *EcoRI* del plásmido pOJ260, generándose pFL940.

5

La construcción pFL940 fue introducida en la cepa conjugativa *E. coli* ET12567 (pUB307) y desde ahí, por conjugación, en la cepa *Micromonospora sp.* ML1. Los clones transconjugantes se seleccionaron con 25 µg/ml de apramicina y a partir de ADN cromosómico de los
10 mismos se comprobó mediante hibridación Southern que efectivamente la región PSV5 había sido interrumpida. La sonda utilizada en este caso fue la banda PSV5.

El mutante *Micromonospora sp.* ΔPSV5 fue crecido en medio MT4
15 de producción de tiocoralina y posteriormente su micelio fue extraído con acetonitrilo y analizado por HPLC, dando como resultado que esta cepa no producía tiocoralina.

Ejemplo 11. Interrupción génica en región PSV6

20

A partir del plásmido pGPSV6 se obtuvo la región PSV6 como una banda *EcoRI* de 1,1 kb y se clonó en el sitio *EcoRI* del plásmido pOJ260, generándose pFL941.

La construcción pFL941 fue introducida en la cepa conjugativa *E. coli* ET12567 (pUB307) y desde ahí, por conjugación, en la cepa *Micromonospora sp.* ML1. Los clones transconjugantes se seleccionaron con 25 µg/ml de apramicina y a partir de ADN cromosómico de los
25 mismos se comprobó mediante hibridación Southern que efectivamente
30 la región PSV6 había sido interrumpida. La sonda utilizada en este caso fue la banda PSV6.

El mutante *Micromonospora* sp. Δ PSV6 fue crecido en medio MT4 de producción de tiocoralina y posteriormente su micelio fue extraído con acetonitrilo y analizado por HPLC, resultando ser productor de tiocoralina.

5

Etapas 3. Obtención y análisis de la secuencia nucleotídica del agrupamiento génico responsable de la biosíntesis de Tiocoralina

10 En base a los resultados anteriores, en los que las zonas amplificadas de los dominios de adenilación PSV2 y PSV5 habían sido las únicas cuya interrupción génica causaba un fenotipo no productor de tiocoralina, se eligieron dos cósmidos solapantes para ser secuenciados, cosV33-D12 (que contiene la región del dominio de
15 adenilación PSV2) y pCT2c (que contiene las regiones de los dominios de adenilación PSV2 y PSV5). El análisis de las 64.650 pb obtenidas a partir de dichos cósmidos reveló la presencia de 36 ORFs completos y 2 incompletos cuya organización se muestra en la Figura 1. La comparación con las secuencias de proteínas existentes en bases de
20 datos de los productos deducidos de los diferentes genes permitió deducir las funciones para la mayoría de ellos (Tabla 1).

Ejemplo 12. Determinación y análisis de la secuencia nucleotídica del inserto de los cósmidos cosV33-D12 y pCT2c

25

La secuenciación de ambos cósmidos se realizó utilizando la metodología habitual y para el análisis informático de la secuencia se empleó el paquete de programas GCG, de Genetics Computer Group de la Universidad de Wisconsin (Devereux et al. 1984, Nucleic Acid Res.
30 12, 387-395).

Se obtuvo así una secuencia de 64.650 nucleótidos, cuyo análisis informático reveló la existencia de 38 ORFs [36 ORFs completos y 2

incompletos] cuya organización se muestra en la Figura 1. Los productos de expresión génica deducidos de dichas ORFs fueron comparados con proteínas de función conocida presentes en las bases de datos empleando el programa BLAST (Altschul et al. 1997, Nucleic Acid Res. 25, 3389-3402), con lo que se asignaron funciones probables para la mayoría de estas ORFs (Tabla 1).

Tabla 1

Gen	Posición	Amino-ácidos	Función Deducida	Notas
<i>ORF1</i>	2-535	178*	Transposasa	SEQ ID NO:2
<i>ORF2</i>	993-1130c	46	Desconocida	SEQ ID NO:3
<i>Tio3</i>	1517-2131	205	Regulador familia OmpR	SEQ ID NO:4
<i>Tio4</i>	2154-2822c	223	Posible regulador	SEQ ID NO:5
<i>Tio5</i>	2970-3791c	274	ABC transportador (subunidad permeasa)	SEQ ID NO:6
<i>Tio6</i>	3794-4777c	328	ABC transportador (subunidad ATPasa)	SEQ ID NO:7
<i>Tio7</i>	4904-5611	236	Regulador familia MerR	SEQ ID NO:8
<i>Tio8</i>	5701-6426c	242	Triptófano 2,3-dioxigenasa	SEQ ID NO:9
<i>Tio9</i>	6426-7688c	421	Kynurenina aminotransferasa	SEQ ID NO:10
<i>Tio10</i>	7733-8524c	264	NAD- o NADP-óxidorreductasa	SEQ ID NO:11
<i>Tio11</i>	8791-10002	404	Quinaldato 3-hidroxilasa (Citocromo P ₄₅₀)	SEQ ID NO:12
<i>Tio12</i>	10022-11590c	523	3-hidroxi-quinaldato-AMP-Ligasa	SEQ ID NO:13
<i>Tio13</i>	11847-13634	596	NRPS	SEQ ID NO:14
<i>Tio14</i>	13734-15005c	424	Desconocida	SEQ ID NO:15
<i>Tio15</i>	15005-16354c	450	V-cloroperoxidasa	SEQ ID NO:16
<i>Tio16</i>	16441-18744c	768	NRPS	SEQ ID NO:17
<i>Tio17</i>	18774-19055c	94	3-hidroxi-quinaldato-Carrier Protein	SEQ ID NO:18
<i>Tio18</i>	19260-20036	259	Tioesterasa	SEQ ID NO:19
<i>Tio19</i>	20146-20880c	245	Tioesterasa	SEQ ID NO:20
<i>Tio20</i>	21188-28969	2594	NRPS	SEQ ID NO:21

Tio21	28979-38398	3140	NRPS	SEQ ID NO:22
Tio22	38449-38661	71	Desconocida (similar a MbtH)	SEQ ID NO:23
Tio23	38642-41263	874	DNA excisionasa	SEQ ID NO:24
Tio24	41835-42368	178	Regulador familia OmpR	SEQ ID NO:25
Tio25	42395-43255c	287	Posible regulador	SEQ ID NO:26
Tio26	43340-43741c	134	Desconocida	SEQ ID NO:27
Tio27	44152-49563	1804	NRPS	SEQ ID NO:28
Tio28	49635-53669	1345	NRPS	SEQ ID NO:29
ORF29	53749-55305c	519	Permeasa de glúcidos	SEQ ID NO:30
ORF30	55384-57222c	613	Permeasa de glúcidos	SEQ ID NO:31
ORF31	57895-58467	191	Regulador familia MarR	SEQ ID NO:32
ORF32	58535-59206c	224	Anti anti- σ factor	SEQ ID NO:33
ORF33	59298-59564c	89	Desconocida	SEQ ID NO:34
ORF34	59611-60114c	168	Anti anti- σ factor	SEQ ID NO:35
ORF35	60202-60888	229	Sistema regulador de dos componentes (Regulador de respuesta)	SEQ ID NO:36
ORF36	60960-62240	427	Sistema regulador de dos componentes (Histidina quinasa)	SEQ ID NO:37
ORF37	62300-62833	178	Desconocida	SEQ ID NO:38
ORF38	62925-64650	574*	Chaperón DnaK	SEQ ID NO:39

* ORF incompleta

Algunas de las proteínas identificadas están implicadas en la formación de la estructura peptídica de la tiocoralina, como por ejemplo varias de las NRPSs identificadas, Tio12, Tio17, Tio18, Tio19, Tio20, 5 Tio21, Tio22, Tio27 y Tio28. Aparecen también varias proteínas que pueden estar relacionadas con procesos de resistencia, como Tio5, Tio6, y Tio23. Los posibles reguladores de la ruta de tiocoralina identificados en la región secuenciada corresponden a Tio3, Tio4, Tio7, Tio24, Tio25.

Aparecen por último también varias proteínas relacionadas con la generación de la unidad iniciadora 3-hidroxi-quinaldato, Tio8, Tio9, Tio10 y Tio11.

- 5 Los genes cuya interrupción génica genera un fenotipo no productor de tiocoralina se encuentran señalizados en la Figura 1 mediante un asterisco (*tio20*, *tio27* y *tio28*).

Ejemplo 13. Interrupción génica en *tio28*

10

Con el fin de demostrar la implicación o no de la proteína Tio28 en la biosíntesis de tiocoralina se llevó a cabo la inactivación por interrupción génica del gen *tio28*, y en concreto del único de los dominios de adenilación que posee.

15

Se diseñaron dos oligonucleótidos iniciadores internos a este dominio de adenilación (FL-T-102up y FL-T-102rp) y se utilizaron para amplificar una zona de 1.428 pares de bases en *tio28*. Las secuencias de dichos oligonucleótidos iniciadores son las siguientes:

- 20 FL-T-102up: 5'-ACCTGAGGTACTGGGCGCAGC-3' (21 nucleótidos)
FL-T-102rp: 5'-CCGATCACCACCACCGTGGC-3' (20 nucleótidos)

- El programa de PCR utilizado fue: 2 min a 94°C, 30 ciclos (30 s a 94°C, 60 s a 53°C, 90 s a 68°C), 5 min a 68°C y 15 min a 4°C. La
25 mezcla de la reacción de PCR contenía: 1 µL de ADN molde del cósmido pCT2c, 1 µL de cada oligonucleótido a una concentración 30 pmol/µL, 7,5 µL de solución dNTPs (dATP, dTTP, dCTP y dGTP) 2 mM, 1 µL de MgSO₄ 50 mM, 5 µL de tampón de reacción para Pfx (Invitrogene), 5 µL de solución Enhancer para Pfx (Invitrogene), 28 µL de agua destilada y
30 0,5 µL de polimerasa Pfx (Invitrogene).

El producto de PCR obtenido, denominado PSV7, fue clonado en el sitio *EcoRV* del plásmido pOJ260, generándose pFL971.

La construcción pFL971 fue introducida en la cepa conjugativa *E. coli* ET12567 (pUB307) y desde ahí, por conjugación, en la cepa *Micromonospora sp.* ML1. Los clones transconjugantes se seleccionaron con 25 µg/mL de apramicina y a partir de ADN cromosómico de los mismos, se comprobó mediante hibridación Southern que efectivamente la región PSV7 había sido interrumpida. La sonda utilizada en este caso fue el producto de PCR PSV7.

El mutante *Micromonospora sp.* ΔPSV7 fue crecido en medio MT4 de producción de tiocoralina y posteriormente su micelio extraído con acetonitrilo y analizado por HPLC-MS (Ejemplo 5), dando como resultado que esta cepa no producía tiocoralina.

Etapas 4. Expresión heteróloga de tiocoralina en otros actinomicetos

Para verificar la implicación de los genes identificados en la biosíntesis de tiocoralina, se ensayó la expresión heteróloga del cluster de genes de tiocoralina en varias especies de *Streptomyces*. Se eligió la región de ADN comprendida entre las posiciones 1.393 (sitio de restricción *MseI*) y 54.301 (sitio de restricción *AcII*) de la SEQ ID NO: 1 como fragmento de ADN a ser clonado en un plásmido replicativo en *E. coli* y posteriormente en un plásmido replicativo en *E. coli*/integrativo en *Streptomyces*. Esta región de ADN contiene todas las ORFs situadas entre *tio3* y *tio28*, ambas inclusive y completas (Figura 1). La elección de esta región de ADN fue debida a que las proteínas Tio3 y Tio28 son las más externas dentro de la región secuenciada que mostraban similitudes con proteínas del metabolismo secundario.

Dicha región de ADN, debido a su gran tamaño, fue obtenida por pasos uniendo 3 fragmentos de ADN independientes, los cuales fueron

obtenidos a partir de 3 cósmidos diferentes (cosV33-D12, cosV19-B4 y pCT2c):

- fragmento A (20,2 kb): *MseI* (posición 1.393 de SEQ ID NO:1) - *NsiI* (posición 21.585 de SEQ ID NO:1);

5 - fragmento B (19 kb): *NsiI* (posición 21.585 de SEQ ID NO:1) - *EcoRI* (posición 40.636 de SEQ ID NO:1); y

- fragmento C (13,7 kb): *EcoRI* (posición 40.636 de SEQ ID NO:1) - *AclI* (posición 54.301 de SEQ ID NO:1).

10 Para facilitar el subclonaje, el fragmento de ADN completo se subclonó primero en el plásmido replicativo en *E. coli* pOJ260 (Ejemplo 14). El inserto se rescató y subclonó en un vector replicativo en *E. coli*/integrativo de *Streptomyces* que contenía el promotor de resistencia a eritromicina (*ermEp*) (pARP) [Ejemplo 16] o sin dicho promotor

15 (pAR15AT) [Ejemplo 15]. Esta región de ADN seleccionada se clonó en dichos plásmidos integrativos de *Streptomyces* pAR15AT, en ambas direcciones (Ejemplo 17) y pARP (Ejemplo 18). Finalmente, mediante conjugación intergenérica dichas construcciones se introdujeron en varios estreptomicetos (Ejemplo 19).

20

Ejemplo 14. Clonación en el plásmido de *E. coli* pOJ260 de la región de ADN seleccionada

A partir del cósmido pCT2c (Figura 2) se obtuvo mediante

25 procedimientos habituales (Sambrook et al. 2001) la región de ADN situada entre los sitios de restricción *EcoRI* (posición 40.636 de SEQ ID NO:1) y *AclI* (posición 54.301 de SEQ ID NO:1). Este fragmento de ADN se clonó en los sitios de restricción únicos *EcoRI* y *ClaI* del plásmido de *E. coli* pUK21 (Vieira et al. 1991, Gene 100, 189-194), generándose la

30 construcción pFL1023 (Figura 3).

A partir del cósmido cosV19-B4 se obtuvo, mediante procedimientos habituales (Sambrook et al. 2001), la región de ADN

situada entre los sitios de restricción *Nsi*I (posición 21.585 de SEQ ID NO:1) y *Eco*RI (posición 40.636 de SEQ ID NO:1). Este fragmento de ADN se clonó en los sitios de restricción únicos *Nsi*I y *Eco*RI del plásmido de *E. coli* pGEM-11Zf (Promega), generándose la construcción pFL1022 (Figura 3).

Estos dos fragmentos de ADN se unieron a continuación. Para ello, el fragmento de ADN situado entre los sitios de restricción *Nsi*I (posición 21.585 de SEQ ID NO:1) y *Eco*RI (posición 40.636 de SEQ ID NO:1) presente en pFL1022 fue rescatado digiriendo con los enzimas de restricción *Hind*III (situado en el sitio de clonación múltiple e inmediatamente antes del sitio de restricción *Nsi*I) y *Eco*RI. A continuación, este fragmento fue clonado en los sitios únicos de restricción *Hind*III y *Eco*RI presentes en la construcción pFL1023, generándose así el plásmido pFL1024 (Figura 3).

Toda la región clonada en pFL1024 fue rescatada como una banda *Spe*I (gracias a estos dos sitios de restricción presentes en ambos extremos del sitio de clonación múltiple de pUK21) y clonada en el sitio único *Spe*I del plásmido pOJ260, generando así el plásmido pFL1036 (Figura 3).

Por último, a partir del cósmido cosV33-D12 se obtuvo el fragmento situado entre los sitios de corte *Mse*I (posición 1.393 de SEQ ID NO:1) y *Nsi*I (posición 21.585 de SEQ ID NO:1) y se clonó en los sitios *Nde*I y *Nsi*I respectivamente de pFL1036, generándose la construcción pFL1041 (Figura 4) que contiene en pOJ260 (Bierman et al. 1992, Gene 116, 43-49) toda la región comprendida entre las posiciones 1.393 (sitio de restricción *Mse*I) y 54.301 (sitio de restricción *Acc*II) de SEQ ID NO:1, es decir, desde la ORF *tio*3 a *tio*28, ambas incluidas y completas. Además, en pFL1041, esta región se encuentra flanqueada por dos sitios de restricción *Spe*I. pFL1041 es un plásmido replicativo en *E. coli*.

Ejemplo 15. Construcción del plásmido integrativo de *Streptomyces* pAR15AT

5 El origen de replicación del plásmido pACYC184 (Rose 1988, Nucleic Acids Res. 16, 355), ori p15A, se obtuvo como un fragmento *SgrAI-XbaI* y se trató con el fragmento Klenow de la ADN polimerasa de *E. coli*. Este origen de replicación se clonó en el sitio *SmaI* del plásmido pUKA, obteniéndose así el plásmido pUO15A (Figura 5). pUKA es un
10 derivado del plásmido pUK21 (Vieira et al. 1991, Gene 100, 189-194) que contiene clonado en sus sitios de restricción *PstI-AccI* el gen de resistencia a apramicina obtenido del cósmido pKC505 (Richardson et al. 1987, Gene 61, 231-241) como una banda *PstI-EcoRI*.

15 Por medio de una digestión *BglII-XhoI* sobre pUO15A se obtiene un fragmento de ADN que contiene el ori p15A junto al gen de resistencia a apramicina *aac(3)IV*. Este fragmento se clonó, utilizando las mismas enzimas de restricción, en el plásmido pOJ436 (Bierman et al. 1992, Gene 116, 43-49), dando lugar a la construcción pOJ15A
20 (Figura 5).

En el sitio de restricción *PvuII* de pOJ15A se clonó el fragmento *DraI-BglII* (tratado con Klenow) procedente del plásmido pOJ260, y que contiene el origen de conjugación *oriT*. Se obtiene así finalmente el
25 plásmido pAR15AT (Figura 5).

Ejemplo 16. Construcción del plásmido integrativo de *Streptomyces* pARP

30 En el sitio de restricción *Ecl136II* de pSL1180 (Amersham Pharmacia) se clonó el gen de la glicosiltransferasa *elmGT* procedente de la ruta de biosíntesis de elloramicina, como un fragmento de ADN *EcoRI-HindIII* tratado con Klenow obtenido del plásmido pGB15 (Blanco

et al. 2001, Chem. Biol. 8, 253-263). Se obtuvo así la construcción pSLelmGTa (Figura 6), en la cual el gen *elmGT* se encuentra bajo el control del promotor constitutivo del gen de resistencia a eritromicina *ermE* (P_{ermE}).

5

Un fragmento *SpeI-NheI* obtenido de pSLelmGTa que contenía P_{ermE} -*elmGT* fue clonado en el sitio *XbaI* del plásmido pAR15AT descrito en el ejemplo 15, obteniéndose la construcción pAR15ATG* (Figura 6).

10

Mediante digestión *XbaI* sobre el plásmido pAR15ATG* y posterior religación se eliminó el gen *elmGT*, manteniéndose el promotor P_{ermE} , lo que dio origen al plásmido pARP (Figura 6).

Ejemplo 17. Clonación en el plásmido integrativo de *Streptomyces* pAR15AT de la región de ADN seleccionada, en ambas orientaciones

15

El fragmento de ADN *SpeI* procedente de pFL1041 (Figura 4) que contiene la región comprendida entre las posiciones 1.393 (sitio de restricción *MseI*) y 54.301 (sitio de restricción *AcII*) de SEQ ID NO:1 se clonó, en ambas orientaciones, en el sitio de restricción *XbaI* del plásmido pAR15AT (Figura 5). Se generaron así dos nuevos plásmidos, denominados pFL1048 y pFL1048r (Figura 7), con el gen de resistencia a apramicina, replicativos en *E. coli* e integrativos en *Streptomyces* mediante el sistema que utiliza la región attP del fago ϕ C31.

20

Ejemplo 18. Clonación en el plásmido integrativo de *Streptomyces* pARP de la región de ADN seleccionada tras el promotor del gen *ErmE*

25

De una forma similar, el fragmento de ADN *SpeI* procedente de pFL1041 (Figura 4) que contiene la región comprendida entre las posiciones 1.393 (sitio de restricción *MseI*) y 54.301 (sitio de restricción *AcII*) de SEQ ID NO:1 se clonó en el sitio de restricción *XbaI* del

30

plásmido pARP (Figura 6). Se generó así pFL1049 (Figura 7), en el cual la ORF correspondiente a *tio3* (SEQ ID NO:4) se encuentra bajo control del promotor constitutivo P_{ermE} presente en pARP. Este plásmido posee el gen de resistencia a apramicina, es replicativo en *E. coli* e integrativo en *Streptomyces* mediante el sistema que utiliza la región attP del fago ϕ C31.

Ejemplo 19. Expresión heteróloga de la ruta de biosíntesis de tiocoralina en distintos estreptomicetos

El plásmido pFL1048 (Figura 7) se introdujo por conjugación desde la cepa *E. coli* ET12567 (pUB307) (Kieser et al. 2000) en las especies *Streptomyces lividans* TK21 (Kieser et al. 2000) y *Streptomyces albus* J1074 (Chater et al. 1980, J. Gen. Microbiol. 116, 323-334).

El plásmido pFL1049 (Figura 7) se introdujo por conjugación desde la cepa *E. coli* ET12567 (pUB307) en las especies *Streptomyces coelicolor* M145 (Redenbach et al., 1996, Mol. Microbiol., 21, 77-96), *Streptomyces lividans* TK21, *Streptomyces albus* J1074 y *Streptomyces avermitilis* ATCC 31267.

Por último, el plásmido pFL1048r (Figura 7) se introdujo por conjugación desde la cepa *E. coli* ET12567 (pUB307) en la especie *Streptomyces lividans* TK21.

Los resultados del cultivo del clon *Streptomyces albus* (pFL1049) en medio de producción R5A (Fernández et al. 1998, J. Bacteriol. 180, 4929-4937) se muestran en la Figura 8A. En la Figura 8B se muestra el espectro de absorción del pico con tiempo de retención de 27 minutos en este cromatograma, y su espectro de masas (Figura 8C), siendo ambos idénticos a los de Tiocoralina purificada.

Reivindicaciones

1. Una molécula de ácido nucleico aislada que comprende una
5 secuencia de nucleótidos que codifica al menos una proteína de la ruta
biosintética de producción de tiocoralina, o un fragmento
biológicamente activo de la misma.

2. Molécula de ácido nucleico según la reivindicación 1, que
10 comprende una secuencia de nucleótidos que codifica la totalidad de las
proteínas de la ruta biosintética de producción de tiocoralina, o
fragmentos biológicamente activos de las mismas.

3. Molécula de ácido nucleico según la reivindicación 1 ó 2, que
15 comprende la secuencia de nucleótidos mostrada en la SEQ ID NO: 1 o
su cadena complementaria.

4. Una molécula de ácido nucleico que hibrida con la molécula de
ácido nucleico de la reivindicación 3 y codifica al menos una proteína de
20 la ruta biosintética de producción de tiocoralina, o un fragmento
biológicamente activo de la misma.

5. Molécula de ácido nucleico según la reivindicación 1, que
comprende una secuencia de nucleótidos que codifica una proteína de
25 la ruta biosintética de producción de tiocoralina, o un fragmento
biológicamente activo de la misma.

6. Molécula de ácido nucleico según la reivindicación 5,
seleccionada del grupo formado por:
30 - la molécula de ácido nucleico que comprende los nucleótidos 2-
535 de la SEQ ID NO: 1 (*orf1*);
- la molécula de ácido nucleico que comprende los nucleótidos 993-
1130c de la SEQ ID NO: 1 (*orf2*);

- la molécula de ácido nucleico que comprende los nucleótidos 1517-2131 de la SEQ ID NO: 1 (*tio3*);
- la molécula de ácido nucleico que comprende los nucleótidos 2154-2822c de la SEQ ID NO: 1 (*tio4*);
- 5 - la molécula de ácido nucleico que comprende los nucleótidos 2970-3791c de la SEQ ID NO: 1 (*tio5*);
- la molécula de ácido nucleico que comprende los nucleótidos 3794-4777c de la SEQ ID NO: 1 (*tio6*);
- la molécula de ácido nucleico que comprende los nucleótidos 4904-5611 de la SEQ ID NO: 1 (*tio7*);
- 10 - la molécula de ácido nucleico que comprende los nucleótidos 5701-6426c de la SEQ ID NO: 1 (*tio8*);
- la molécula de ácido nucleico que comprende los nucleótidos 6426-7688c de la SEQ ID NO: 1 (*tio9*);
- 15 - la molécula de ácido nucleico que comprende los nucleótidos 7733-8524c de la SEQ ID NO: 1 (*tio10*);
- la molécula de ácido nucleico que comprende los nucleótidos 8791-10002 de la SEQ ID NO: 1 (*tio11*);
- la molécula de ácido nucleico que comprende los nucleótidos 10002-11590c de la SEQ ID NO: 1 (*tio12*);
- 20 - la molécula de ácido nucleico que comprende los nucleótidos 11847-13634 de la SEQ ID NO: 1 (*tio13*);
- la molécula de ácido nucleico que comprende los nucleótidos 13734-15005c de la SEQ ID NO: 1 (*tio14*);
- 25 - la molécula de ácido nucleico que comprende los nucleótidos 15005-16354c de la SEQ ID NO: 1 (*tio15*);
- la molécula de ácido nucleico que comprende los nucleótidos 16441-18744c de la SEQ ID NO: 1 (*tio16*);
- la molécula de ácido nucleico que comprende los nucleótidos 18774-19055c de la SEQ ID NO: 1 (*tio17*);
- 30 - la molécula de ácido nucleico que comprende los nucleótidos 19260-20036 de la SEQ ID NO: 1 (*tio18*);

- la molécula de ácido nucleico que comprende los nucleótidos 20146-20880c de la SEQ ID NO: 1 (*tio19*);
- la molécula de ácido nucleico que comprende los nucleótidos 21188-28969 de la SEQ ID NO: 1 (*tio20*);
- 5 - la molécula de ácido nucleico que comprende los nucleótidos 28979-38398 de la SEQ ID NO: 1 (*tio21*);
- la molécula de ácido nucleico que comprende los nucleótidos 38449-38661 de la SEQ ID NO: 1 (*tio22*);
- la molécula de ácido nucleico que comprende los nucleótidos 38642-41263 de la SEQ ID NO: 1 (*tio23*);
- 10 - la molécula de ácido nucleico que comprende los nucleótidos 41835-42368 de la SEQ ID NO: 1 (*tio24*);
- la molécula de ácido nucleico que comprende los nucleótidos 42395-43255c de la SEQ ID NO: 1 (*tio25*);
- 15 - la molécula de ácido nucleico que comprende los nucleótidos 43340-43741c de la SEQ ID NO: 1 (*tio26*);
- la molécula de ácido nucleico que comprende los nucleótidos 44152-49563 de la SEQ ID NO: 1 (*tio27*);
- la molécula de ácido nucleico que comprende los nucleótidos 49635-53669 de la SEQ ID NO: 1 (*tio28*);
- 20 - la molécula de ácido nucleico que comprende los nucleótidos 53749-55305c de la SEQ ID NO: 1 (*orf29*);
- la molécula de ácido nucleico que comprende los nucleótidos 55384-57222c de la SEQ ID NO: 1 (*orf30*);
- 25 - la molécula de ácido nucleico que comprende los nucleótidos 57895-58467c de la SEQ ID NO: 1 (*orf31*);
- la molécula de ácido nucleico que comprende los nucleótidos 58535-59206c de la SEQ ID NO: 1 (*orf32*);
- la molécula de ácido nucleico que comprende los nucleótidos 59298-59564c de la SEQ ID NO: 1 (*orf33*);
- 30 - la molécula de ácido nucleico que comprende los nucleótidos 59611-60114c de la SEQ ID NO: 1 (*orf34*);

- la molécula de ácido nucleico que comprende los nucleótidos 60202-60888 de la SEQ ID NO: 1 (*orf35*);
- la molécula de ácido nucleico que comprende los nucleótidos 60960-62240 de la SEQ ID NO: 1 (*orf36*);
- 5 - la molécula de ácido nucleico que comprende los nucleótidos 62300-62833 de la SEQ ID NO: 1 (*orf37*);
- la molécula de ácido nucleico que comprende los nucleótidos 62925-64650 de la SEQ ID NO: 1 (*orf38*); o

10 fragmentos de las mismas que codifican fragmentos biológicamente activos de proteínas de la ruta biosintética de producción de tiocoralina.

7. Molécula de ácido nucleico según la reivindicación 1, que comprende una secuencia de nucleótidos que codifica dos o más proteínas de la ruta biosintética de producción de tiocoralina, o
15 fragmentos biológicamente activos de las mismas.

8. Molécula de ácido nucleico según la reivindicación 7, que comprende dos o más genes seleccionados entre los genes identificados como *orf1*, *orf2*, *tio3*, *tio4*, *tio5*, *tio6*, *tio7*, *tio8*, *tio9*, *tio10*, *tio11*, *tio12*,
20 *tio13*, *tio14*, *tio15*, *tio16*, *tio17*, *tio18*, *tio19*, *tio20*, *tio21*, *tio22*, *tio23*, *tio24*, *tio25*, *tio26*, *tio27*, *tio28*, *orf29*, *orf30*, *orf31*, *orf32*, *orf33*, *orf34*, *orf35*, *orf36*, *orf37*, *orf38* y fragmentos de los mismos que codifican fragmentos biológicamente activos de proteínas de la ruta biosintética de producción de tiocoralina.

25

9. Molécula de ácido nucleico según la reivindicación 1, que comprende una secuencia de nucleótidos que codifica al menos una proteína de la ruta biosintética de producción de tiocoralina, o un fragmento biológicamente activo de la misma, o un mutante o variante
30 de la misma, en donde dicha proteína se selecciona del grupo formado por las proteínas identificadas como ORF1 (SEQ ID NO: 2), ORF2 (SEQ ID NO: 3), Tio3 (SEQ ID NO: 4), Tio4 (SEQ ID NO: 5), Tio5 (SEQ ID NO: 6), Tio6 (SEQ ID NO: 7), Tio7 (SEQ ID NO: 8), Tio8 (SEQ ID NO: 9), Tio9

(SEQ ID NO: 10), Tio10 (SEQ ID NO: 11), Tio11 (SEQ ID NO: 12), Tio12 (SEQ ID NO: 13), Tio13 (SEQ ID NO: 14), Tio14 (SEQ ID NO: 15), Tio15 (SEQ ID NO: 16), Tio16 (SEQ ID NO: 17), Tio17 (SEQ ID NO: 18), Tio18 (SEQ ID NO: 19), Tio19 (SEQ ID NO: 20), Tio20 (SEQ ID NO: 21), Tio21 (SEQ ID NO: 22), Tio22 (SEQ ID NO: 23), Tio23 (SEQ ID NO: 24), Tio24 (SEQ ID NO: 25), Tio25 (SEQ ID NO: 26), Tio26 (SEQ ID NO: 27), Tio27 (SEQ ID NO: 28), Tio28 (SEQ ID NO: 29), ORF29 (SEQ ID NO: 30), ORF30 (SEQ ID NO: 31), ORF31 (SEQ ID NO: 32), ORF32 (SEQ ID NO: 33), ORF33 (SEQ ID NO: 34), ORF34 (SEQ ID NO: 35), ORF35 (SEQ ID NO: 36), ORF36 (SEQ ID NO: 37), ORF37 (SEQ ID NO: 38), ORF38 (SEQ ID NO: 39) y combinaciones de las mismas.

10. Molécula de ácido nucleico según la reivindicación 1, que comprende una secuencia de nucleótidos que comprende un *orfs* seleccionado del grupo formado por *orf1*, *orf2*, *tio3*, *tio4*, *tio5*, *tio6*, *tio7*, *tio8*, *tio9*, *tio10*, *tio11*, *tio12*, *tio13*, *tio14*, *tio15*, *tio16*, *tio17*, *tio18*, *tio19*, *tio20*, *tio21*, *tio22*, *tio23*, *tio24*, *tio25*, *tio26*, *tio27*, *tio28*, *orf29*, *orf30*, *orf31*, *orf32*, *orf33*, *orf34*, *orf35*, *orf36*, *orf37*, *orf38* y combinaciones de los mismos, o de las correspondientes regiones, mutantes o variantes de los mismos.

11. Molécula de ácido nucleico según la reivindicación 1, aislada de *Micromonospora sp.*

12. Una composición que comprende, al menos, una molécula de ácido nucleico según cualquiera de las reivindicaciones 1 a 11.

13. Una sonda que comprende una molécula de ácido nucleico según cualquiera de las reivindicaciones 1 a 11 o un fragmento de la misma.

14. Un vector que comprende una molécula de ácido nucleico según cualquiera de las reivindicaciones 1 a 11 o una composición según la reivindicación 12.

5 15. Una célula hospedadora trasformada o transfectada con un vector de la invención.

16. Célula hospedadora según la reivindicación 15, en la que dicha célula hospedadora es un microorganismo, preferentemente una
10 bacteria.

17. Célula hospedadora según la reivindicación 16, en la que dicha bacteria es una bacteria Gram-positiva, preferentemente un actinomiceto o un estreptomiceto.

15 18. Una proteína codificada por la molécula de ácido nucleico de la invención.

19. Proteína según la reivindicación 18, seleccionada del grupo
20 formado por las proteínas identificadas como ORF1 (SEQ ID NO: 2), ORF2 (SEQ ID NO: 3), Tio3 (SEQ ID NO: 4), Tio4 (SEQ ID NO: 5), Tio5 (SEQ ID NO: 6), Tio6 (SEQ ID NO: 7), Tio7 (SEQ ID NO: 8), Tio8 (SEQ ID NO: 9), Tio9 (SEQ ID NO: 10), Tio10 (SEQ ID NO: 11), Tio11 (SEQ ID NO: 12), Tio12 (SEQ ID NO: 13), Tio13 (SEQ ID NO: 14), Tio14 (SEQ ID NO: 15), Tio15 (SEQ ID NO: 16), Tio16 (SEQ ID NO: 17), Tio17 (SEQ ID NO: 18), Tio18 (SEQ ID NO: 19), Tio19 (SEQ ID NO: 20), Tio20 (SEQ ID NO: 21), Tio21 (SEQ ID NO: 22), Tio22 (SEQ ID NO: 23), Tio23 (SEQ ID NO: 24), Tio24 (SEQ ID NO: 25), Tio25 (SEQ ID NO: 26), Tio26 (SEQ ID NO: 27), Tio27 (SEQ ID NO: 28), Tio28 (SEQ ID NO: 29), ORF29 (SEQ ID NO: 30), ORF30 (SEQ ID NO: 31), ORF31 (SEQ ID NO: 32), ORF32 (SEQ ID NO: 33), ORF33 (SEQ ID NO: 34), ORF34 (SEQ ID NO: 35), ORF35 (SEQ ID NO: 36), ORF36 (SEQ ID NO: 37), ORF37 (SEQ ID NO: 38),
30

ORF38 (SEQ ID NO: 39), y combinaciones de las mismas, o fragmentos biológicamente activos de las mismas.

20. Un procedimiento para producir una proteína implicada en la biosíntesis de tiocoralina según cualquiera de las reivindicaciones 18 ó 19, que comprende crecer, bajo condiciones apropiadas, un organismo productor de tiocoralina, y, si se desea, aislar una o más de dichas proteínas implicadas en la biosíntesis de tiocoralina.

21. Un método para producir tiocoralina que comprende crecer, bajo condiciones apropiadas para producir dicho compuesto, un organismo productor de tiocoralina en el que se ha aumentado el número de copias de genes que codifican proteínas implicadas en la biosíntesis de tiocoralina, y, si se desea, aislar la tiocoralina.

22. Un método para producir tiocoralina que comprende crecer, bajo condiciones apropiadas para producir dicho compuesto, un organismo productor de tiocoralina en el que la expresión de los genes que codifican las proteínas responsables de la biosíntesis de tiocoralina ha sido modulada mediante manipulación o reemplazo de uno o más genes que codifican proteínas implicadas en la biosíntesis de tiocoralina o bien mediante la manipulación de las secuencias responsables de regular la expresión de dichos genes, y, si se desea, aislar la tiocoralina.

23. Método según cualquiera de las reivindicaciones 21 ó 22, en el que dicho organismo productor de tiocoralina es un actinomiceto, preferentemente *Micromonospora sp.*

24. Un método para producir tiocoralina que comprende crecer, bajo condiciones apropiadas para producir dicho compuesto, una célula hospedadora según cualquiera de las reivindicaciones 15 a 17, y, si se desea, aislar la tiocoralina.

25. Método según la reivindicación 24, en el que dicha célula hospedadora es un actinomiceto o un estreptomiceto.

26. Un procedimiento, basado en la utilización de genes responsables de la biosíntesis de tiocoralina de *Micromonospora* sp. ML1, para la producción de dicho compuesto en otro actinomiceto, que comprende:

- 10 (1) la obtención de mutantes afectados en genes específicos de la ruta de biosíntesis de tiocoralina;
- (2) el aislamiento de la región del cromosoma de *Micromonospora* sp. ML1 que contiene el cluster de genes responsables de la biosíntesis de tiocoralina;
- 15 (3) la obtención y análisis de la secuencia de nucleótidos del cluster de genes responsables de la biosíntesis de tiocoralina; y
- (4) la producción heteróloga de tiocoralina en otros actinomicetos.

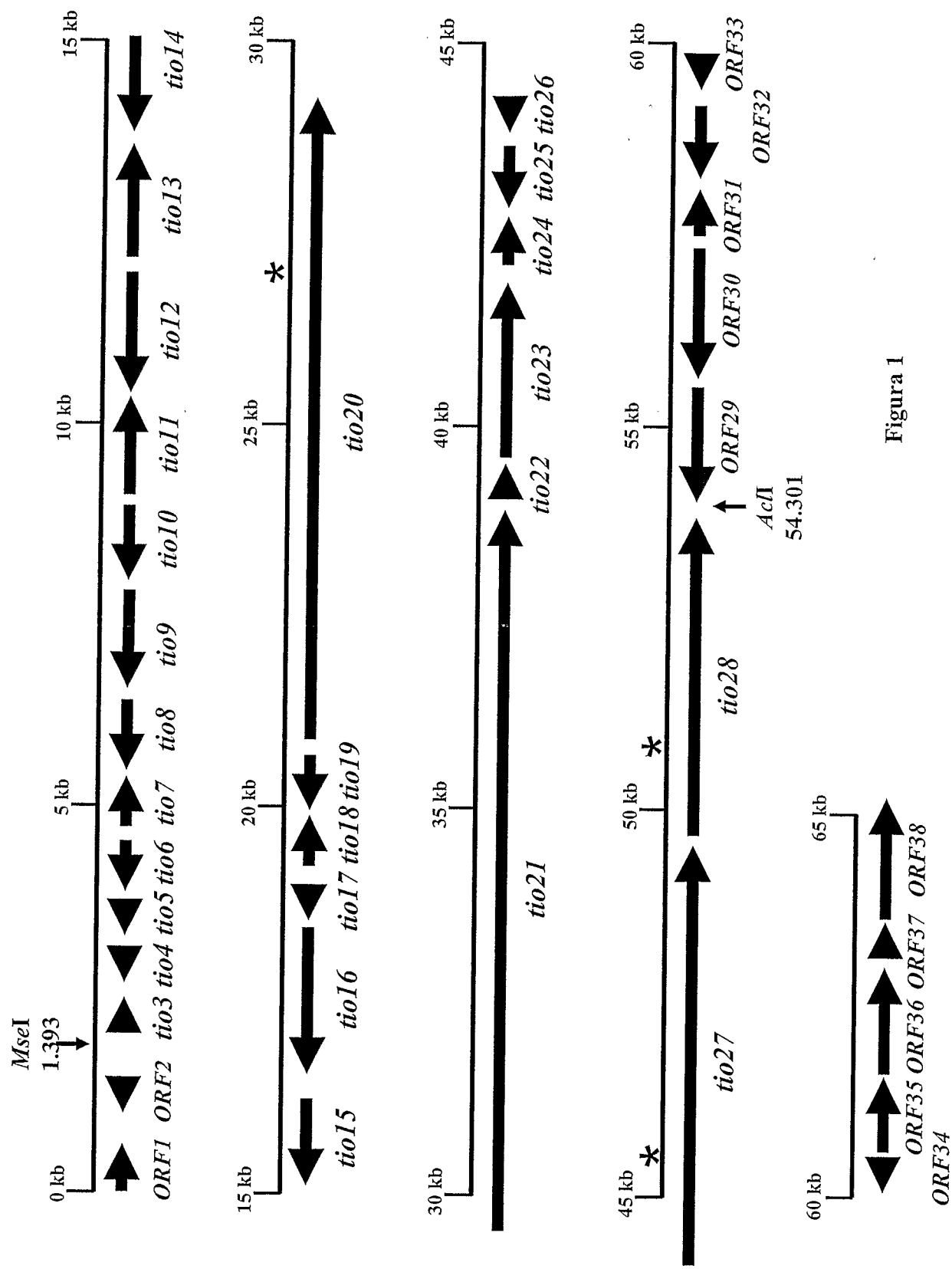


Figura 1

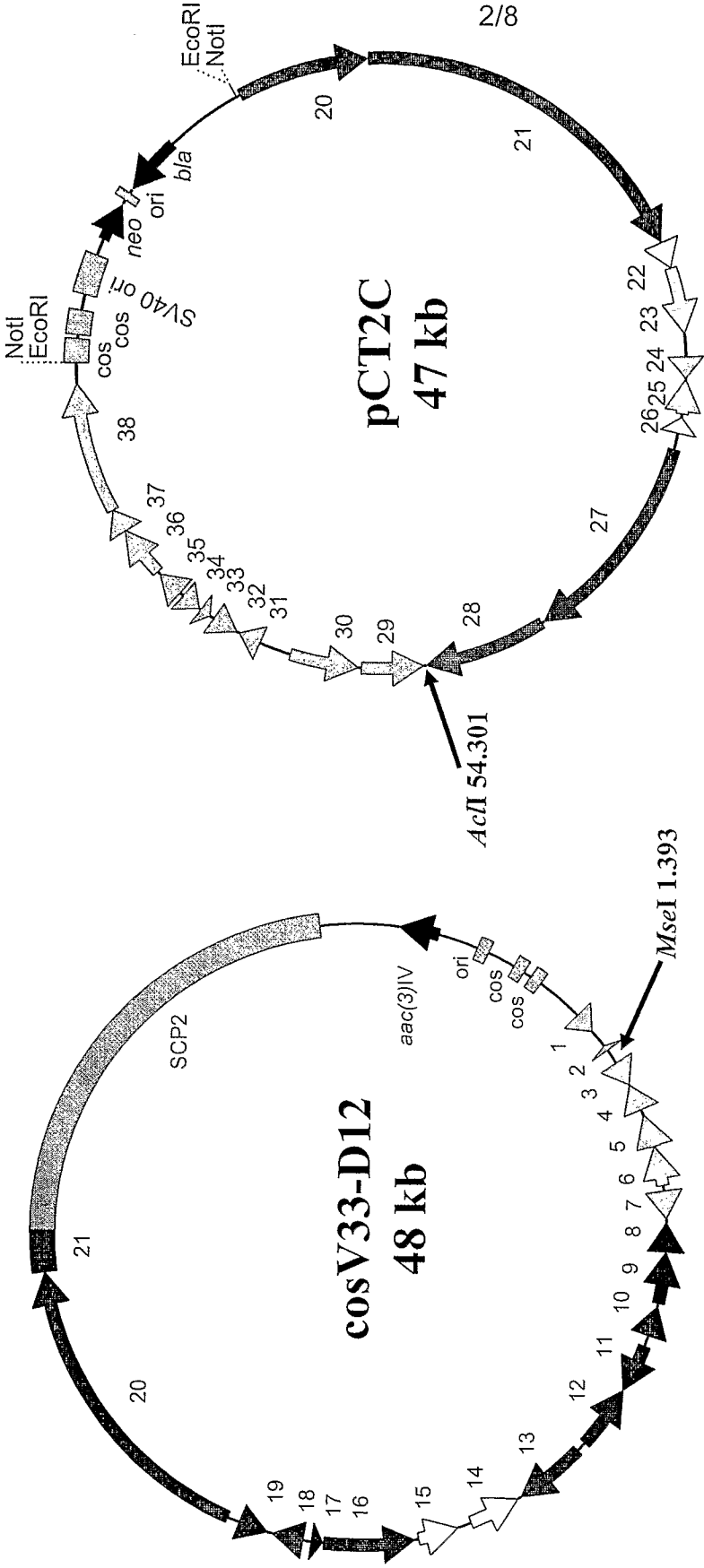


Figura 2

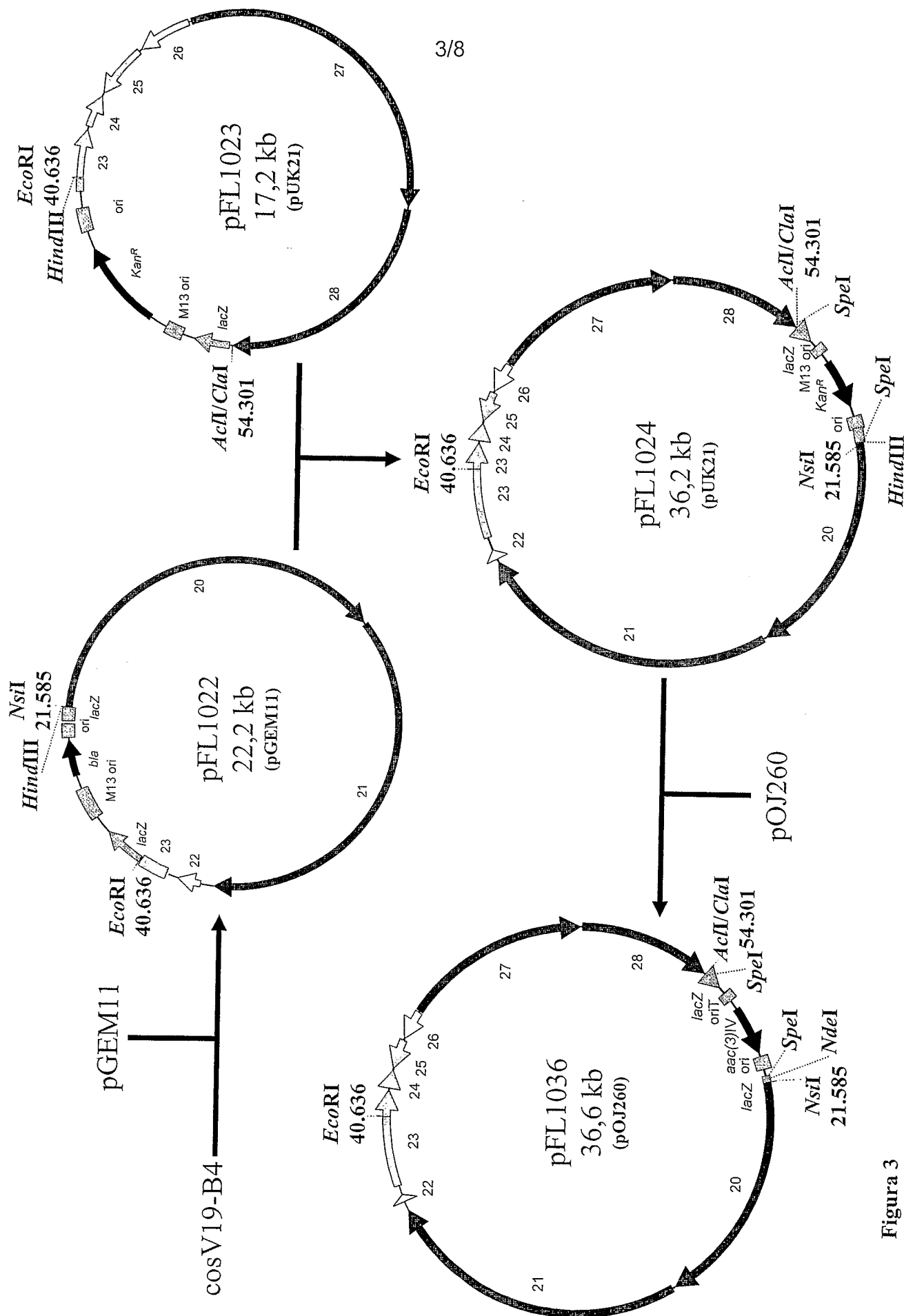


Figura 3

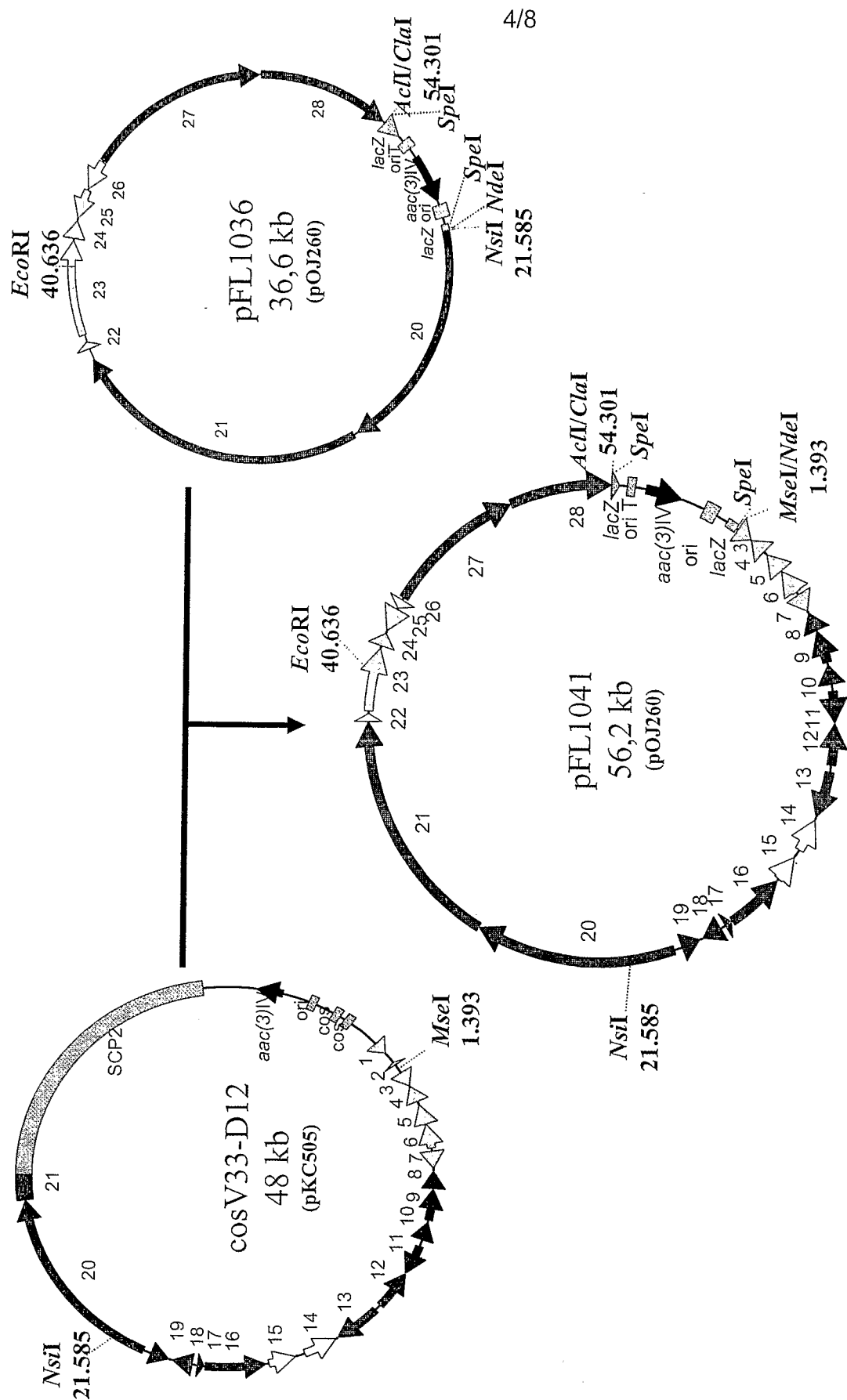


Figura 4

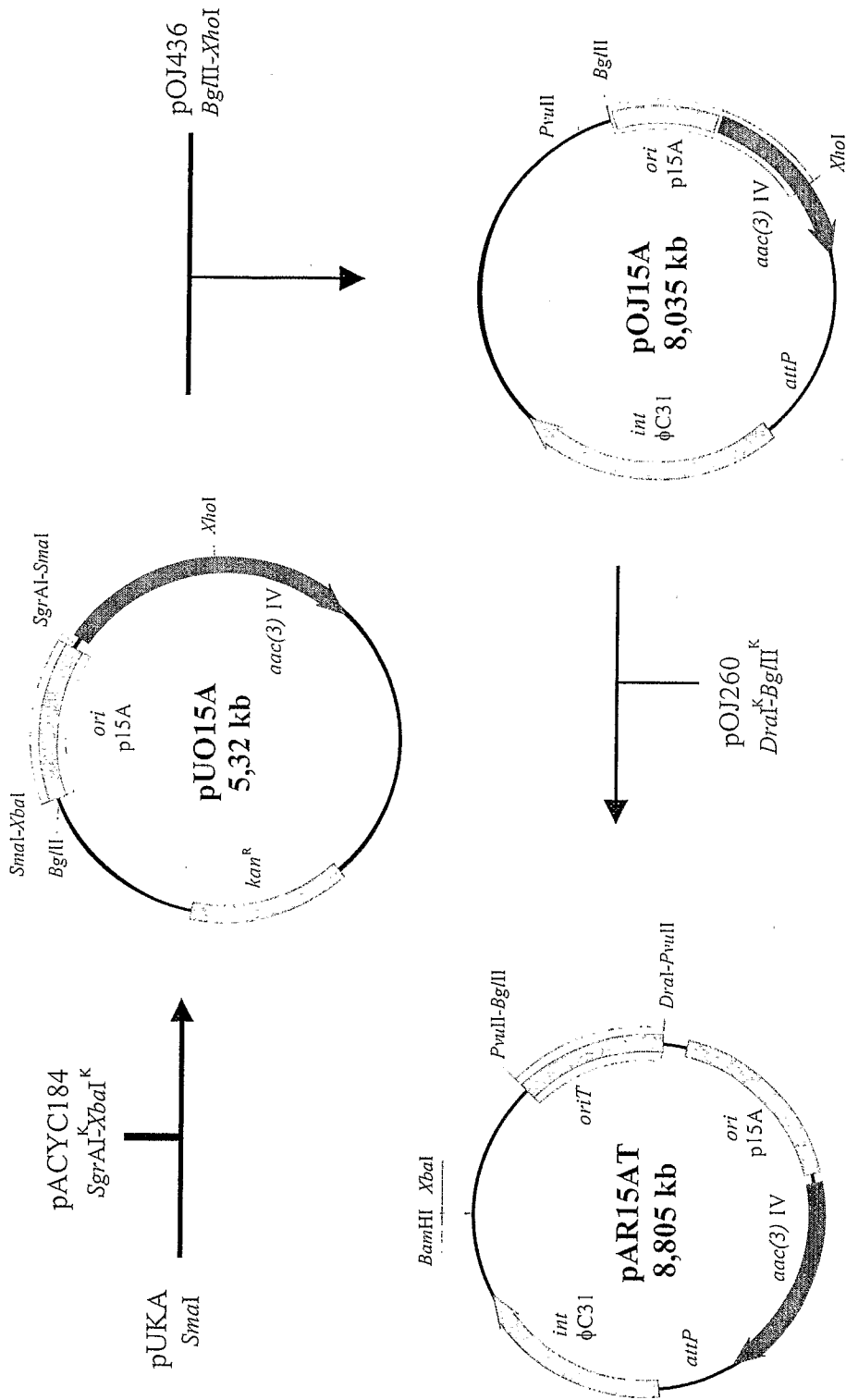


Figura 5

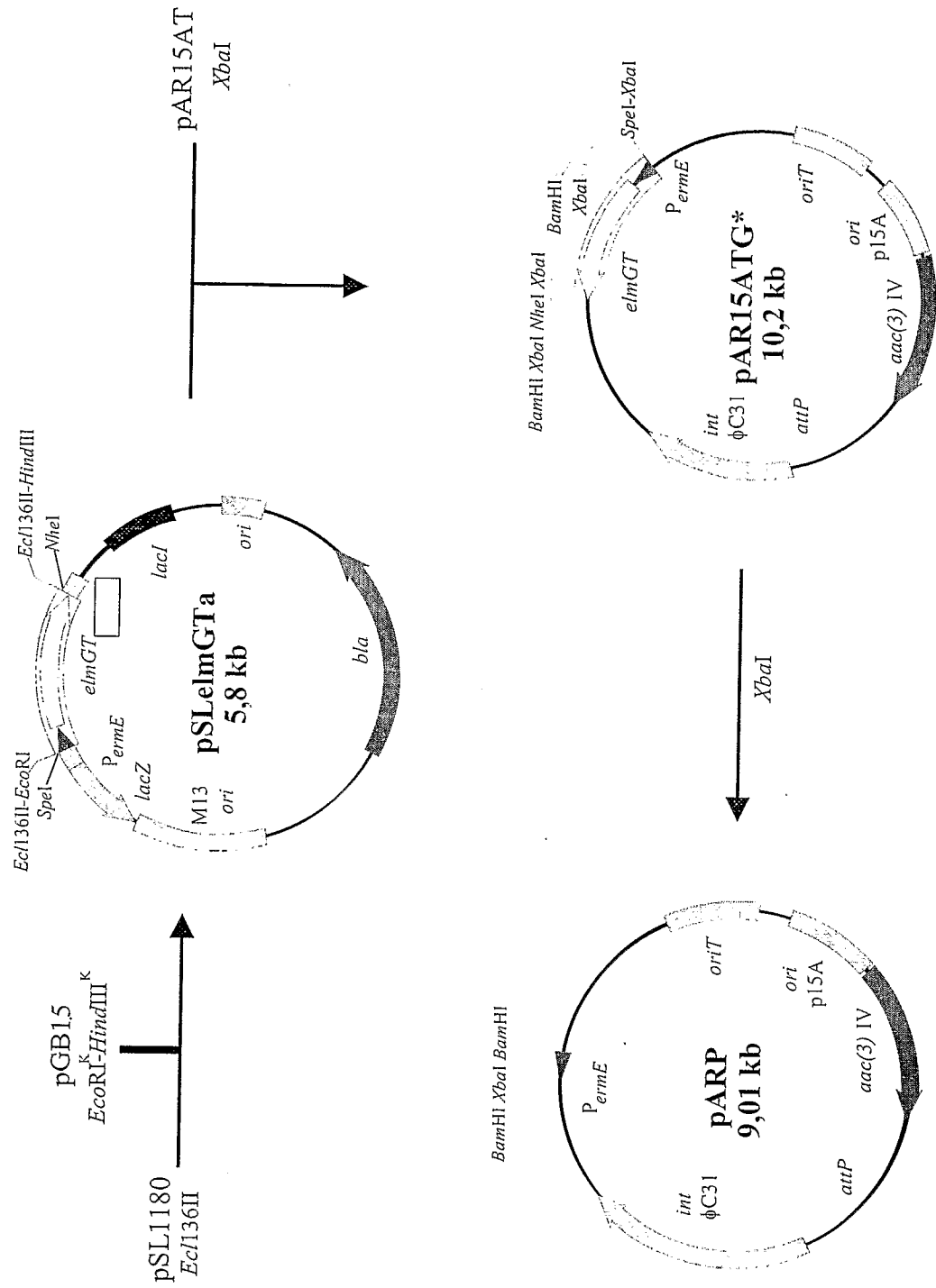
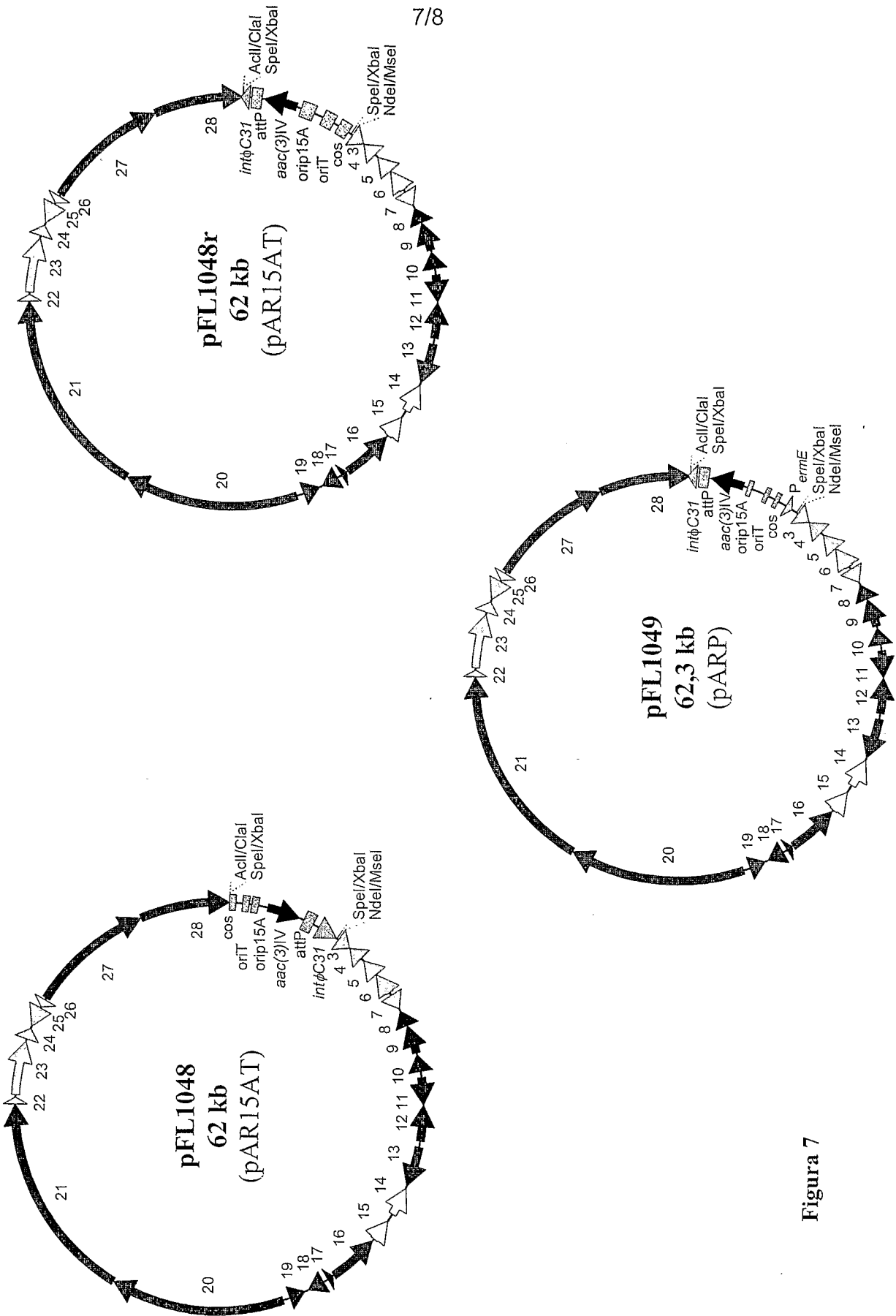


Figura 6



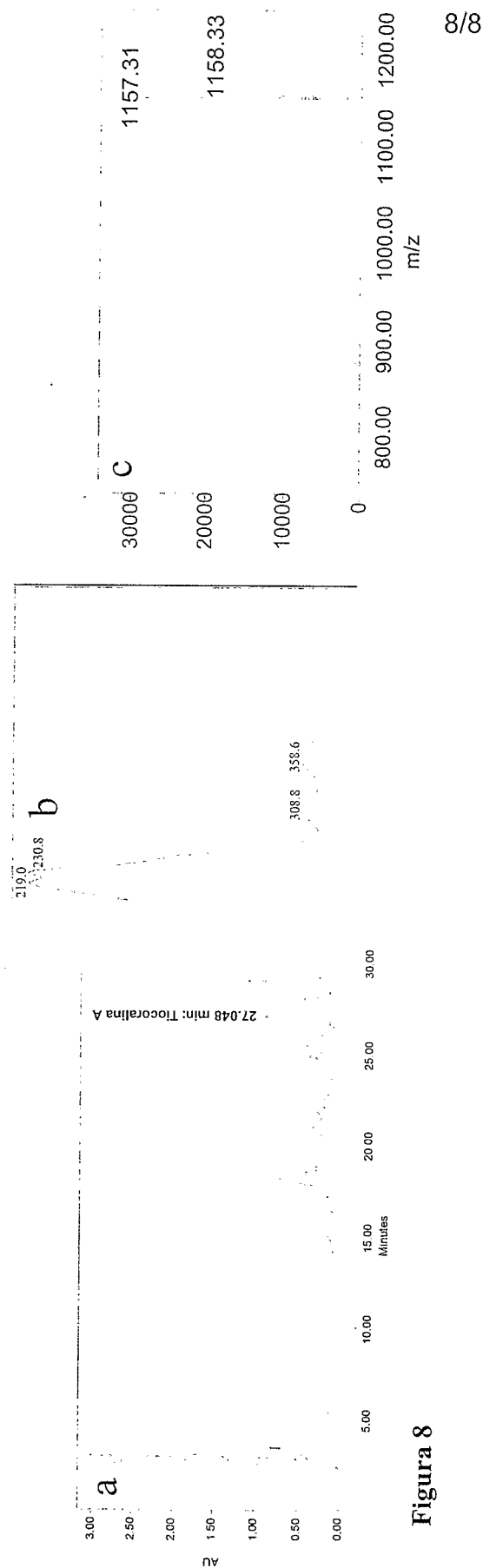


Figura 8

LISTA DE SECUENCIAS

<110> Pharma Mar, S.A.

<120> GENES INVOLUCRADOS EN LA BIOSÍNTESIS DE TIOCORALINA Y
PRODUCCIÓN HETERÓLOGA DE LA MISMA

<160> 39

<170> PatentIn version 3.1

<210> 1

<211> 64.650

<212> ADN

<213> *Micromonospora sp.*, cepa ML1

<220>

<221> CDS

<222> (2)..(535)

<223> Fase de Lectura Abierta de *orf1*

<220>

<221> CDS

<222> (993)..(1130c)

<223> Fase de Lectura Abierta de *orf2*

<220>

<221> CDS

<222> (1517)..(2131)

<223> Fase de Lectura Abierta de *tio3*

<220>

<221> CDS

<222> (2154)..(2822)

<223> Fase de Lectura Abierta de *tio4*

<220>

<221> CDS

<222> (2970)..(3791)

<223> Fase de Lectura Abierta de *tio5*

<220>

<221> CDS

<222> (3794)..(4777c)

<223> Fase de Lectura Abierta de *tio6*

<220>

<221> CDS

<222> (4904)..(5611)

<223> Fase de Lectura Abierta de *tio7*

<220>

<221> CDS

<222> (5701)..(6426c)

<223> Fase de Lectura Abierta de *tio8*

<220>
<221> CDS
<222> (6426)..(7688c)
<223> Fase de Lectura Abierta de *tio9*

<220>
<221> CDS
<222> (7733)..(8524c)
<223> Fase de Lectura Abierta de *tio10*

<220>
<221> CDS
<222> (8791)..(10002)
<223> Fase de Lectura Abierta de *tio11*

<220>
<221> CDS
<222> (10002)..(11590c)
<223> Fase de Lectura Abierta de *tio12*

<220>
<221> CDS
<222> (11847)..(13634)
<223> Fase de Lectura Abierta de *tio13*

<220>
<221> CDS
<222> (13734)..(15005c)
<223> Fase de Lectura Abierta de *tio14*

<220>
<221> CDS
<222> (15005)..(16354c)
<223> Fase de Lectura Abierta de *tio15*

<220>
<221> CDS
<222> (16441)..(18744c)
<223> Fase de Lectura Abierta de *tio16*

<220>
<221> CDS
<222> (18774)..(19055c)
<223> Fase de Lectura Abierta de *tio17*

<220>
<221> CDS
<222> (19260)..(20036)
<223> Fase de Lectura Abierta de *tio18*

<220>
<221> CDS
<222> (20146)..(20880c)
<223> Fase de Lectura Abierta de *tio19*

<220>
<221> CDS
<222> (21188)..(28969)
<223> Fase de Lectura Abierta de *tio20*

<220>
<221> CDS
<222> (28979)..(38398)
<223> Fase de Lectura Abierta de *tio21*

<220>
<221> CDS
<222> (38449)..(38661)
<223> Fase de Lectura Abierta de *tio22*

<220>
<221> CDS
<222> (38642)..(41263)
<223> Fase de Lectura Abierta de *tio23*

<220>
<221> CDS
<222> (41835)..(42368)
<223> Fase de Lectura Abierta de *tio24*

<220>
<221> CDS
<222> (42395)..(43255c)
<223> Fase de Lectura Abierta de *tio25*

<220>
<221> CDS
<222> (43340)..(43741c)
<223> Fase de Lectura Abierta de *tio26*

<220>
<221> CDS
<222> (44152)..(49563)
<223> Fase de Lectura Abierta de *tio27*

<220>
<221> CDS
<222> (49635)..(53669)
<223> Fase de Lectura Abierta de *tio28*

<220>
<221> CDS
<222> (53749)..(55305c)
<223> Fase de Lectura Abierta de *orf29*

<220>
<221> CDS
<222> (55384)..(57222c)
<223> Fase de Lectura Abierta de *orf30*

<220>
 <221> CDS
 <222> (57895)..(58467c)
 <223> Fase de Lectura Abierta de *orf31*

<220>
 <221> CDS
 <222> (58535)..(59206c)
 <223> Fase de Lectura Abierta de *orf32*

<220>
 <221> CDS
 <222> (59298)..(59564c)
 <223> Fase de Lectura Abierta de *orf33*

<220>
 <221> CDS
 <222> (59611)..(60114c)
 <223> Fase de Lectura Abierta de *orf34*

<220>
 <221> CDS
 <222> (60202)..(60888)
 <223> Fase de Lectura Abierta de *orf35*

<220>
 <221> CDS
 <222> (60960)..(62240)
 <223> Fase de Lectura Abierta de *orf36*

<220>
 <221> CDS
 <222> (62300)..(62833)
 <223> Fase de Lectura Abierta de *orf37*

<220>
 <221> CDS
 <222> (62925)..(64650)
 <223> Fase de Lectura Abierta de *orf38*

<400> 1
 gatccccagc cgcgaatggc agccacacag ctgcggcccc ggggccaaag gaccacgcga 60
 gtacctgtgg gcctggatca ccaccgccac cagcgcaggt gagcaccggt ggctgctgat 120
 gcgcccgaac cgcagcaccg gtgagctggc cttctacctg tgctgggtcac cccgcccggc 180
 gccgtgcac accctcgtga ccgtcgtctg ttcccgctgg agcatcgagg agctgttcca 240
 gtccggcaaa ggccaggtcg gcctggacca ctaccaggtc cgcggctgga ccggctggca 300
 ccgccacgtc accctggcca tgctcgccct ggccgtcctg accatcctcg ccgtcggca 360
 gcccgacccc gaccgggaga tcctcgcgct gaccgtcgcc gaaatccgcg ggctcctcaa 420
 cgccttcgtc ctgcgcgtga cgctccacc ccgcacacc ctgcactggg cgacctggcg 480
 acgaacatcc caagcccgcg cccgacgagc ccaactaccg cgcagacagg cgaagtgcg 540
 ttggagtact agccattgag ccgataattc agcgcacgac ttgccggaat cggtagggat 600
 ggtgtggcga ccatggccgc gatcgatgcc tgggcgcgtc ggcccggtcg acgcgcagtc 660
 gggcgacgcg ctggcgggtg aaggtggccg cccgtcgaac ttacgactt gaataccagt 720
 atcgggtgaag aggccgaaga ccgatatctg ccattacgta ccgatcatgt tgcaggaatg 780
 cccggaacct ttcggtatgc aggtgccag ttggcaagcg ggtgggtgcc ctctcccgca 840
 gggcttcgcg tagtcgcctg aacgtcttcc gacctcgcgt tcgtgggtgtc gcggacctga 900
 ttggagggtg atttagcacg ggtgtgggtc atcggtgatc tgcagttgct gagacaacat 960

gagacggaga	aggaccacac	cctcgcatgt	catcgctcgt	gatcgacgta	accggtcagc	1020
atgccccgcc	ccgcccgaact	caccgaactgc	tggccgcgag	gcgcttggag	gccctggcgc	1080
gggttcccca	tccgcgtgac	ccgtgccaac	cgccgccac	acgacctcat	cgccaccgtg	1140
aacagcagct	acccactat	gcaatgaccc	tgctctgtcc	cgtcgacggg	acagagcgcc	1200
cgccgcccga	ggaagaaacc	tggtgccgtc	gcgggatgct	gattatattc	agcctcatga	1260
ccctgctcat	gaagcggcgg	gcttcggctc	cgatgcgttt	gttccgccga	tatctcacta	1320
aacgccggga	ctggacagt	tgaacatcat	tgcacttcta	gctggttgac	aacgtatcta	1380
cggtcgccat	ggttaattgt	ggttgacggc	gcgtgtcgcg	actgcggagc	gccatgtcgt	1440
cgaatctcgg	ccagccgtcg	ccgtgacatt	tgtcaaccag	ctccgatgaa	ttctgtgcac	1500
tctgggggag	ttgctgatgt	cggtttcggg	aagccgccgg	agcgaaggct	cctaccgtcg	1560
ttcgggtggt	tgtgcccagg	gccgctcgcc	ccggccagcc	agggaaagac	tcaagttcgt	1620
ccgttgccgg	gccgagcggc	acctacgcga	cctgtgccgt	caggagcgga	tcccgtgcct	1680
cctggtggtc	gagggcggag	cagagccgcc	ggtctgcagt	gatccgcgtg	aggactgggt	1740
tcgtacgccg	atctcgccat	ccgaccttga	cgctcgggtg	accgctctcc	ggcaacgttc	1800
tcacgaccgg	cgtagccctg	ttctcgaccc	ggccgggacg	ttgtcgttca	ggactgcgtc	1860
cgtcgtcatg	tccagtacgc	acctcgaact	catggatctg	ttcgtcgaac	gtctcggcca	1920
ggtggtctac	cgctgggag	tgatgacct	cctcacgaag	cgcgactcga	catcaacgcg	1980
aaagcccttc	gatctgcaca	tcatgagaat	gaggaagcgg	atcgccgccg	tcgatctgga	2040
tatagttact	gcttggggcc	gtggctacgc	gctggaagaa	aagtcctcat	gggcccgggt	2100
ggccagcaag	cccgcggctg	tcgatcattg	ctgacatcgc	cgcccgatgg	ctactcagtt	2160
tcgtcgtggg	tagtcttgtc	gagggctctc	gctgcggccc	ggtcgagggt	ggagacgac	2220
cactgatcga	gttcgacgct	tgattgggtc	gcgccctggc	gccatcgggt	aaggcgtcgg	2280
tcgggtatgg	gaatgcgggg	catcagtgcg	gtcttgatgg	tttctgttgt	gcggtcggag	2340
cggtcgcctt	tgatggccgc	gcggagtgtt	ttgaccgacc	acatctgggt	ttcggcttcg	2400
tcgagccaac	ggtcctgatc	gtcgggagac	attgcggcta	cttcggcgtg	gtgttgga	2460
ctgaggccgc	ttcgacggcg	ggccagtggg	aagcggccgg	cgaccagggc	gtagttgcgt	2520
agggctctgt	agtccagtcc	tacagcctgg	acgacgtgcc	gatagcggtc	ggcgtattgt	2580
tccttgccgt	acaccagcca	gtcgcccagg	caccaggacg	aggagtgcac	gataccggag	2640
agttgggcgc	cggttcgctc	ccagtgttcg	aagctcattc	gttcgggcaa	ccgtagaccg	2700
accctggtca	gcatcacctg	ggccggctct	gctgggcgtt	gcttcggaac	gccagccgtt	2760
gctgcaggga	ccacccgga	cgctgctacg	gcgttgctgt	acgatcggct	accgcgctc	2820
atcccaactca	acctccccga	cagtcggata	actcgctatc	tctgctctct	ttttccgtct	2880
cgcgcgctg	atgtgtgatc	ggcgcgcgcg	ggatgccacc	cgtaaggccg	cccttcgcac	2940
aaaggcggcc	ttgcgggtgt	ttccgggttac	ctcgaccgca	ggctgtactt	acgggtcgcc	3000
agcggagcga	agatgggtgat	cagcaggacc	gaccaggcaa	ccgtcagcag	cacatccagc	3060
cccggcggat	tgccgttgaa	caggtcgcgc	agcgcgttga	ccgtcgcggt	catgggactg	3120
gcatcgccgt	acacccgag	ccaggtcggc	atcgactcgg	taggtacgaa	cgccgacgag	3180
atcagactga	acgggaagat	ccagatcagg	ccggccgact	gcgcgcctc	cggttgggc	3240
accgagaggc	cgatcatggc	gccgaccag	gtgagcgcgt	agccgaacag	caacagcagc	3300
gcgagtcgcg	cgagggcaga	caggacccc	ccctggaagc	ggaagccgat	cagcagacc	3360
atgagcacca	cgaccagcat	cgacacgacg	ttgcgaagga	tctccgacag	cgtagccccc	3420
acgaggaccg	cggaccggct	catcggcatc	gagcggaaac	ggtccatcag	gccgcggctc	3480
atgtcgggtg	agatgccgat	cgtggtgccg	gcgaccgaac	cgaacaacac	catctgcacc	3540
aggatgccgg	gcagcaggaa	ctgggggtag	ttcccgcctt	cgacattgat	cgccccgcgg	3600
aagacgtagg	cgaacagcag	gatgaacatg	acgggtgaa	ccagcgagaa	catcaccagt	3660
tcggggctgc	gcgtcagggt	ccgcagattg	cgacgcgcca	tctgccaggc	gtcggacagc	3720
caccaggaca	ggccggtctg	ctgctgtgcg	gcggcggcat	cgagccgcgg	tttcagggac	3780
gtggagacca	tcaccgctta	ccacccttct	tcttgctgtt	cacgtcttcg	gacgcccggg	3840
ccgtggcggg	gtgcccggtg	agggcgagga	acacgtcatc	gagcgtgggc	cggcggacgg	3900
ccacgtcggc	gattgcgatg	ccggaatcgc	gcagcggatt	gagtagctcg	ttgagcgtca	3960
taccctcggg	gaggggcgcg	tgcacctgca	tgccttcggg	atcgccgacc	ggctcggctc	4020
ccaacctctc	gcgcaggatc	tgggtgggtc	tctcgaccgc	ggcgccgctc	atgagagtca	4080
cgtagcccca	cgccctggccg	acctccgcct	tgagttcgtc	cgccgcgcgg	ctggcgatca	4140
ccttgccgct	gtcgatcacg	gacagcgagt	cgccgaggtg	gtcgccctcc	tcgaggtact	4200
gggtcgtgag	cagaaccgtg	gtgccctcct	cgaccaactc	gttgacgatg	cgccacacgt	4260
ccatgcgact	gcgcggatcg	agggcggcgg	taggttcgtc	gaggaagatg	agctggggcc	4320
ggctcatcag	ggtggcggcg	atgtccagcc	ggcggcgcgt	accgcccag	tacgtcttcg	4380
cctgccggcc	ggccgcctcg	gtcaggtcga	agcgtccag	cagttcggtc	gcccgcacct	4440
tgatgtccgg	cgtagacatc	cggaacagct	tgccgaccat	gtgcagggtc	tcccgcggcg	4500
tcagcatgtc	gtcgacgggtg	gtgtactgac	cggtcaggcc	gatgatggag	cgcacctgta	4560
gcgcgtttcg	ccgcacgtcg	aagccgaata	tctcggcctt	ccctcatcc	ggcctgagaa	4620

gggtggtcag	aaccogaacg	gtggtggtct	tgccgcaccc	gttaggtccg	agcagaccgt	4680
gcaccttacc	ctccggcacg	gtgagttcga	ggtgatcgag	agcgggtcacc	gagccgaatc	4740
tcttgccacag	tccagctact	tcaatcgctg	ctggcatctc	caggattcct	ctcaaagggg	4800
ccaacgaagg	tccggcgtgt	gatcctggaa	tctacccctg	aacttccctt	agaagctttg	4860
accgcagaat	gagcggcagc	gagtcgctat	agtctcaaac	gtgatgtatt	tccgtccggc	4920
tgatctcgcc	catgagcacg	gcctttcccc	gcaatcggta	cgcaactacg	agcgtgatgg	4980
tctcatcccg	cccgtcggc	gtacggagag	cggctaccgg	cggtagaccg	agaagcacgc	5040
cgcagcgcta	cggggttacc	gcgccctcat	cccggcgcac	gggtacgccc	aaagcggcgc	5100
gatcatgccc	gacatcactg	ccggacgggt	gaacgaggca	ttgacaacca	tcgaccgcag	5160
tcacgccgaa	ctgctgcgag	accgcggcac	cctggacgct	gtcgcgcagg	tcctgactca	5220
cctgacggga	cgctcgaccg	gcacctggcg	tgccccgaag	agcattcagc	cctacagcat	5280
cggggagttg	gcgcaccggc	tccggcgtcaa	tgtggcgacc	atccggaaat	gggagacgtc	5340
ggggtgtgtc	gcgcccgctc	gcgggccgga	gaccggggcac	cggatgtacg	acgccagcga	5400
cgtccgcgac	gcggaaactg	cgcacttcct	gcggcgcggt	cgtatccgc	tcgagctgat	5460
agccaccgtg	gtgcagcagg	tccgggcggc	cggagacacc	caggcgttgg	ccgcggccct	5520
cgcggactgg	cggcagcgcg	tcaccacgcg	cggcctggcc	atgctgaacg	ccgcgcgaca	5580
actcgccagc	tacctggcga	tggaggagaa	ctgatccggg	gcccgcagcc	cccgtgacct	5640
gctgcgcggg	gactgttccg	ggcgggtggg	gcccgtaccg	ctccaccac	cccgtactca	5700
ctgctgtgtg	gcagtcggct	gctcccacac	caacgcgtcg	tggacctgcg	cccgaagctg	5760
ccgcagctgg	gcgaaggcca	gggtgatccg	gcgggccaga	tagtcgacgc	cgggtgtgcc	5820
gcccgtgccc	ttgcgattgc	ccaccatctt	cgagaccagc	gccaggtggg	tcactctcca	5880
ccgccagtag	ccgttggcca	cgtcgagcag	gctgtcggcg	atgcggttga	gcacgccggc	5940
gcccggggccc	cgtgcgcaga	cctcggcgac	gctgatgccg	ttgtgctcgg	tcagccgctc	6000
gaaggccccg	tagaggcggc	cctcctgacg	ctcgcgcggc	agcagccgcg	cgagcaggcg	6060
gaactgcgcc	gactgcgcac	cgtcgcggcc	gtcgaggaa	gggcgggaat	cggcgaagtg	6120
acgcagcggc	aaccgttcca	gcgcgaccac	ctgctcgtgc	agtaccggga	gcagctcagc	6180
ggtgcgctgg	agaaattcca	gcgccagctc	ctcgtcctcg	ggtccgtagg	gcggcgctcat	6240
cgcgtccgcc	gcggcgctga	ggtcgtagct	cacctgcttg	agccacagct	cgcaggcttg	6300
atgggtgacg	atgaagaagt	gctcagccag	taccacggac	cggctcgcgg	tgtccggggc	6360
gcgggtctcc	tgacgggaca	gcagcgtgtc	gacctgcagg	tagtccgcgt	aggtcagggg	6420
cgtcatcggg	cgggtcccgg	gcgttctcgt	gcgaactcca	gcagctcgcg	tacgcgcgtg	6480
cgcagcaccg	acggcgctac	gcacatggcg	atccggatga	ggtgacgcga	ccagggcggc	6540
ggttcgacgg	tgatgaccgg	gccccggggc	tccaggcggg	gcacacagcg	gaaccgggtg	6600
ccgggaacca	ccgcgacgtt	ccgggtctcc	agcagccggc	tggcgaactc	ttcggcggtc	6660
aggccgggtc	gcgcgacgtc	gagcactgcg	taccagccac	cggcgggcag	ttcgggcagc	6720
aggccggcgt	cgcgtacggc	gggcagcacc	gcacgcgggt	tgctgagcac	catggcacgg	6780
ttctcgcgta	ccacgtgtcc	cagccgcagc	gcctccagcc	ccgcgtattg	caccgggggtc	6840
gacatcgcga	tgagggttgg	ctcctgcacg	atccggaaag	cgtgcgcggg	ccggctcgtt	6900
gcgggtgtga	cgtaccccaa	tccgttagcc	gtcatcgcca	ggctcttcga	gaaggagtag	6960
gagctgtgca	cgatccgctc	ggagaccggc	acatcccact	ccagtgcgcg	ggtcgagacg	7020
tgttcgtcgt	cgaagacgaa	gtgctcgtac	gctcgtcgc	tgatgacctg	ccagccccgg	7080
gcgcgcgcca	tgctgatcag	cgtgctcagg	gtcgcgggtg	cgaggacggc	cccggtcggg	7140
ttggcggttg	agttgatcaa	cagcactctg	gtgttcgcgg	tcgaggccgc	cagcaccgac	7200
tccgggtccg	tccggtagtc	gcgtcccagc	gggtagaaca	ccggacggag	accggcgagc	7260
agcacctgct	gcaggtggat	gggccagtg	agctccggca	gcaggatctc	cgcgtcgggt	7320
tccggccacg	cctgcatcag	cgcggccagg	ccctgacagg	agccgggggt	gacgaagagc	7380
cgggacacat	cgggtgtccac	gccgttcttg	tcagcgagct	tctccaccag	acgctcgcgg	7440
agctccagca	gccccctgac	cgcctgttag	tccgtgcgcc	ctgcggtcac	cgcctccggc	7500
agcgcctgcg	cgacctcggg	tggcggggtg	aagtagggct	cgcgcgacat	cagcttggtc	7560
accggaacac	cgtctgatgc	ctcgcgctcg	ccggcatcgc	ggaagacgcg	atgcagctcc	7620
gacgacgaaa	tcgtcgacgc	gaccgagggg	ccgtcgagca	gggtgggata	ggggcttggt	7680
aggggcatct	aatctccggt	tcgatcgtga	cggcggtacg	gggagggcat	caggaggtgc	7740
gaccgccgtc	gacgggaatc	agcgtccggg	tgatccagcc	ggaaaactcc	gggtcgagca	7800
ggccgcacac	ccaccggcg	atctcggcgg	gctggccgaa	ccggcccaaa	ggggtcgcgg	7860
cgagcagcga	cgaccgcagc	aacgcggact	tcacggagtc	caggcccatg	tcgtcccaca	7920
tccgtgtgtc	caagggtccg	ggtaggacgg	cgttgaccct	gacgtccggg	gccagttccc	7980
gggccagcga	tccgggtcaga	ctgttcagcg	cggccttgct	ggcgccgtag	aaggaccgcc	8040
ccggcatggc	cagcacggca	ccgacggagc	acacgttcac	caccgcgccc	cgatgctcac	8100
gcaggtgggg	tagggccgcg	cggatcaggg	cggccgggtg	gcggacgttg	gtggcgagca	8160
gccggtcgaa	gtcctcgtcc	gacgccgtgt	cgagcgtcgc	gaagcggggc	agcccggcgt	8220
tgttgaccag	accgtgcacc	gtaccgaagc	gggtgggtcg	ggcggcgacg	acgtgttccg	8280

ccgcgccccg	ctcgggtgatg	tccggcggcga	cggtcgcgac	ctgcccccca	ccgtctccgg	8340
tcagctcccc	caacgcgcc	tcgcgacgcc	cgaccgcgac	cacgttcgcg	tcgcggggcca	8400
gcagtcgttc	gccgatcgcc	cgtcccagtc	ccgtgcgggc	cccggtgacg	atcaccgta	8460
gaccggcgag	gccgtcgccc	cgctcagcca	cgccggggcca	tgccctcggt	tcgggttgggg	8520
acaccgacgg	cgtggctggc	tgggacggcc	ggccgcccag	gtgtgcacgg	ccgatatccg	8580
cgtccggtaa	cgcgatgagc	ggttgtgacc	atggcctccg	cctcgaccag	actgggtgttc	8640
gagctccaat	ctccgcctg	ggaggcggat	cggcattccat	tccgacagaa	ctgagccaga	8700
lgcgcccgcg	gccctaccga	ggcttgtgac	gatgcgggca	gttacgactt	cgccgcccga	8760
tctccggatc	ccgtacggaa	ggttctgtcg	atgagttcgc	ccacggccag	caccagcgcg	8820
ctcgacctca	ccgaccggac	gacgttcgtg	cgccacgaca	cgcacgtgtt	ctggggccgag	8880
gtgcgtgacc	acaatcccgt	ctactgggtac	cagggccgcg	aggaccgccc	cggtctctgg	8940
gtggtatcgc	gatatgtcga	cgtggctcg	tcctacaccg	acgctgcgag	gttgagctcg	9000
gcgcgcccga	ccgtgctgga	cgtgctgctc	cggggcccag	actccgcccg	cgggcgggatg	9060
ctcgcggtca	ccgatcggcc	gcgccaccgt	gaactgcgca	acgtcatgct	ccgcgcgttc	9120
tccccccggg	tgtcggccg	ggtcgtcgag	caggtccatc	gacgcgccga	cgagctgac	9180
cgcggggtca	ccgtaccgg	cgccctcgac	ttcgccaccg	agggtggccga	gagcattccg	9240
atgggacaaga	tctgcgacct	gctgtcgatc	ccaccggcgg	accggccgga	cctgttgccg	9300
tggaacaaga	agccctctc	ctccgacgag	gccgacgcca	acctgtatgc	cgcgctggag	9360
gcccccaacc	agatccgtct	ctatttcgatg	gatctggccg	agcagcgccg	ggccagtccg	9420
ggcgacgacg	tgatcagcat	gatcgccact	gccaccgtcg	gcggcgagcc	gctctccatc	9480
gacgacgtcg	cgctcaactg	ctacagcttg	atcctcgccg	gtgacgagtc	ctcacggatg	9540
tcggcgatct	gcgcgggtcaa	ggcgcttcg	gacttccccg	accagtggcg	ggccgtacgc	9600
gacggtgacg	ttgcgatcga	caccgcggtc	gaggaggtgt	tgccgtgggtc	cacgcgggcc	9660
atgcacttcg	cccgtaccgc	caccacggac	ttcgagctgc	ggggccagca	ggtgcgggcc	9720
ggagacatcg	tcaccctgtg	gaacctctcg	gccaaacttcg	acgagcgggg	gttcgaccgg	9780
ccctaccggg	tcgaggtggg	gcgtaccccg	aacaagcacc	tgctggttcg	ccatggcccc	9840
catttctgcc	tcggcgcccta	ccttgggccgc	gccgagctgc	aggcgttgct	caccgcgctg	9900
gtcggaaaccg	tttcgcgcat	cgagtccgcc	gggtcgcccc	ggcgcgctcta	ctcgaacttc	9960
ctcaacggcc	acagctcgct	gcccgtcgcc	ttcacgggtc	gctgaccggg	ccaggccctc	10020
accccgatgt	gcgcaggggtg	accttgctga	ccttgccgac	ggcggtgccc	gggaactcct	10080
cgatcggcac	caccgatcc	ggaatcttgt	agggtggccag	accgcgctcg	cggaggaagg	10140
acttgacagc	cgccggactg	atcggtcgt	cccggaggat	gacgaaggcg	acggtcttct	10200
cgcccagcac	cgggtcgccc	acgccgacca	cgcgcgctc	gcggatcgcg	cggtgggcca	10260
gaaggtgctc	ctcgatctcc	tcggcgggta	ccttctcccc	gccgcggttg	atgatgtcct	10320
tgatccgccc	gaccaccacg	atgctgccgt	ccggccgccc	gcggaccatg	tcgcgggtgc	10380
ggaagtagcc	gtcgggggtg	aacgagcgcc	ggttctgttc	ctcgccccgg	tagtagccgc	10440
ggatcgtgta	gggaccccgc	accagcagct	cgccctcgcc	gccatcgggg	acgggtcgac	10500
cgctcggtgtc	gacaacgagg	atctcgctcg	gcgcggccag	cggacgtccc	tcgggtggcg	10560
gaacgacctc	ctccggatcg	tccaggcgcg	tgtaggtcag	cagcccttcc	ccgacgccga	10620
accactgtgt	cagcacggca	ccgaggctgg	tccggacctc	ggcgggcgcg	gcggggctga	10680
acttgagct	gcccacctgg	agcatgaggt	tggggaaccg	gacgtccgac	ttgcggggcg	10740
cgtcgagcca	cagccgcacc	acgggcccga	cgagcgtcgt	gagcgtcacc	cgctcccgt	10800
ccaacagctc	gaacacctcg	tccggccgga	cgctggagggt	gagcactgcg	gtgcgcgccga	10860
ccagcagtg	gccgaggacg	ccgggacagc	cgagcgccga	attgtggccg	atcgggttga	10920
ccgccaggta	cacgtcctcc	gcgccgaacc	gcagcgcate	agccgtggcg	cgcatgttgt	10980
aggcgtagtc	gtcgtgctg	cgtgggatga	gcttggggcag	gccggtcgtg	ccgcgggaga	11040
gcaggaacag	cgcgggctcg	ctggggctga	cggcaggcag	gtcacgcccc	cgccggtaga	11100
ggtccgcccag	cgcggtgaac	tcctcgggcg	caccggcgac	cacgacatgc	ttcagtgccg	11160
gcacctcg	ccgtacctcg	cgggcccagc	cgcggtggtc	gaagcccttg	acctggctga	11220
cgacgaccag	ggcgaccgcc	tgcgaaatgtg	cgcacagggtg	ggtcagctcg	accttacggt	11280
gtcccggcag	ggccatgacg	gggatggcgc	cgatccgcag	cagcgcgtag	accaccacca	11340
cgaactcggg	cacgttgggg	agctggacga	cgactcgggtc	caacggttcg	atgcccacgt	11400
ccaacagccc	gcccgcgacg	tctgcggatg	ccctcgacag	ctcggcgtag	gtcatccgcc	11460
gctcgccgca	gacgagtgcg	gtcttgtcgg	cgtgccgggc	acatgcctcg	gtcagcaggt	11520
ctcccagcgg	ccgtccctgc	cagtacccct	cgcgggcggt	gagcgcgcca	agttcctcgg	11580
gccagggcac	acaaccatcg	agcatcgctt	ccaccaaca	gagtcgaggt	agttgaacag	11640
acacggcgtc	tcgacgccat	tgccacagcg	tgaccaagc	cgtgcggccc	ggacggcacg	11700
gcgacaggga	cgcgaaagg	gcgtccgtgg	cggcacttgc	cctgtcgcaa	tcctttcgtg	11760
gccgggcacc	ggcgcgcgag	cgagcagcta	ggttcggggc	ccgtgcagcg	ccctcgtaac	11820
cagtaccgag	aggacggagt	gagcaggtgc	ccaccgttgc	ggaattgttt	gagtcggtt	11880
ttgcgcgctc	gggcgaccac	gtggcggtga	tcttcggcga	ccgcctcgtc	agctatgcg	11940

aactcgacgc	ccgcgcaaac	cggttggcgc	gggtgctggt	cgagcgcggc	gtcgccccc	12000
acaccaccgt	ggcgggtggc	atgccgaagt	gacacgagct	gatggtggtg	ctgctggcgg	12060
tgctcaaggc	cggcggcgcc	tacctgccgc	tggaccgcga	gtatcccgc	aagcgactgg	12120
cgtacatggt	gcaggacgcc	cgccccacgc	tgctggtccg	gatggcgag	gtgtcgctgg	12180
aactcggggg	gcccggctcg	gaacttgtcg	tggacgaccc	cgacacggcg	cgggagaccg	12240
agtcgcggcc	agatcaccag	ctgaccgacg	ccagccgcaa	cgctccgctg	cgcccgga	12300
acctgatgta	cgatcatctac	acctcgggct	cgacgggtac	gcccgaaggc	gtcgccgtga	12360
cccacaccgg	cgtcgccgac	atcgctggccg	cccagaccgc	ttccatcgcg	ccccggccgg	12420
gcgaccgcgt	gctccagtgg	gcgtcggtca	gcttcgacgc	ggcgttctgg	gactggtcgg	12480
cggcggttgc	gtccgggtgcc	acgctgatca	tggccccggc	cgacgaactg	ctcccgggcc	12540
aaccgctgcg	cgacaccctg	cgccgggtacg	cggtcaccca	cgccgtgctg	cccccggtcg	12600
cgctgagcgt	caccgagccc	gacgatgtgc	tcgtcgccgg	cacgctgatg	tcgaccggcg	12660
atgctgcac	ccgcgcgctg	gtggccaagt	gggcgcggcg	ccggcggatg	tacaacggct	12720
acgggcccgc	cgagaccacg	gtcgggtcga	ccatcgccgg	gcccacgcgc	atctccgacg	12780
agatcaccat	cggtagcccg	tggtcgggca	acagggtgta	cgctcctcgac	gagcgccctgc	12840
gcccgggttc	gaacggccgc	gacggcgagc	tgtacctggc	cggaacggcg	ctagcccgcg	12900
gctacctgaa	ccggccgggc	ctgacggcga	gccggttcgt	gtcgcacccc	ttcggccccc	12960
ccggcgggcg	gatgtaccgc	tcgggcgacc	ggggccaccg	ccgcgcggac	gggcagttgg	13020
tcttcgccag	ccgcggcgac	gaccaggtga	agatccgtgg	cttcgcgcgc	gagctggcgg	13080
aggtcgaggc	gcggctggcg	gcccaccctg	cggtggatgt	cgccgcgcgtg	gtcgtggtcg	13140
gcgacctcgc	gggcgcggcg	ctggccgcct	tcgtcagcac	cggtggccggc	accgggggtt	13200
cggctgccga	actgagcgca	cactcggcgg	agaccctgcc	ggaacacatg	gtgccgtcat	13260
cgggtgcagg	gttacagcgg	ctcccgggtca	ccgccaacgg	caagatcgac	cgggcgccgt	13320
tgccggacct	cgccgagtcg	gcggcgccac	ccgacgcagc	agcctcgccg	ggtcctccc	13380
gcgaggatcc	gacctcgggtg	gagagccggc	tatgtctgct	ggtcagggag	cttctggcgg	13440
tgccggacgt	caggccggggc	gacaacttct	tcaaactggg	cgccactcgc	gtgctggcga	13500
cccggctcgc	cggacgggctc	cgcgacgagt	ggggaatcac	cggtccgac	cgggcggtgt	13560
tcgaggcggc	caccctgtcc	gagctcgccc	tggtcgtgca	ggcgtcgaag	gcatccatcc	13620
ccaacgggcg	cccctgacgc	cgacgaggaa	ccgaaccagc	gttccggcca	cacaggaaa	13680
accccctcgg	gcgatgcggg	accggcaccg	ccgaaggggg	gatccatcgg	ctacctgggg	13740
ctacagtgcc	tggggcagtc	gccacgcggg	ggcgggcccc	accaaccggc	ggtgttgtgg	13800
ttgtttccgt	tgccgctcgc	gaccgcgaag	ccggcgggacc	ggcccgaggc	ggtgaccgcg	13860
atgacgtacg	agcccaggtc	gttgtcgcgc	aggcgggtcg	tgccgtcccc	gcgcggctcc	13920
cacatgaccg	cgctgaacgg	gtacgcgcac	gagccggtga	agttgcggat	gccgacgatc	13980
tggccggcgt	cgttgacccg	gtacgccacc	cgaccgggaa	gctccaccgg	ctggcgcttc	14040
gcgctggggc	agtagatggc	cggtggtggac	tggccgggtg	cgttgatccg	ccaaccgacg	14100
atcaaccctg	tctcgttgcg	gtcgggtggcg	tacgtgtcgg	cgagaccgcc	ggtgggaagt	14160
tgctgcacgg	cgagggtgcc	gggctgccat	accaccgcgg	tggtgtgcca	ggaactgctc	14220
ccgacctggg	tgaccagccc	accggccact	ccgcgcgcgt	cgccacgcgt	gtgcgcgagg	14280
gccaccggtc	cggtgaccga	cgtggtgggt	ggcggcagcg	tggtgtacgt	gcggctggcc	14340
gggttccaca	ggtacgagcc	cacgggaaga	ccgacccccc	tggcgttggg	gccgaccacg	14400
aggaccgtgc	ccttgtcgtt	gatgtcgacg	gcctgggcac	cgcccgcgcc	gggcggcagc	14460
ggcaggacgt	cgagcccggg	gcgcgcttcc	cagacgaaag	gctgaagcag	cggttggtcg	14520
ccgtccggcg	cgggcgctcg	cgccccgatg	cccagcccga	gcgtgccagc	gacgacgccc	14580
aagcgggtga	tcgcctgcgc	tcgcgagcta	ccgcctcgcg	acggcgtgat	ttcgatcagg	14640
cggccgcocct	tgtgcgcggg	gttcgcgatg	aaggcgtggg	ccgtctgggt	gcccgggagc	14700
gacgtcacgc	cgaccacgtc	gccgcggctg	ttgaggctcg	tggcctcact	cggcgtaccg	14760
ggtccgggca	gcgaccggag	gtcgtggaac	cggctgagcg	cctcggcggg	tgccgacccc	14820
ggcccggaag	acgtgctgcc	ggtgccccct	agatgggggt	tcgcgatgag	cgctggtgtg	14880
gtctgtgtgt	cggcgagcgt	ggcgacgccc	agcacgtcgc	cggacgcggc	gaccggttcc	14940
acctgatgcg	aggctgcggg	cagtgcaggt	atcgcgacga	ccgtcagggt	cgcgactccc	15000
gtcaccagg	caccgcgacc	ggtccgacgc	caccgcgcgc	caacagggtc	aggctcgtag	15060
gggcgacctg	gcggccgagc	gcggcgccga	cctcgtcgga	gtagcggaag	tggacacccg	15120
ccaggacccg	ggcgtcgatg	ttctccttcg	tcaacgtcga	ccacggcggt	ccacgcgggt	15180
agttgcgggt	caccacccgg	ccgtcacccc	gggcgaagg	gatcgagaac	gccatcgggg	15240
tgagggttcc	gaccagctgc	gaaagcacct	gttcggcggc	cccagcgggt	accgcgtggc	15300
cgctggggta	ctcgggatgc	ggcggcgtcg	gcaggaaacga	gttcagggtg	gggtcggggg	15360
cggtagcggg	atcgccgtcg	gtgtcggcgt	cgcggatggc	cgtagccggc	cgccagtgca	15420
ggagggcgta	cttggcctcg	aagggtggcg	tgctggaatc	gacgatagcc	gcgcgagcgc	15480
ccgccagtag	ccgcaccttg	ctcctgcggg	accgggcccga	atcgccgacc	agcaggcgaa	15540
gcaccggcac	gtactgcgac	aggggggtcca	gccacgcgag	gtcgtctctgc	gcctcgggtgc	15600

gctcggaaac	ctctccgccg	agacggcgta	cctcgtccag	ggcccgcg	taaggggccc	15660
tgcccagggc	cgggcggtgt	ggcgggcgga	accgatcggc	cgggccgagc	aggaacgggc	15720
ggggcagccc	gtaaccggcc	gccaccgtct	gggtcgcccc	gggcgtgagc	cggtagaccc	15780
ccggtgcctc	ggggcgccagc	gtgaagggca	cgttcaccga	cgccagggtcg	aggccgtcac	15840
cggtacgcac	cgcgaccatc	gccgcgcggg	cctcccgaac	ggcggcgatg	ccgccttct	15900
tggcctcgct	cgcgccagc	ccggccagcg	agtcggccag	caccccgctcg	agcttgctcg	15960
cctggccggg	gaccagtgcc	ttcagcacac	cgtagcaccg	cgttgcgggtg	gcggcatacct	16020
cgaacaccgc	ccgctgcccc	ggtgcgcgca	ccggtgcccgc	ccgcgctgcg	ctgagcgccg	16080
cccacgcggc	aatccagctc	atcgcccaca	tccgctcggg	cagcccggcc	gtcgggtgccg	16140
ccgcgaacgt	cgccagcatc	tctcgaacc	aggggcccac	gacgtccggg	cggaaccgt	16200
gccggggcga	cgccgtggcg	gcctcggcct	tgggtggcag	gatcaacagg	cccgcgggtgc	16260
cggtcgcggt	gctggccggc	gcgatcagca	ggccgcgtcg	ggagagcccc	ggaaccaggg	16320
gggataggga	cggttcggta	gatcttgtcg	tcacgccat	gcctttccat	cggatacggg	16380
acgggggttc	ggtgaggggt	ccggttcggg	accggtcggg	ttcagccaca	ggctgggttca	16440
gagccggctg	agcagcgccg	ccgggtcgac	cttccccgtg	ttggtcaggg	gcagttcgctc	16500
gagggcgacg	atgcgcgtgg	ggaccatgta	cgacggcagg	atcctctgga	gggcaccgac	16560
gacctcgccc	gtgggaagcg	gctcgccgct	cggaacgtag	aaggcgcca	gcgtccgtcc	16620
gcccgcggcc	gagggagcca	ccaccgagge	ggcccgagcg	ccgcgcact	gtaggaggcc	16680
gctctgcacc	tcgcccagct	cgattcggta	cgccgctacc	ttcacctggt	cgtagaggcg	16740
ccccaggaac	tccaggtcac	cgcttgccag	ccaacgggcg	cggtcaccgg	tgtggtacat	16800
gcgttcgccc	gtgcggtggg	cccacgggtc	ggggaggaat	ttggcgggcg	tcaggccggg	16860
ggaccgggca	tagcccaggg	cgacgcattc	accggcgatg	tacagatcgc	cctccgcgcc	16920
ggtggggcag	ggccccaggc	cgtaggtcgag	tacgtagtag	cgggcgttct	gcacggctt	16980
cccgtatgga	atgctcgccc	accgggggtc	gacgtcgctt	accgggtagt	ggttcgacca	17040
gacggtgcac	tccgtcgccg	cgcccagggc	caccacctcg	gcgttcggga	aagcgggccc	17100
gatctggtcg	ggcagcgcca	ccggaatcca	gtcgccgctg	aggaggacca	ggcgagcggt	17160
gccacgcccg	gccggctcgc	ggacctggac	gaaggcgatg	accatggcca	gggcggcggg	17220
tgccgagtec	cagatggtga	tctcctcctg	ttcgaggagg	tccaccaggg	tgtccggttc	17280
ggccagttcc	gccgctgacg	ccagccgtac	ggacccaccg	gcggccagca	cgccgaagat	17340
gtcgtagacc	gagagatcga	agcagaacga	ggtgacaaag	aggatctggt	cggtcggccc	17400
gacccggtag	ctgcggttca	cccagtcgat	gaggttcact	gccgaccggg	ggctgacgac	17460
gactcccttg	ggtgcgcccc	tcgagccgga	agtgaagatg	gtgtagcaga	cgctcgtcggc	17520
cgtagggcgg	gtggcgagcg	gctcgctcgg	gcgcagctcg	atgtgatggc	cgggtccacag	17580
ccgctcgcg	cccgcgccc	ggcggtcgcc	ggatggcgcc	acgcggatca	cggctcgta	17640
cgctggggcc	agctcgccgg	tgaacgtctc	gacctggcgg	cggtggacgc	tgccgcgggt	17700
caccccgggc	aggaactcgg	gaagacgttc	gatgaagtgc	ggtgcgaccc	ggatgccgga	17760
atagtcactg	agatccgggt	cgatcaggtc	ggcgagcacg	atgacgccgt	cgtaggcgag	17820
cagtccgagc	agcgacccga	ggacgctgag	caaatagtcg	aggtccggca	ggaactgcac	17880
ggtgctggcc	agcaggatga	cgtcgaacgg	accctggacc	cgctcggcta	cctcgtgggc	17940
gaagcacact	acgccgtcga	cctcgccgcc	ggcgcggaac	ccgccctcct	ggctgcgcgt	18000
cacggacacc	ggtgaggggt	cggcggcgac	gtagcgcgac	gcgcgcgggg	ccagggttc	18060
gacgatcaga	ccggacccgg	atccgatctc	gaggatcgag	ggtcgtgccg	cgcccgctgc	18120
ggccagggcg	aaggcagcca	catggtcgcg	gtactgctgc	acatcgttct	cgtagaagg	18180
acgctccgaa	cggcgcattg	tgaagccggc	ggcctcgagc	tggctcgggg	cgctcgacac	18240
gaagtcgaag	aagttacca	cgaggtcggg	ccggaagggtg	gccgaccag	agaactcctc	18300
ggcgatgtcg	gggcagagca	cgtaggcgag	cgagcgaccc	tgccaccgaa	gctcctgcat	18360
ggagcgacgg	tgggtgcggg	cggtcagcac	cgcgccacg	tcgagctggt	cgagcagccc	18420
gaccaccggg	cgggcgggcc	acgacggatc	gatcggtagc	taggcggcac	cgacttgag	18480
cacggccagc	agcccgacga	gggcccgtgc	gtcacgcctc	ccgaagagcg	ccacggtgtc	18540
gccccgtccc	acgccgagat	cgcgtagcag	ccgcgcagg	cggttgccgc	gccggtgcag	18600
ctcccggtgg	gtcaggacct	gcttgctcga	ggtcacgac	gccgggcggt	ccgggtgccg	18660
ttcggcggcc	tcagcgagca	gatcggggat	cgtagcgtcg	gcgcggaagt	gtcgtcggt	18720
ctggttccat	acgctccagg	gcacgcccgg	agacagggct	ggcggggccc	tcacggggcc	18780
agcgacccga	acgcctcgtg	ggccgcccagc	gcgtcgggtca	gctcgttgac	tgtcggatgg	18840
tcgaagaaga	gcttgagcga	caagcggggtc	ccggccagcg	caccgagccc	ggccatgatc	18900
tcggcgatca	gcagggagtg	cccgcgatg	gcgaagaagt	cgtagctggt	gtcgaaccgt	18960
tcgtgttgga	gtgcggctgc	ccacccggcc	cgtacgtggg	cacgcagcaa	ccctcgccgc	19020
gagccgtcgt	tgccgtcgcg	atccgggttc	gccacctcgg	ccgatccgat	gtcgttcacc	19080
actcgcttcc	ggtcctctcc	cggtgccctg	cctctgcaac	tgccggccgg	tgccgcccgg	19140
tccccgacgg	tgcccgggcg	gtgccagctc	caccgcgggtg	gtccgcggtt	gtctgcggat	19200
ctttcagagg	tcagtcaggg	ttggcgctga	tgggcgtcgg	cgcggaatc	ctcacgtgca	19260

tggaacgag	aacgaagaat	cggtggttcg	tagagggtctc	gcgcagtcgg	ctcaaggagt	19320
gtgccttgta	cgccattccg	cacgcgcggc	gcggggtggc	cacagtgccg	agtctctgtg	19380
tcgaactcgg	cgacgcgttc	gacaccttcg	gcggtgcgct	gccgggacgg	gagtcgccggc	19440
tcgggtgacc	ggtggtgacg	gtcatgtccc	ggctcgccga	cgaactcgcc	ggcgaactgc	19500
tccgccatgc	ccagggggcg	ccgggtggtgc	tgtacggcca	ctgctccggc	tggcgatcg	19560
cctacgaggt	agcgcgcggt	atggaccccc	cccagctgcg	gggactcgcc	gtctcctcgc	19620
atccggcgcc	cggcagcggt	cggtggtgcg	ccacctggga	actgccgcga	caggagttcc	19680
tcaagcaggt	cgtgaccgac	gggtacctgc	cgcccgagat	catcgccaac	gaggaactgc	19740
tcgacatcta	cgagcccgcc	gtacgcgcgc	attacgagct	cgtcgagacc	tacgagttgg	19800
acctgttcgg	gggtggggccc	gtcgaggtcg	tgcgcgcgcg	gacggtggcc	atccacggtc	19860
gcgacgacca	caccatcgac	ccggtccaca	togatgcctg	gtcgggccctc	accaccgggc	19920
cgtttcgggt	cgtgtcgctg	gcgggggggc	acaacctgct	cctcggccga	ccggcgaggc	19980
tggcgcgggc	catctgccag	ggcctggcct	cagacaaccg	actccatgat	caccggtagc	20040
ggccggccca	ccgcccgta	cgtggcacag	gtgcctgccc	tgtcggcccg	actggtcgag	20100
cagggcagcc	caccggtgat	cgagggccat	caggggaagt	gtcacgccc	actcccggcc	20160
catgatgtcc	gagaccagcc	cgtcgacatc	gatggacaca	tggagtggtc	cgcccggtcg	20220
caccgtcagc	cggaaactcgc	ctgtggtgac	gtccgcccac	tcggtcagcc	tgcactcgct	20280
caccagcggg	tcgcccagcg	cgtgataaca	actcacgggt	cacgacagcg	gcgtggttcc	20340
ggagccgggg	cggtagttgg	cgtgcgcgcg	caggtccgag	cgacgcgtcg	cgcatgaaca	20400
ctcgcgtagc	tccgcgatct	cggcgacccg	ggggtcgatg	cgccccagag	cgcatagggc	20460
cgcccagaac	tgcgtcatcg	gcttcagggt	aacgtcccga	ttcttcatct	ggccgggtgg	20520
gaggcaggcc	gaggcgcaga	gcctcctcgg	ttccaccttc	atgtcccga	gcaacatcgc	20580
ggtctcgtac	gccaccatgg	cgcccaggct	gtgtccgaac	agcacctggt	cgcggcaggg	20640
cagctcaagc	aactccacca	cgatgtgccc	ggccatctcg	gccacgtcgt	ccacgaggtg	20700
ctcggcgaat	cggtcacccg	gtccggggta	ctggacggcc	acgagttcga	cccctgggtg	20760
catcgcggtc	gccagtcgcg	tgaaggcggt	cgccgagcct	cctgagtgcg	gaaaacacac	20820
caagcgcccc	gtgggcttct	cgaccgtggt	cagcgtgcgt	atccagccct	tcttccgcat	20880
ccagctcate	gatccgtcct	tccgattggt	ttccgctttg	cactggatgg	tgcgtcgttct	20940
gacgtcactc	aggtcattcg	tgaagaaat	tggtcagaca	tagctggtgc	aacggttgct	21000
ggccccaggg	ttggttcttg	ccgacatctt	tcaggccgtt	tgaggactca	acaaccggtt	21060
catagtgtgg	ggctcacgaa	cgacggcggc	ttccgcgggc	cgtggctggg	gagctgcgta	21120
cgggaggacc	caactgcgct	ccgcccgcgag	tcagtgcaga	tcgagagtta	cttattcggg	21180
agtctgtagt	ggttacggcg	gtgtttcgtc	ggcgaggag	gcaatgtggc	tggaaacagaa	21240
tcttgacact	cagggtgccga	acagcaccac	gaccatttgg	gatgtgcggg	gcgagtgca	21300
cctgcctgat	ctcgaggacg	cgttgcgga	cgccctgggc	gagacgagcg	cactcctggt	21360
gaacttccgg	cgggagagcg	atgagctgcg	gatcgtcccc	cggaagctgg	acgcgtggga	21420
gcccttccgc	attgacgcga	gcgacgcgcg	cgacctgcg	gcgcgggccc	gggaatttct	21480
caccgaacag	attcgccgcc	cgttcgacct	cggtgcggat	gcgctgttcc	ggatcggaac	21540
ggtccggctg	ggcgacgggc	atcatctgct	gtgcctgac	atgcataca	tcctgaccga	21600
cgcggttcgg	gtgttcaccc	tgtcttccca	gcgcgtcgcg	gagatctacc	gcgcctgcg	21660
cgccgattcc	cctgtgcccg	accggccgga	cacgcctgcc	gagatcgccg	gcgaacggga	21720
ggcgacgtac	cggtcgtccg	aaaggttcac	ggacgcggcg	cggttctggc	gtgattacgt	21780
cacggccggc	ggcacggctg	cccgaactgc	gaccgggatg	cgggcgggcg	cgaccctgac	21840
ggcgagcgtc	gggttctggg	acggtctgac	cgcaccgctg	ggcgctcgca	gtacacccgc	21900
cagcgtggcg	gcgagcgagg	tcgcccgggt	gcggggcgct	gccgattcga	ccgaaacgag	21960
cgtcccggac	ctgctcacct	cggcggccac	cgccttcttc	caccgcctct	gcaaacgcgc	22020
cagaccgatg	ctctccttca	ccgtcaacta	ccgcaggggg	gacatcaaac	aggcctgcgg	22080
cctgtactcg	aacgccttgc	cgtggtgat	cgaggtgcgc	gcggcaggcg	acttcgtcga	22140
cctggcgaag	tcggtgcgtc	tcgagcgtct	gcgggtgctg	cggtagcccg	agtacaacgt	22200
ctccctgata	aagcgcgcgg	ccggtcacgc	gggcgacgta	cgcagcccg	tgggcccgt	22260
cgtcaacgtc	attcccttcc	tccgggaact	cgacctcgcc	ggcgccgctg	gacggttcgc	22320
ggcgggcacg	ttcgggctac	ccgacgaggt	ccgtatctcc	acctacgtgg	acggcgggcc	22380
ggacagcgac	ctgtacatcc	gcttcgacgc	gccgggacgg	ctgtacaccc	acgaggacat	22440
cgccgggctg	gccgagaggt	tcgtgacctt	cgtccggggc	gccctgaccg	acccacgggc	22500
gcgggtccgt	gacatcagcg	ccatcgggga	gcaggaaaga	gtcgcgctcc	tcgaccgctg	22560
tgcgggcccg	gtagtcgagc	tggccggcgt	gacggtgccc	gtgctgctcg	acatgcaggc	22620
gctggccaca	ccgatgcgg	ccgcgctcgt	gcacgagaac	gccgtactga	cctaccggga	22680
gctggccgcg	aggtccgacg	ccctggcccc	cagactcgcc	gagtgccggc	cgggaccgga	22740
acgtctcgtg	gcggtcgccg	tgccccgggt	gatcgccctg	gcggtcgcgt	tgtgcgcggg	22800
cctcaaggcc	ggcgccgcct	acctgcccct	cgatccggac	tctccggcgg	agcgcgtgga	22860
gttcatgctc	ggcgacgctc	ggccgacgct	gctgctcacg	gtcgccggga	ccgacggact	22920

ggccatcacg	cccctgcacg	gagcgcagcc	acgctgggag	gccagcgtta	tcgagcggac	22980
gaccgcagcg	tcaccggcgg	ccacgtcgca	ggctggcggc	accggcgagc	acgccgcgta	23040
cctgcggtac	cggtcggggc	cattcggtcc	ggcgcccggc	gtcgtggtga	cgacgcgcgc	23100
cctcgtcaac	gagctgctct	ggatgcagca	gcgataccgg	ttgaacgaca	ccgaccggyt	23160
gctgcacaag	gccccggcat	ccctcaatgc	ctccgggtgg	gagatcttct	ggccgctgat	23220
cgccggtgcc	accgtcgtca	tggcgaccca	ggacgtccac	ggggacgcga	ccgcgctcgc	23280
ggcgatgatc	cggcggacgg	gtgtcaccac	cgtgcacgtc	gtcccgtccg	tgctcgccga	23340
gctggtgcgc	gacgaagccg	ttgtcaccgc	cggcaagggt	ctgcggcacg	tcatgtgcag	23400
cggggaagtg	ctcccggccg	acctggcagc	ccggttccac	gagagctgcg	acgtcccggg	23460
gcacagtctt	tacggggccc	tcgaggcgat	gccggtcgtc	acgtcgtgcg	actaccgacc	23520
cggctcgcgc	acggtaccgc	tcggacgacc	gatctcgaa	gccggcgctc	acgtgctgga	23580
cagccaactc	cagctgacgc	cgccgggagt	ggtcggcgag	gtgtacgtca	ccggcggtgg	23640
tctggccccg	ggctacctgc	accgcccggc	cgcgaccgcg	gaccacttcg	tcgctgccc	23700
gttcggcccc	gccggcacc	gcattgtacc	caccggtgac	ctggcgcgct	ggaccgacga	23760
cggggaactc	gagcacgccg	gcagcgccga	ccagcgggtg	cggatacgcg	gcttctgggc	23820
cgcgttctc	gagatcgaag	cagttctgac	cgcccacccg	ggagtggacc	aggccgcgct	23880
gatcgcccg	ccgacggcc	ccgaccgcgc	cgcccagctg	atcgctacg	tggtaccgc	23940
gaccggcggg	cggatgggtg	cggccggcct	ggacgtggac	ttcagcgccg	ggctggacgt	24000
ggtcgagctg	cggcggttcg	tcgctgcccc	gctgccgaa	cacctcgta	ccgccaccat	24060
cgctgtactg	gaccggctgc	cgctcacggc	cgaagggaag	cgggaccacg	ccgcgctgcc	24120
cgaacccccg	gccacccgac	gggcccaccg	ggtggccccg	accgcccagg	aacgcctcct	24180
cgcggacgcc	ttcgccgagg	tgctggggcg	cgcccgcatc	ggtctggacg	acgacttctt	24240
cgccctgggc	ggcgacagca	tctcggcgat	gcgggtggtg	gccctcgccc	gtgccggcgg	24300
cctcgcgctc	accgcgcgta	ccgtcttcga	atgccgcacc	gtcgccgccc	tgccggccgt	24360
ggcggcgccg	gccgacgaac	cagcggtcgc	cgtgcggcc	gagggcgacg	gcggacgcga	24420
cgacgcgcgc	ccgatcccac	ttcgccgat	ggcgcggctc	ttcgccgaac	gtggaccggg	24480
cctggaccgg	ctggcccagt	ggctggtgct	gacgctgccg	gcggagatga	gccccgacct	24540
gctcaccgga	gcggtccgag	cggtcctgga	ccggcacgat	gtgctgcgct	cccggctggc	24600
cgacggtcag	ctgctgatcc	agccgcgggg	atcggtggac	gtcgacaagc	tcatccggca	24660
ggtgccctgt	tcgggtgact	ggagcggggc	ctcgtggccg	gcgttactgg	cggacgaggc	24720
cgcgaaggcg	gtcggtcagc	tcagcgccct	ggccgggcag	gtgatccgca	tggtctggtt	24780
cgccgcccgc	ggcggtgcac	ccggaagact	gctcatcgcc	gcccaccacc	tggtcgtgga	24840
cggaatgtcg	tgccgggtgc	tcgtcccggg	cctcgccgag	gcccgggcac	agatccgata	24900
cgggcgtacg	ccggcactgg	ccccgggtgg	cacctcgctc	cggtcctggc	tggaacgaact	24960
cgcggccgaa	gcggtacggc	ccgagcgcgt	ggccgagctg	gaactctggc	gcggcactct	25020
cgcacccctc	ggcgatccgc	tggtgtcccg	acccctcgat	ccggcggtcg	acctgaccag	25080
gacggcccag	acgtcgcggg	tcgaggttcc	cgcgaactc	agcagggcc	tgctcacctc	25140
ggtgccggcc	accttccggt	gcgggttcga	cgatgtgctg	ctcaccgccc	ttgcgctcgc	25200
gttgccagc	cgctcgcaaca	acagcgccgg	agtcgtccgg	ctggaagggc	acggacgtca	25260
ggaggaaact	gttcccggca	ccgacctgac	ccgtacgggt	ggctggttca	ccacggtcca	25320
cccggctccg	ttcgacctgt	ccggcctcga	cctcaccgac	gcgatcaacg	gtggcccggc	25380
cgccgcggag	gcactccggc	gggtaaagga	gagccggcgc	gtcctgcccg	atcggggcat	25440
cggctactcg	ctgctgcggt	acctcaacca	ggcgaccggc	gccatgctcc	gcgacctccc	25500
cgccgacgag	gtcggcttca	actacctcgg	tcggttctcc	ttcgaccggt	cccgacggg	25560
ccgcgagatg	gcgtggacgt	ccgcgcggga	atccgcgcga	ctcgtggcgg	caccggatcc	25620
ggacatgccg	ctgatctccg	cactcgaact	gaactccctg	gtcgtcgaca	ccgaccacgg	25680
cgaacagcta	caggcgctgt	tcaccttcgc	caccgggggt	ctcgccaccg	acgaggtccg	25740
gcacctcgcc	gacgactggc	aggccgcgct	gcgtggcatc	cgagcacagg	cggccacggg	25800
cgcgagcggc	ctgaccccat	ccgacgtgcc	gttgggtcgg	gtgcggcagc	aggacctgga	25860
cgactggcag	agccgtttcg	ggcggttcga	cgacgtgtgg	ccgctgactc	cgctccagca	25920
cgggtctgct	taccacagcc	tgctggggcg	cagcgccgcc	gcgtcgtacc	agacctagtt	25980
cgctcctccg	atcaacggac	cggtggtacc	ggcgcgtctt	cgtggcgcgg	cacagacctt	26040
gctgaaccgg	cacccaacc	tgcgggtggc	cttctgacc	aagtccaccg	gcgagccggt	26100
gcaggtggtc	gtggacggcg	tcgaggtgcc	cttcgcgcac	gtcgacctcg	ccggtcaacc	26160
gcaggccgag	gagcgactgg	cccggtacct	ggccgaggaa	cgggaggccc	agttccgcca	26220
cgacagcgca	ccgctgctgc	ggttcagcct	catcaccctc	accccgacc	aggccgaact	26280
cgcgctgacc	gcacatcacg	tgctgttcga	cggctggtcg	ctgcgcgtca	tcgaacaaca	26340
gctgatgcac	ctgtacgcgc	gcggggccga	cgcgtggaa	ctggtcgagc	acggctatcg	26400
cgagttcctg	cgctggcact	cgcggcagaa	cgcgacgag	gcgctcgcgg	cctggaccga	26460
ggaactggcc	gacgcgacgc	agcccacgct	cctggcatcg	gggttcgccc	ccgcgagca	26520
ggctgggacg	agcaccgaga	tcggccaggt	cgacgtcctg	ctcaccgcgc	aggaggcggc	26580

atccgcagtg	cgtcggggccg	cggatttgccg	catcacaccg	aacgtcctcg	tgcagggcgc	26640
ctgggagggtc	gtgctcgccc	agctcacccg	aaacgcccag	gtcctcttcg	gcagctccgt	26700
caccgtacgc	ccgcccgaac	tgccggggcgc	gcacacgggtc	gtcgggatgt	tcaccaacac	26760
cgtgcccgtg	cgggtacggg	gcgcgcccgc	cgacagcctc	gaaggcatgc	tccaggccgt	26820
acaggatgcc	caggcgagga	cgttgatca	ctaccacgtc	gggctcggcg	acatccagcg	26880
ggcgaccgga	ctcactaccc	tcttcgacac	catcgtcac	ttcgagtcgt	tcccgggtgga	26940
ccgggaggcg	ttgtogaagg	ccagcgcgtc	ggcggaactg	accgtcaccg	gcacccggcc	27000
gttcgcgcgc	acgcactacc	cgtgaccgt	gctcgcggcg	gcggatccgc	tggttgagcct	27060
gacctccag	taccaccccc	acgcgcttgc	accgagcacc	gtgcaggcga	tcgcggaccg	27120
tttcgcccgc	gtgttgccgc	agttcgccgc	cgacccgcgc	acccgcatcg	ccgccatcga	27180
cgtgctctcc	gaggccgagc	ggcagctggt	ggtgcacgac	tggaacgcca	ccgcgctgga	27240
ggtgcccgc	gagttcgtgc	ccgagcagtt	cgcccggcag	gcggcagcga	caccggacgc	27300
ggtggccgctc	gtgggtggggg	agcagcacct	cacctacgcc	gagttgagcg	agcggggccaa	27360
ccggatggcg	cactggctga	tcgcgcaggg	aatcggcgc	gagagccggg	tcgtggtcct	27420
gctgccccgc	tcggcggacc	tggtggccgc	gctgctggcg	gtctggcagg	ccggcggatg	27480
ctacgtaccc	gtggatcccg	actatcccgc	cgcgcgggtg	cggctccgtga	tcgacgactg	27540
tgcggccacg	ctggtgctgg	acgagaacct	gctggaccgg	accgacctgt	cgcgtacccc	27600
ggcacacgcc	cccgcgatcg	ccgtcgcgcc	cggccaggcg	gcctacacca	tctacacctc	27660
cgactcgacc	ggcaagccga	agggcgtggt	catcggcac	gacgccctga	ccaactttct	27720
ggccagcatg	cagcgaacgc	tgccgctgtc	ggcgcaggac	cgcttcgccg	ccgtcaccac	27780
gatgccttc	gatatcgccg	cgctcgagtt	gtttctgccg	ctcacgaccg	gggcgcgcgt	27840
ggtgctcgcc	aaccgggagc	acgtcgcgaa	tccaccggcc	atgctggacc	tcgtcgagca	27900
gaccggcgctc	accgtcatgc	aggcgacccc	cgcgctgtgg	cagatgctcg	cgacgcacga	27960
cgtgcgtcgc	ctgtccggcg	tacgcgtcct	gaccggcggc	gaggcactgc	cccggccgct	28020
cgcgcagggc	ctctgcgcgc	acgcgcaca	ggtgatcaac	ttgtacgggc	cgaccgagac	28080
gaccgtctgg	tcgaccttg	cggacgtcgc	gacctgcaag	ccgccgtcga	tcggtacgcc	28140
catcgccaac	accaggtcc	tgatcctgga	ctcggcactg	cggcccgtcc	caccgggggt	28200
cgggtggagac	ctgtggatcg	caggggacgg	cctggcccgc	ggctatcacc	ggcaaccggg	28260
gctgaccgcc	agccgggttcg	tggccaaccc	gttcggccct	caggggtgcc	ggatgtaccg	28320
caaccggcag	ctggctcgct	ggtcggagag	cggaaggatc	gagttcctcg	gccgggcccga	28380
cttcagatc	aaactccggg	gattccggat	cgagccgggc	gacgtcgaac	acgccctgac	28440
ccggcatccc	gcggtcgcgc	aggccgtcgt	gacgatccgg	gaggaccagc	ccgacgacaa	28500
gcggctcgtc	gggtacgtgg	tcgtcgcgga	ggatgccgag	accccgccga	cccgggagtt	28560
gcagcagttc	gtgcgggagc	agctgcccga	ctacatggct	ccgtcggcgc	tggtaccctt	28620
gcccgcatt	ccgttgacgg	cgaacggaaa	gctggaccgt	tcgcgactgc	cccgaaccga	28680
ggcgagcagc	acgaagtacc	gggcgcggga	atccgcccgc	gaggtcgcgc	tctgcgggat	28740
cttcgcgaa	gtcctcggcg	cggagcgggt	cggcgtcgac	gacgacttct	tcgcgctcgg	28800
cgggcaactc	ctgctcgcca	cccggcttgc	cgtcggctg	cggctccgacc	tgggggtcga	28860
ggtcccgatc	cgcaccatct	tcgcgcacc	gaccgtcgcc	gaactcgcgc	ggcgggtggag	28920
cggctctgtc	acctccgtcc	ggaagcctct	gcgcaggatg	accgaaaggt	aacagagaat	28980
ggcgcgcgtc	tcgtatgcgc	aacaacgaat	gtggttcatc	aatcggttcg	agggccagac	29040
cgcgacctac	aacatcccga	tcctgatccg	gatgcgcgga	cagatcgacc	gagacgagct	29100
gcgggctcgc	ttcatggacg	tcctgtcccg	gcacgagatt	ctgcgtaccg	tctacgacga	29160
gctcgacggg	cagccgctgc	aacggatcat	cgagccggcc	gagctgagcc	tgccctggac	29220
cgaactggga	tgggtgtccc	cacagcaggt	ggacagcatc	gtcgcgcgt	cggctccgca	29280
gggcttcgac	ctgtcgggtg	acctgccgat	ccgggcgtac	ctgctgcgtc	ccgagccgga	29340
cgtggccctg	ctcgcgggtg	tgatccacca	catcgccggc	gacggcggct	cgtcggccc	29400
cttcacccgc	gacctctcgc	ccgcctaccg	cgcgcggctg	gccggggtgg	cgccacctg	29460
ggaagcagtg	ccgggtgcagt	actccgacta	cgcgctcgtc	cagcgtgagc	tgctgggcga	29520
cgagtcgac	ccggacagcg	tcttctcgac	ccagctgcgc	tactggcggc	aggaactgga	29580
cggcgtctcc	cggccgggtc	cgtcgcgcgt	cgaccggccg	cgtcccgcgc	gggcccagcta	29640
tcgcggcgac	aaccacggat	tcgagattcc	cgcggcgcgt	ttcgagaagg	tcgagcgcct	29700
cgcacagtcg	tacaacgtga	cgggtgcccat	ggtctttcag	accgcgctgg	tggtcctgct	29760
gcaccgcctc	ggcggcgacc	gcgacatcac	gctcggctcg	ccgatcgcg	gccgcaccga	29820
ccaggccctc	gagaacctcg	tgggcttctt	cgtcaacacc	tgggtgctgc	gggtgggggt	29880
gtgcaccggc	gactcgttcg	cgcaggtcct	ggaacaggtc	cggcagaagg	ccatcgccgc	29940
gtacgaccac	caggacctgc	cgttcgagcg	gctggtggag	ctgctgcgac	ccgggcgctc	30000
agtcgcgcac	cacccctgt	tccaggtcac	cctcgcctgg	cagaacagcg	tccgccagga	30060
actcgacctg	ccggcgctca	cggcctcgat	ggagccgggtc	gccaccggca	cggaagggtt	30120
cgacctgttc	gtcaacctga	tgccgaacca	ttcgcgcagg	tcggtgaccg	tcgacctgga	30180
atacgccacc	gacctgttcg	accccgccac	cgtcgaggcg	atcgcgccac	ggtacctgct	30240

cgtgctggag	caggtggtgg	agagcccgga	cctcggcgctc	ggcatcctcg	acgtaccgcg	30300
gatgccggcg	acgcgcggcg	agcggacggt	cgacgcggcg	gaacgcgaga	agatcctcta	30360
ccgctggaac	gacaccgacc	gcagctttcc	ctgcccgga	ccgatccacc	tgccgttcga	30420
gcagcaggcc	gccgcgcgtc	ccgacgcgat	cgccgtgcgc	tggtcgggcg	gcaccatgac	30480
ctaccgcgag	ctcaaccggc	aggccaaccg	gatcgccctg	accctcaagc	gccgcggcgt	30540
gcggccggag	accgtggtgg	gcgtggccgt	gcggcgcgcc	ccgctcatgg	tcgccgcctg	30600
cctcgggggtg	ctcaaggccg	gcggagcgta	cctgcccatc	ggctccacac	tgccctccga	30660
ccgggtcgcc	ggcatgctgg	ccgacgcgtc	cgcgacgctg	gtgctgacca	ccgaggacac	30720
caacaggtgg	acgccaccgg	atggcggtga	actcctcgac	gtcggccggg	ccggcctggc	30780
cctgtcgctc	gacggcgaga	tcaaccccga	gccggtggcg	agtgccgaca	acacggccta	30840
catcatcttc	acctcgggaa	gcaccggaaa	gcccaagggt	gtcactgtgg	cccaccggcc	30900
ggtgcacaac	ctgctgaact	ggtgctaccg	gaccttcgac	ttcggccccc	acgatgtgag	30960
tctctgcgtc	acctcgcgtg	acttcgacct	ctccgtgttc	gacatcttcg	gcctgctcgg	31020
ggcaggcggc	gggttctaca	tcgcggaagc	caccagcag	cgcgatccgg	agatgctgct	31080
cgacgtcctg	ctgtcggagc	cgatcacctt	ctggaactcg	gtgccgggca	cgctgaacca	31140
actggccccc	ctgctgccgc	agaccgcgg	cagcccagg	gtgcaggatc	tcgcctgggt	31200
ctttttctcc	ggcgactaca	ccccgtgtc	cctgccggac	caggtccgcg	cggtcttccc	31260
gcacgccgag	atcatcagcc	tcggcgggcg	gaccgaggcg	acggtctggt	cgaacttctt	31320
cccggtcggc	gcgatcgacc	cgcagtggcg	cagcattccc	tcggggcgcc	ccatcgacaa	31380
cgcgcgctac	tacgtcctcg	acgagaagct	ggagccgtgc	gcggtcgggt	tggagggcga	31440
cctctacatc	ggcggggccg	tgatcgctct	cgggtacgtg	aaccgtcccc	acctgacggc	31500
cgagcgggtt	atcgccgacc	cgttcggcgg	ggttcccggg	cagcgcctct	acctgacggg	31560
cgaccgcgcc	agcttcttcc	cggacggcaa	catctgttct	cttgggtcgc	ccgacggcca	31620
ggtgaagggt	cgggggttcc	gcacggaact	gggcgagatc	gagtagcccc	tcggccgcca	31680
cccggcggtg	cgccagggtg	tcgcgatcac	ccgcggggac	tcgctcgggt	acctgcgggt	31740
ggtcagctac	gtcctgcccc	atttgacac	ggtggagagc	gtgggtcggc	ccgacgaaca	31800
ggtccgcgaa	tggcaggaga	tctacgacca	gggtacacct	gaggtcaccg	accaggattt	31860
cggcgacgac	ttcaacctct	gggtgtcgag	ctacaccggc	gagcccattc	cggtcggcca	31920
gatgcgggag	tggcaggacg	ccgccgtcga	ccggatcctc	agcttcacgc	cgcggcgcgt	31980
actggagggt	ggtgccggaa	ccgggctgct	gctcgcgcgg	gttgccgggca	gcgtggaggc	32040
gtactgggcg	accgacttct	cggagccggt	catcgagcgg	ctgggcccgg	aggtgaccga	32100
ggcgggctgg	gccgagcggg	tacggctgct	gtgccgacgc	gcggacgacc	tggacgggat	32160
cccacggatc	ttcgacaccg	tagtgctcaa	ctcggctcgt	cagtacttcc	cgaacgagcg	32220
ctacctggag	caggtgctcg	acggcgctcg	ggcgatcgct	gagccagggt	gcgggctcgt	32280
gtcggcgac	atccgcgggg	cccgtcgtct	gcgcgccttc	caggtggccg	tcgacgaggc	32340
caagcacggc	aacctcccgc	cggcgcaact	gcgcagcgcg	gtcgaacagg	gcctgctcct	32400
ggagaaggaa	ctggtcatcg	acccggagtg	gttccagcgc	tgggcccagc	gcgcggggcg	32460
ggccggcgct	gacgtgcgcc	tgaaggaggg	cgcttccag	aacgagttga	cccgccaccg	32520
ctacgagatc	gtggtgcaca	agccggggcac	aacggaggcg	ggccggccct	acgcggtgga	32580
taccgtcccc	cgcttgaggt	ggaacggcga	cctggacggc	ctggccgagc	gcacccgcac	32640
gctgggcggg	ccggtggtgc	gcacgcgcgg	catccccaac	gcgcgcgtcg	cgcaggagggt	32700
ggcggcgggc	cgcgacctgg	ggctggagga	gtccgagccg	ccgacgtcgg	tgccgggtgga	32760
cccgcacgag	atggcgacct	gggcccggcg	gcagggtcgg	tcgatcgcgc	tgacctgggt	32820
cggaagcgcc	gtcgaccagt	ttgaagcagt	cctgttcacc	gacgctacga	ccgagcaccg	32880
ggcactgtcg	ggcacctacc	tcccgggtgg	cgccgcggcg	acctggatca	acgaccgggc	32940
cgccgcggcg	ggcatcaccg	cactcccggc	ggtcctgcgg	gagtacctgt	cgaagacgct	33000
gcccaggtac	atggtgccat	cggcgatcgt	gccgatcggc	cgggtgccgc	tgacgccgaa	33060
cggcaagctg	gaccggggcg	cgctgccggg	gccggagtag	gccgccgcca	gtcgcggcgg	33120
gcccgcgtgt	acctccgagg	aggagtgcgt	ctgcgagatc	ttcgcggcg	ttctcgggct	33180
cgaccgcgtc	ggcgtggacg	acaacttctt	cacgatcggc	ggccactcgc	ttctggcgac	33240
ccgggtgggtg	agccggatcc	gcgcgcgcga	cgggtcgcag	atcccgatcc	ggacgatctt	33300
cgaggcgccg	accgtcgccg	aactggtggc	ccattttgac	ggccagcgga	cggtccggcc	33360
gccgctgcgg	ccgcagccgc	ggccggaccg	gctcccgggt	tccttcgccc	agcaacggct	33420
gtgggttcac	caccgcttcg	aagggccatc	gccgacctac	atcattccgc	tctccatgcg	33480
tctgcgcgga	aaactggacg	tgcccgcgtt	gcggcaggcg	gtgcacgatg	tcgtgggtccg	33540
gcacgagtcg	ctgcgcaccg	tcttcgacga	ggtggacggg	gcaccctacc	agcgggtact	33600
cgacccgagc	gggttggcga	tcccgtggca	ggagcacgag	gtcaccgagt	ccgcccttca	33660
ggaggcgctg	cgcctggagg	cacgtcggcc	cttcgacctg	gccgcggaga	ttccgggtacg	33720
cgcgcgcgtg	ttccgcgtcg	gcgagcacga	cgcggtgctg	ctgcttccgc	tgaccacat	33780
cgcggccgac	ggctggtcgc	tggggccgct	ggccgaggat	ctgaccgccc	cgtacaccgc	33840
caggtgcgac	gggcagcgac	cgcaatgggc	cgactgcgcg	gtccagtagc	ccgactacac	33900

gctgtggcag	cgtgaccttc	tcggcggcga	tgacccacac	agtctgtacc	ggcggcagtt	33960
ggactactgg	acggggcaac	tcgccggact	gccggaatcg	ctgcaactgc	cgacggaccg	34020
ggcccgaccc	gccgtcgcg	cgttcgccgg	cgaggtgctg	cccgtagaac	tcgacgaatc	34080
gctgaccgag	gacctccgca	agttggccct	gcgtaccggg	accaccgtgt	cgatgggtgct	34140
ccaggcgggt	ctcgcggggc	tgctcagcaa	gctcggcgcc	ggtaacgaca	taccgatcgg	34200
gtcgccgata	gccggggcga	cggacgaggg	gctcaaccga	ttgggtgggct	tcttcgtcaa	34260
cacctgggtg	gcgcgcgtcg	acaccagcgg	aagccccagc	ctgatcgagc	tggtggaccg	34320
ggtacgggag	gtgagcctcg	ccgcctacga	ccaccaggac	atcccgttcg	agcacctggg	34380
cgaggcgctg	aatccgggtg	gctcgaccgc	gcatcatccg	ctgttccagg	tggtgcctggc	34440
ccttcagaac	aacgtgcagc	cggccttcga	cctgcccgga	ctgaagggtg	cccacgagcc	34500
ggtcgacatg	ggcgtctccc	gtttcgaact	gttcttgaat	ctgatggagg	agaccaccgg	34560
cgaggggacg	agccggatcg	ccggagtcgt	cgagtacgcc	acggaactgt	tcgaccggga	34620
cacggtcacc	gggtctgctg	agcgtcggcg	gcgcctgctg	cggcagtggtg	tcggcgctcc	34680
ggacgccccg	atcgacgcgg	tggagatcct	caaccgggac	gacctggccg	cccttgagcg	34740
ctggaccggg	cgcgctcgcg	gcaccgatcg	tgctcgtcgg	acgattcccg	agcgtttcgc	34800
cgccgtgggtg	gccgagcagc	cggaggtcgt	cgccctgggtg	gccgcccagc	gcgaggagag	34860
ctggacgtac	ggcgaaactcg	accggtgggc	caaccgcata	gcccaccatc	tccacgcgcg	34920
cggcgctcgg	cggcagcaac	gggtggcggt	ggtcattggag	cgggtccccgt	tgctgggtcgc	34980
ggcgggtgctg	ggcacgctca	aggcggggcg	ctgttatgtg	ccggtgggaa	ccacctggcc	35040
gcggggcccg	atcgacctcg	tgctcgccga	cctcgatccg	gccctcgtga	tcgacgagcg	35100
gctggccgag	gaggacctga	ccggctaccc	gacgcgcccg	ctggacaccg	ccgacgtcgg	35160
cgggggagcac	ctgcctacc	tgatgtacac	ctccggctcc	acggggagcg	cgaaagggtg	35220
cgaggtgagc	catcgcaacg	tgctcagttc	cgcgctcgac	ccgtgctggg	ccgatgctga	35280
ccaccagcga	gtgctgggtg	atgccccgcc	gaccttcgac	gcacgcagct	acgagatgtg	35340
ggtgccgctg	ctgcacggcg	gagcggccgt	ggtcgcgcgc	ccgggaaagc	tggaacgcgg	35400
ccgactggcg	acgctcatcg	ccgagcgggg	cgtaaccggcg	ctgtggctgc	cggcgggcct	35460
gttcgatcta	atcacgcagc	atcacccgaa	gagcttcgtc	caggtaacgc	aggtgtgggc	35520
cggcggcgac	gtaactgtcc	cggcggccgt	gcgacggctg	gtacgcgacg	acggcacgct	35580
caccgtcgtg	aacgggtacg	gtcccaccga	gaccaccacc	ttcgcggccc	ggtaccggat	35640
gtccgcaccg	gcgcgggtga	aggacccctt	gccgatcggc	gagccgatgg	cgggcagccg	35700
cctctacgcg	ctcgacgacc	ggctccggca	ggtgcgcgag	ggcgtcatcg	gcgaactcta	35760
cgctcgtggc	gacgggggtg	ctcgcggtga	cgcgaacctat	cccccgctga	ccagcgaacg	35820
attcgttcgc	gatacccttcg	gacgcocggg	cgaagcgatg	taccgcaccg	gtgacctggt	35880
ccggtggaac	cacgacgggc	aactcgagtt	cctgggcccgc	gtcgacgagc	aggtcaagat	35940
ccgcggttcc	cgggttgagc	cgggagagat	ccgagcggcc	ctccgggaa	gggacggcgt	36000
ggcacaggcg	gtgggtggtc	cccgtacgga	ccgactcggc	gagcggcgcc	tggtggccta	36060
cgtggttccg	gaggtgcccg	cgggcgcgga	cgaagactcg	accgaacatg	tcgagaagtg	36120
gcgcgccatc	tacgactcga	tgtacgacga	gaccgagggc	gacgctaccg	agatcggcaa	36180
cgacttcacc	gggtggaaga	gttccctaac	gcgcgacaa	atcccgtctg	ccgagatgcg	36240
ccggtggcgc	gactccgtcg	tcgaggaggt	ccgcgggtctg	cgggcgcgcg	ggatccctgga	36300
gatcggggtc	ggttcggggc	tcctgctcgg	gccgttggcg	ccggaagcgg	aggcgtactg	36360
gggcaccgac	ttctcgttgc	cggtcacatga	gcgcctcgag	gtgcagggtc	gcaccgatcc	36420
gtgcctgaag	gagaagggtg	cgtcgcgggtg	ccagcacgcg	gacgtcgcgg	acggactgcc	36480
ggtgaagtac	ttcgacaccg	tcatectgaa	ctccgtcgtg	cagtaactttc	cggacgccgc	36540
ctatctctcc	cgggtactgg	acgtcgcctt	ggaccgggtc	gcgcggggtg	gccgatccct	36600
cgctcggtag	gtccgggaact	acggcacccct	gcgggagttc	ctgacggccg	tccatcacgc	36660
gcaacaccgc	caggacagcg	cgtcggccgt	ccgggcggcg	gtcgagcgtg	ccgtcctggc	36720
cgagaaggaa	cttgtcacog	atccggattt	cttcaccgag	tgggcacgga	cccggcccga	36780
cgtcgtggcc	gtcgacatcc	ggctgaagcc	gggcgcggac	cagaacgaac	tcacccggca	36840
ccgttacgag	gtcatcctgc	acaagcagcc	gagccaaaca	ctgcgcctcg	cggacgtgcg	36900
gacggcgaac	tggggctcgg	agggtcccga	cctgtccggg	ctggagaccg	cactggcccc	36960
gcacggcggt	cgcctccggg	tggcccggat	ccggaacgcc	aggctgggtg	gcgaggcggt	37020
ccagtgcggc	gtgcccacca	acgtaggcgg	cacgcccctg	gacccgcacg	agttggcgag	37080
ctggggaggg	cagcgggggt	actcgggtcca	ctgcacgtgg	tcggcggagg	cgcgccgctg	37140
gttcgaggcg	gtgatcatcc	cggtcgacag	cgggcactgc	cgcgacgggtg	tgtaccggcc	37200
ggtggggccc	cggccacggc	agttgggtcaa	cctgcccgcc	goggcgcgcg	gggtgtctcg	37260
gctgccctcg	tggctgcgcg	aggaaactgg	ggcggagctg	ccggagcacc	tggtgccggg	37320
cgacatcgtg	gtcatggaac	gactgcctgt	gaccaccaac	gggaagatcg	accacagcag	37380
acttcgggag	gtggagcagg	ccgaccggga	gtacggctcg	ccgcgcacac	ccctggaggc	37440
ggaactggcc	ggtctcttcg	ccgaggtcct	cggcgtggac	cgggtgggca	tcgaggacga	37500
cttcttcgac	tgccggcggt	gttctctcca	ggtcatccgg	ctcatctggc	ggatccgcgc	37560

ggagctgggc	ttogacatto	cggtcagaac	catcttccag	catccgaccg	tgcccgaggt	37620
cgcggaacat	ctggcgggccg	gcagggagga	cgctcaggttc	gacgacccct	tctcgggtggt	37680
gctaccgatac	cggaccgaacg	gggaccgggc	accgatctgg	ttggtgccgc	cggttgccgcg	37740
actctccttg	gcctatctcg	gcttcgcccc	gcacctggac	cgctcgcgcc	ccatctacgg	37800
cctccaggcc	cgtggatttg	tcggagaccc	gcgggcgacc	tcgatcgaga	gcatggtgga	37860
cgactacgtc	gaccagatcc	tgcggttcca	gccggaaggt	ccctacaacg	tgctgggttg	37920
gtcgctgggc	ggcccgatcg	cccaggcggt	ggccgcgcaa	ctccagcgcc	gggagtgcga	37980
ggtggatctg	ctggccgtcc	tcgactcggg	tcggcgagc	tacttcgccg	acttccagac	38040
tcccgcagag	aagttggtgc	gcaggtacct	ggcgccactac	atggggccatc	tcgcgggaat	38100
ggcggagttc	gagtcgctcg	tgacagacatc	caccaccctg	ttcatcgagc	acacggacct	38160
gatgacgcaa	ttacgctcgc	cgcggttccg	tgccgatctg	atcttctttg	ccgccctcat	38220
cgaccagggc	acccgggaac	gacggcaact	cgaggtggaa	ctggacgtcc	aatggcgga	38280
atacaccgag	gggaccgtgc	ggcgcttcga	catcgaatgc	gcccacaacg	aatgatgtg	38340
gcccgagaac	gcggccatga	tcagccgtat	cgtaacagag	atcatcgatt	cgccccgata	38400
accgcttccg	ccctgacctg	gagcggacgt	gacaactacg	gaggaaagat	gagcgtcaac	38460
ccgttcgatg	acgaggacgg	cgagttctac	gtgctggtca	acgacgagga	gcagcactcg	38520
ctgtggcccc	ccttcggtga	cgtgccgac	ggctggcgaa	tcgtgttcgg	gccggccggc	38580
cgagcggaga	gcgtggcata	cgtggaggag	aactggaccg	acatgcggcc	caagagcctg	38640
cgtgaggcga	tgctggcagc	atgacgtacg	tcgtgggctc	gcaggcagac	gcgtccgcct	38700
gcgggccccac	gattttgaac	ttccgcgacg	ccggtggcg	aaggtgcgta	cagaagtctt	38760
cacaggatgt	ctaggtatag	aactgggatg	tggaagccg	gagatcgccg	ggggagtgat	38820
cagcccggtg	caccccgccg	gctccgcgtc	gcgcgccaga	cactttggag	gccccatgac	38880
tggtgctcg	cccaggcaga	gccgagaaga	cctatcgctc	gtcccgaca	cccatgacgt	38940
catccgcgtg	gtcggggcac	gggagaacaa	cctcaaggac	atccgactgg	agatcccga	39000
gcgcgggtg	accgtgttca	ccggtgtctc	cggtcgggg	aagagttccc	tggtgttcga	39060
cacgatcgcc	gcggagtcac	agcggctgat	caacgagacc	tacagcgcat	tcgtgcaagg	39120
gttcatgccc	acgtgtctgc	gccccgacgt	cgacgtgctc	gatggcctca	ccacggcgat	39180
catcgtcagc	caggagcgga	tcggcaacaa	tcgcggttcc	accgtgggca	ccgtgacgga	39240
cgccaacgcg	atgctgcgca	tcctgttcag	ccggtcggc	gagccacgac	tggtctccc	39300
gcggtcgtac	tcgttcaaca	ttccctcgat	cagcgtcgcg	aacgcggtca	ccgtcggcca	39360
agggagccgg	gcggtcgagg	agcaacgcgg	cttcgacgtc	gccggcgga	tggtccccacg	39420
gtgcgaaggt	acggggctcg	tcaacgacgt	cgacctcacc	cagctctacg	acgccacgaa	39480
gtcgctcgac	gagggtgcca	ttcggtgcc	cggttaccag	gtcggcggct	acaaccaccg	39540
gctctacgcc	ggttcggggc	tgctcgaccc	cggaagccc	gtcgagaagt	acaccgagaa	39600
ggagctcaac	gacttctctc	accgggaacc	gacccggctg	aagatcgggc	actaccacac	39660
gaactacgaa	gggctggtcg	caagggtcca	gaagtgcgtg	ctcaccaagg	acatcgaggc	39720
gcttcagccg	cacgtacgcg	ctttcgtgga	gcgtgcggtc	acgttcacgg	cgtgcccaga	39780
ctgcggcggt	acccggctca	acgagcaggc	ccgggcatcc	aggatcaagg	gcatcaacat	39840
cgccgacgcc	tcgcgcgatg	agatcaacga	cctcgccgaa	tggttcggcg	acctggccga	39900
gccatccgtc	gcaccgctgc	tgccacgct	cctgcacacg	ctcgactcgt	tcgtgggtat	39960
cgggctgggc	tacctctccc	tcgaccgggc	ctccggcaca	ctctccggcg	gcgaggcaca	40020
gcgggtcaag	atgatccgcc	acctcgggtc	accgctcacc	gacgtcacct	acgtcttcga	40080
cgagccgacg	atcgggcttc	acccccacga	catccaacgg	atgaacaaac	tcctgctgcg	40140
gctccgcgac	aagggaacaa	cgggtgctcgt	cgtggagcac	aagccggaga	ccatcgccat	40200
cgccgaccgc	gtcgtcgatc	tcggtcccg	cgccggcacc	gcaggcgga	ccatctgctt	40260
cgagggcacc	gtcgcggaac	tgcttccgc	cgacacgctg	accggccgcc	atcttgacga	40320
caaggcgtcg	ctgaaggaca	gggtgcgtac	cccgtcgggc	tcgctcgagg	tcggggcgcc	40380
gagcaaacac	aacctgcgag	acgtcgacgt	cgacatcccg	ctcggcgtgc	tcgtcgtcgt	40440
gaccggcggtg	gcaggctccg	gcaagagctc	actgatcgaa	gggtcgggtat	cgggacgcga	40500
cggcgtggtc	acgatcgacc	aggggtcgat	caagggtccc	cgtcgaaagca	accccgccac	40560
ctacaccggc	ctcctcgatc	ccatccgcaa	ggcgttcgcc	aaagccaacg	gcgtgaagcc	40620
ggcgtgtttc	agctcgaatt	ccgagggggc	ctgtgccacg	tgcaagggtg	caggcgtcat	40680
ctacaccgac	ctggcgatca	tgccgggcgc	ggcgaccacc	tgtagcgaact	gcgagggcag	40740
gcggttccag	gacttcgtgc	tcagtagccg	tctcggcggt	cgcgacatca	gcgagggtgct	40800
ggccatgtcc	gtggccgagg	ccgggcgttt	cttcgctgcg	ggtcaggcac	gtacgcccgc	40860
cgccgctcgtc	atcctcgaac	ggctcgccga	cgtggggctt	ggctaccttt	ccctcgccca	40920
gccactcccc	acactctccg	gaggcgagcg	gcagcggtc	aagctcgccg	cgacacctggc	40980
tgacgaaggt	ggcatcgctg	tgctcgacga	gcgaccacc	ggcctgcacc	tcgcccagct	41040
ggagcagtta	ctcggctctg	tcgaccggct	cgttgacgct	ggccgggtcg	tcattggtcat	41100
cgagcatcac	caggcggtca	tgcccacgc	cgactggatc	atcgacctcg	gtccggccgc	41160
cggccacgac	ggcggctcgtg	tcgtcttcga	aggctgccc	gccgacctcg	tagccagccg	41220

cgccaccctc	accggcggaac	acctcgcggc	gtacgtaggg	gtctgacggc	ccgcctgaca	41280
cgggtttccg	ccgcctgccc	tggcgggccg	tcaccgctcg	tacaagcacc	cgctctgtcg	41340
gcctcggccc	acgcgtgcca	tggcatccca	tggcggtacc	cgaccgccc	ctcaggctcc	41400
atcctgccct	gaccaggccg	ttctccagca	cgcgcgatc	ggttcggggc	gatccgggta	41460
gcccggcaca	ggtgtttctg	ctgtccccgt	acctgaccga	acccagctct	tcgtggccct	41520
ggcgaagttc	tttcatcctt	gtggcatgcg	tgttgcgtag	cccatgggtc	gcgttgagta	41580
ttgggaatcg	taattggata	tctcgccctg	ggggggatgc	ggtgatggag	gcgactggcg	41640
ttgcggtg	gcgggtgcca	gcaccggccg	agccggcccg	tcgctgtcgg	ggtcgaggtc	41700
gtcatggaca	cacagccggg	gcgatgccgc	cacggatccg	tgcccatgac	gctctgacag	41760
attcgttgac	cgcgaccctg	cgccgatggt	gacgggagca	gggcctttcg	cgagatgttt	41820
tgggggattt	gcttgtgggc	ggtcggaaact	ccgatgtcgt	cttcatccgc	tggcccgcctc	41880
aggcggacct	gagaaccccg	tgccggaagg	agggagtgcc	gcggctgtct	gtcgtggagg	41940
gcggtgcca	gccgccgatc	agcaatgacc	cggtcgagga	ctgggtcccg	ccgccggtgt	42000
ccaggaacga	cctggaggca	cgggtcaagg	cactacaaca	acgcttcaac	agccgacggc	42060
tgccgaccat	cgactctacc	gggacgttga	ccttcggccg	tcactcgatc	gtcatctcca	42120
acaccagggc	ggagctgatg	cggctgctca	tcgaacgctt	cggcgaagtc	gtgtaccgca	42180
tcgagttcgt	gcagtgcctc	gccatccgga	cgcccgcgct	gaacagaaac	tcgctcgacc	42240
tgcacatcat	cgcgctgccc	cgcgggtccc	taccgatcga	cctggcgatc	cggaccgtgt	42300
ggggtcgggg	ctacatgttg	gcgtatgcgg	cctccgcca	ggccatcatc	agaaacgcca	42360
gcggcatgtg	atgatgcctg	ccgaaaacag	ctcaaactgc	cgtaccgtcg	agctgaaggt	42420
tgcggtgctg	ggctaccgcc	ggtccggccg	aagcagcgag	gagcgcgctc	agcgcgccga	42480
gccggccctc	gcggacgaac	cattcccggc	cgtaggctgc	ggcgaagcat	gaggggtcga	42540
ggtctgccgc	cccgaacctg	ctggcggtgc	tggccatcgc	ggccatcttc	gcgagcaa	42600
cggcaccttc	cgcgtcgaca	gcgagcgaga	tccgggacgc	cggcaccccc	atgtccgtca	42660
ggtcgccggc	agcacgtg	aacacgtgac	gtggccccgc	gcgcagagct	gagggatcca	42720
cgggtggcctg	gtatcccgcc	agctgaaggc	tgcgggtgac	cagagcgccc	acccggtgca	42780
cctgaacgtg	gtccacgtgc	cggtagatgc	cgccgggatc	atagtgaacc	agcgtggctg	42840
caccctcacg	gctggcgatc	cgggccacct	gtcgggcgac	cctcgtggcg	gacgcccgcc	42900
cgagggtgcc	ggcccggtag	gggccggcgt	gcgcaccgga	atcggcgtag	cccagcacct	42960
ccaatctggc	tacgccgagc	aggtcacagg	cacgttcgag	ttcggcgaga	cggcggtgct	43020
gcaaactctc	gccgcgtgac	aacggcacgt	gtggaacacc	ggcggcacca	tcggtggcca	43080
tgacgaggat	cacgcggggc	ccagcgctcg	ccaggcgacg	catggtcagc	ccggtgaaga	43140
ttgcctcgtc	gtccgggtga	gcgtgcagga	cgatgaccgt	gtgcgaggag	agcatgatgt	43200
gatccctttt	gatcgcgccg	gttgagcacg	gtcgcgacga	cagcgcgctc	gccacgcgct	43260
gcgacgcacc	ccgctcgaa	cgggaagccga	agaagccgac	gtgtggtgcc	cgcgcgtgtg	43320
gctgcgcgga	gaacggctag	gtcatagcta	ccgcacccca	ctcgcgtggt	ggtacgtcac	43380
ggagggtgtc	cgtgaactcc	cacaagtgac	cttcgagggt	acggaccagg	tactggcgca	43440
gtccccacgg	cttgtcctgc	aggggctgca	cgatgtcggc	acctgccgct	gtggagcgca	43500
tgaagtgtc	gtcgacgtcg	gagacccaga	cgatgacctg	cttgcaagg	tgcccgccgg	43560
cgcagctgac	cgtgtagggc	tcgccgggtac	gccggaccat	gacgatccca	ccgccggtgt	43620
ccagatcgcc	gtgtcggtg	gtcccgatgg	ccggctgccg	gagcagcagg	cggaaacccga	43680
agacccgtag	cagccagtcg	acagcacgat	cgggggtccg	gtaggcgagt	tcgacgacca	43740
cgtgagggaa	gccgcggggc	gggtcgtaca	cagagtctcc	ttcgttcggc	gagcagcttg	43800
ccgacgtcaa	cggttcgacg	tcaccgtttc	gggacagggt	cttcggatgt	cggaaagtct	43860
cgacatccct	gtcgaggcag	gctacccaag	ttttgagaca	ccaggggccc	cggcgaagca	43920
tctgcgaaag	aggtgccaac	cgtcccatcg	gacgcgcgcg	ctcgcggtgg	aggcgggtccc	43980
ggccgcagag	gactgcaaga	cggtcgggac	gtcacgggtg	gtgccggcgt	gcagttctga	44040
tcggcccgat	ctgagggcga	cggttgattg	gttggcagga	atcaagccgg	agagtgtcgg	44100
atctggcgga	tcagccgggt	ccgcgacccc	tcgcggaacc	cgctgttcca	ggtgggcctg	44160
caactgctgg	gcagcgccac	ccagagcacc	gagccgaacc	tgcccgggct	caccgtcacc	44220
tcgctggacg	ccaacccggg	cgggcacccg	ttcgacctgt	cgatcacctg	caccgagagc	44280
cccgaaccag	tgcggtggtg	cgcgcactac	tcggctcgacc	tgttcgacga	gccgcggatg	44340
gcgcggctgc	tgcggcacct	ggaacgggtg	ttgcggaccg	ccgtcgacgc	ccccgacacc	44400
ccgctgtcgg	cgtgcgcgtt	gctcgacgag	gaggagcaaa	cgtcgtgctg	gcacacctgg	44460
cagggaccgc	agggggagcg	cgcgcgcgaa	ccgggtgcacg	tgcatatcgc	gcggctggcc	44520
gccgaacgac	ccgacctggt	ggcgccacc	ctgcacgggtg	ccgagctgac	ctacggcgag	44580
ctggatcgcc	gcgcgggggt	gctggcgcac	cggttccggg	cgtcggcggt	gggcggggac	44640
gacatcgtgg	cgttgtgtgt	ggaacggggc	ttcgacctga	tcgtcgccat	ggtcgcccg	44700
cagaaggccg	gcggcgcggt	cgtgggtgatg	gaccggggcc	acccggtacg	ccgcacgcag	44760
ttcatcctcg	aagacaccgc	ggccaaggcg	gtgggtcacc	gcagcgcgct	ggccgaccgg	44820
ctgcccgagg	gtgcgcgcgc	cacgcgggtg	ctcgtcgaca	cggagtggga	cgagttgtcg	44880

gcggcggccg	ccggcgcggc	gccgctggac	gaactggccg	acgagaacte	gctggcctac	44940
gtgctctaca	cctcgggctc	gaccggcaaa	cccaagggcg	tgatgatcga	gcatacgcga	45000
ctgaacacgt	tcctgctctg	gctcggcaac	atcttcgact	tcggccccgg	cgaccgggtt	45060
ctgcaacaca	tggcgcccat	cttcgacttc	gccgagggcg	agatcttcac	cgcgctgacc	45120
cgtggcggtga	cgatggtctt	cgtaccggag	gaacaccgca	ccgaccggga	cgtgatcggg	45180
cagctgctgg	tctcggagag	gatcacctac	atcggcggac	cgccggccat	cctggggccg	45240
atcccgcggg	gcagctatcc	cgacctcaag	tacatgatcg	ccggaggtga	ggcggtcacg	45300
ggtgatctca	tcaaccgctg	gaacaccccg	ggtcgccggg	tcatcaacgg	gtacggaccg	45360
accgaggcgg	cgctcggctg	catcttctac	gagtgcgagc	accgcacctg	gatcggccaa	45420
ccgccgatcg	ggcggggccat	gcccacccgg	gtcgccctac	tgctggaccg	gtggaacaac	45480
ctgcaaccgg	tgggggtgcc	cggtgagggt	atcaccggcg	gggaggggtc	cgcccggggc	45540
tacctcaacc	ggcgggagtt	gaccgcccag	aagttcatcg	acaaccgcta	ccggccgggg	45600
gagcggatgt	accgcaccgg	cgacctgggc	gtgtggaccg	aggacgggca	gatccagttc	45660
ctcggccgca	tcgacaccca	ggtcaagctc	aacgggctgc	gcatacagct	ggaggagatc	45720
gagtcgacgc	tcaccgcgca	tcccggcgtc	gccgagggcg	cggtgccgct	gcgcgaggac	45780
agccccggca	gcaagcgcct	ggtcggctac	gtggtcccgg	ccggcgccgc	ggcaccaccg	45840
accgaggaac	tgccgcgagc	cctgctcgcc	gacctgccc	cgtacatggt	gccgcacgtc	45900
ttcgtggccc	tggacgtgct	accgctgacc	agcgtcggca	aggtggaccg	caaccggctc	45960
ccggttcccc	aggagaccgg	gtcggccgag	gtggagctgg	tcgcgcgcgc	caccccgagg	46020
gaggagcggg	tcgcccagat	cttcgcccgc	gtgctcgggt	tcgagcgggt	cggcgcgcac	46080
gcgaacttct	tcgagctggg	cggcaactcg	ctgatggcgg	cccgggtgct	gtcccggatc	46140
gccgaactcg	gcggggccag	cgcgacctg	cgggacttct	acaccaagcc	gcgggtcagc	46200
gagatcgcg	agctgctcgc	cggaaacggc	tcgagcaccg	accggcgcac	cccgaccgac	46260
gcggcaccgg	tggccgacca	gacccgggcc	cggttggagc	aggagatcgc	cgacctggaa	46320
cagcggctac	gggcccgcgc	ggcggccctc	gccgggcagg	acggtccggt	gtccggcggg	46380
cacgagggac	agcggctgcc	gctgtcaccg	acgcaggagc	agctgtggtt	cctcgaacag	46440
ctcgcaaccg	gccagccgac	ctaccacatc	ccggtgccgc	tacggctcac	cggaccgctc	46500
gacgtcgacc	tgctcggctg	ggcgtgacc	gccgtggtgg	cccgccacga	ggcgtgcgc	46560
accacgttcg	gcaccgacga	cgacaccccg	taccaactgg	tgtacccgc	cggtgagggt	46620
ccgctgccc	tgacagagct	ggacgcgacg	gcagccgacg	accgggagca	ggcgtgcgc	46680
gtcgagctgg	aacgcgacgc	caccgaaccg	ttcgacctga	gtcgcgacct	gatgctgcgc	46740
gcccggctga	tccgcctcga	cggccaggag	cacgtgctca	gcattggtgt	gcaccacatc	46800
tgccgcgacg	gctggtcggc	cgggctgctg	gtgcgtgata	tcgcccagct	gtacgacgcg	46860
ttgcggcacc	accggcagcc	gcagctgccc	gccctgcggg	tcaggtacgc	cgaccacgcg	46920
ctggcgcagc	gccggcagct	cgacgcccgt	gacctggagt	cgagttggc	gtactgggcg	46980
cggcagctgg	ccgacgcgcc	catcctggac	ctgcccagcg	accggccccg	cccggccacg	47040
gccaccaacc	gtggcgccct	gctggtgctg	cggatgccc	ccgacgtgct	ggagggcctg	47100
cgccgggtgg	cccagcagga	gcaggccagc	acgtacgcgc	tgctgctcgc	cggcttcaag	47160
acggtgctga	tgcgctacac	cggcgctcgc	gacgtggtgg	tcggcaccgc	cggcgccggt	47220
cgcggccgct	cggagctgga	ggatctggtc	ggtttcttcg	tcaacatgct	ggccctgcgt	47280
accgacctct	ccggcgaccc	gagcttcggg	gaggcggtcc	gccgcaacct	gggcacctg	47340
ctggccgcct	gggaccacca	ggacgcgcgc	ttcgagcggg	tcgtcggggc	gctcggctga	47400
cgccgcgacg	ccagccgcaa	cccggctctc	caggctctca	tcgatctgct	cagcggcgac	47460
ttgggtgatt	tccggttccc	cgggctggcc	gccgagttcc	tcgacgtcga	tcccgggggt	47520
tcccgccttc	acatggcgat	caacgcgtac	gaggagaccg	acggcatcac	cttcgggatc	47580
gagtacgcga	cggacctgtt	cgacaccgac	cggatcgagg	cgatgttcgg	ccacctggag	47640
cgggttctgc	gggcccgcgt	ggcggcaccg	cagacgcgcg	tgctcgccct	gccactgctc	47700
gacgacaccg	agcgaaccga	actgctggcg	ctggcccagg	ggaagcgacg	gccccaccgg	47760
ccggcgaccc	tgccggagct	gttcgcggcg	agcgtggccc	gcaccccgca	ggccaccgcg	47820
gtcatcgccg	ccgagggcac	gctgagctac	gccgaactcg	acgagcgacg	ccgtcggctc	47880
gccgcgcggc	tgccgcgcgc	cggagtgggc	accggggacg	tcgtgcgggt	gctgctggaa	47940
cgcagcctcg	cggaggtggt	gagcgtgctg	gccgtgctgc	gggcccgtgc	cgcgtacgcg	48000
ccgatggacc	cggccgcgcc	gccgagccgg	ctggccttcc	aggtgcgcga	caccggcgcc	48060
cgggtggtgg	tgggcgaccc	gacgcggctc	gacacggtgg	tcgggggtgac	catcgtggac	48120
ccggcgggcg	agctgcccga	ggcggccgac	gagccgtggc	tgcccgagct	gacccccgct	48180
tcgctggcgt	acgtgctgta	cacctccggt	tccaccggcg	agcccaaggg	cgtgctcatc	48240
gagcacggca	tggtcgccgg	ctacctggag	tggtggtgct	cggacttcgg	cgtcgagcca	48300
cgtccccgga	tgatgcactg	ctgcgcgcgc	gtgttcgacc	tgcccgctcg	tgaactgttc	48360
ggtgccctca	ccgcgggtgc	gacctgggtg	gtggccgggt	acgacgacgt	gcaccgcccc	48420
ggcgcggtga	ccgcgctggc	ccgcgagacc	ggcgcgacgc	acatcttcac	cactcccacc	48480
gtgctgggtc	tgctcgatcc	ggccgaactg	ccgcaactgc	ggtcgggtgat	gatggccggc	48540

gagccggcgc	cgccgtcgct	ggtggccgc	tggcacgtcg	acggtcggcg	ggtgctcaac	48600
ctgtacggcc	cgcccgagac	cacggtgggc	tgcacctggt	acgagtggcc	ggccggcacc	48660
gcacccgcgg	cgacggtacc	gatcggtcgt	ccgatgccga	accggcgggt	gtacgtcgtc	48720
gacgacagcg	gtgacctggc	gccggtcggt	gtgcccgggt	agctggtcgt	cgccggcaac	48780
gggggtggcg	gtggctatct	cgaccgaccc	gagctgacgg	cctcggcgtt	cggcggcgac	48840
gagcggggcg	gggaggggacg	ggcgtagccg	accggtgacc	tggcccgggt	gaaccgggac	48900
ggccagttgg	agttcctcgg	ccgcccgcgac	gggcaggtea	agctgcgtgg	gctgcgggtc	48960
gagctggggc	aggctcgagca	tgccttgcgc	acggcgggcg	gcgtgtccgg	ggccgcggcg	49020
gcggtgcacg	gcacggggcg	cgaggcggtc	ctggtcgggt	acgtggtcgg	cgagtgtgtc	49080
gacgacgacg	cggaggcgct	gcggacgacg	ctcgccggga	agctgcccgg	gtacatgggt	49140
cccaacgtgg	tgctgcgcct	gccgcagctg	ccgctgaccc	gtaccggcaa	ggtggaccgc	49200
aagcggttgc	ccgcgcccga	gtccgggtgc	gggcaggcgg	cgtaacgtgc	gcccggcacc	49260
gacgacgagc	ggctcatcgc	cggggtcttc	gccgacctgc	tcggctctgga	acgggcccgc	49320
cgcgaggacg	agttcttcct	cagcggcggc	aactcgctgc	aggcgatccg	ggcggtgggc	49380
cggtcaacg	acgcactcgg	tgtgccgctg	accgtgcggg	acttctacgc	ggcaccgacc	49440
gtcgcgggcc	tggccgagct	ggccggacgc	cgacggggcg	ccgagcaggc	ggcgagtggt	49500
caactgctcg	acgagatcga	gcagctcagc	gacgacgagg	tggaaacggct	gctggcgggc	49560
gactgatcgc	cggctcgctc	cgcgggcccc	gccgcccgtc	gcggcggggc	cgcggtgtgc	49620
ctgaaaggag	tgagatgtcc	gacgagacct	tcgacggcag	ggtagcccga	ctggacgagc	49680
gccggcggtc	gctgctggaa	cggatgcggc	aacagcgccg	gcagccaccg	gcgacgatcg	49740
gcgcccgccc	ggccggcgac	gagcgggtgc	cgctgtcgcg	ggcccaggaa	cagttgtggt	49800
tcctcgcccc	gctcgccggc	gccgagccga	cgtacaacct	ggtgcaggcc	tcgtacctga	49860
ccggcgaggt	ggaccccggc	gcgctgcggc	ggcggttgga	cggcgtggtc	gcccagcacg	49920
aggcgttgcg	cacggtcgtg	gagagcgtcg	acggcggtgc	gtaccagcgg	gtgctgccgc	49980
cggggcccgc	cagcctgggtc	gtcgacgacg	tgtcccggct	gcccaggag	cagcgccggc	50040
cggcggcgct	caccctgctc	cagcagcagg	aggtgcaccg	cccgttcgac	ctggccaccg	50100
gcccgttggt	ccggggcccgc	ctggtcccgc	tcggctccac	cgagcacgct	ctcgccgtcg	50160
cggcgcacca	catcggtggtc	gacggctggt	cgatgggcct	gatcctgcgg	gagctggccg	50220
ccgcgtacga	cgccgaccgg	cacggcacgc	gagcgcggct	ggacccgccc	gccctccagt	50280
accgcgactt	cgcccgcgtg	cagcatcagc	agctggccac	cgccgaactg	gacaaccacc	50340
tgaggtagctg	ggcgacgagg	ctggccgacc	tgcccaccct	ggaactgccg	gccgaccgac	50400
cccggcccgc	caccgcgtcg	ttccgggggtg	cgctgctgga	gcaaccactg	gatccgcagc	50460
tgcacgcccg	ggtgcaggcg	ttggccaagg	agaccggcgc	gagcccgttc	atggtgctgg	50520
tggcggcctt	cgcggccatg	ctcgcccgtc	acaccggcac	cgacgacctg	gccgtcggca	50580
ccaccttcag	cgggcgagct	cgctccggagc	tggagagtgt	ggtcgggttc	ttcgccaaca	50640
tgggggtgct	gcgtctggac	acctcgggcg	accccagttt	cgccaccctg	gtcgaccggg	50700
cccgggacac	ctgcctgggc	gcgtgggagc	accaggacgc	cccgttcgag	cggttggtgc	50760
agcgactcgc	cccaccccgc	gacccgggcg	gcaacccgct	gttccagggtg	gccgtgcagc	50820
tgctggacgg	ctccacctgg	ggcggaacga	cgctgcccg	caccgacagc	accccgtcg	50880
acctgcggct	ggacccggtcc	cgcttcgaca	tgatggtcag	cctcgtcgac	cacggcgacc	50940
ggtactcggt	gctcaccgag	tactccaccg	acctgttcga	cgcccggcg	gtacgcgcca	51000
tgctcgacca	cgctcgcccgc	ctgctcgacg	cgggcctggc	cgacccggct	acgacgtgtg	51060
cgcggttgcc	gatgctcgac	cccgcgaac	gcgagcgggt	cctggccctc	ggcgcggttg	51120
aacgacgtga	cgtggcccct	gccacgcctc	accagctggt	ccgggcgcag	gccgaacgca	51180
accccgacgc	ggtggccgct	cggaacggcg	atcacgagct	gacctaccgg	gagctgctgg	51240
accgggcgcg	gcggcgggcg	gccggcctgc	gogaocgggg	cctgcgaccc	ggcgatcggg	51300
tgcggcgct	gctggaccgg	ggcgtcgacg	aggtgctgac	gccgatggcg	atctggcacg	51360
ccggtgggg	gtacgtgccg	ctggacaccg	ccgcccgcgc	gaaccggctg	cgccggatcc	51420
tgaccaactg	tgcgcgccgg	ctcgctgctca	ccgcacccga	gcacgcgcgc	ggggtgccgg	51480
aggggccgtg	gcggaccctc	ctgctcgaca	ccggcgacct	ggactccggc	gcggggtcgc	51540
ccccggaacc	gtccaccggc	ccggacgacc	tggcgtagct	gctgcacacc	tcgggcacga	51600
cgggtgaccc	gaagggtgtc	cagctgcgcg	accgggggtt	cgtaacctac	ctggactgga	51660
tgcgcgacca	gtggcgctgc	gggcccgggtg	accgggtgtt	gcacggctgc	gcgcgggtgt	51720
tcgacctggc	ggcggtgag	gtgctggcg	cgctcacctc	cggcgccacg	gtggtggtga	51780
tcggcaagga	ccggctgctc	tcgcggggcg	gtctgcgcga	ggtcctggcg	gccgaacgga	51840
tcacgcacct	gttcttgacc	ccgaccacgc	tcggtctggt	cgaggacgac	accgaccggt	51900
tcgacgcgtt	gcgggaggtg	ctggtcggtg	gcgaggtctg	ccccgccgac	ctggtggccc	51960
ggtgggcgcg	ccggggggcg	cgctcggttcg	tcaacctgta	cgccccgacc	gagggcaccg	52020
tgggtgcct	cgcccacgac	tgacggact	ggtcggcggc	cacccacccg	ccgatcggcg	52080
cgacgatgcc	gaaccgcgcg	acctacctgg	tcgaccgctg	ggacaacccg	gtcccggctg	52140
gggtgccggg	ggagatcctg	gtcggtggcg	cggcgctgag	caccggctac	ctcaacgacg	52200

acgagctgac	cggacgcccg	ttcaccgccc	atccgttcga	tccgcagggg	cgggtctacc	52260
gcaccggcga	tcgtggtcac	tgggacgcgc	agggccggct	gcacttcggc	ggccggctgg	52320
acggccaggt	caagttgcgc	ggcctacgcg	togaactcga	cgagatcgag	gcactgctga	52380
gccggcactc	cggggtggcc	cgccgcgcgc	tgaccgtggt	ccgcgacggc	aacgggggtgc	52440
agcggctcgt	cgggtacgtg	gtggacgcgc	ccgagccggc	cagcgacgcg	gacctgcgca	52500
cgcacctgag	cacggagctg	cccgcaccac	tgggtcccgc	ccggttcgtc	cgactgcccg	52560
ccctgccgtt	gaccaccagc	ggcaagggtca	accggcgggc	gctgccggca	ccggaggcgg	52620
cggacaccga	cgacaccggc	gcggccccc	ccaccggggc	cgagcgccag	gtcgcggaga	52680
tcttcgcgga	ggtgctgggg	gtggcgccgg	tgggcgcgga	gacgtccctc	ttcgacctcg	52740
gcgggcactc	ggtgcaggcg	gcgcacgtgc	tggcccgcat	cgccggacgg	atcggcacgt	52800
cgggtgctgc	caagcagttc	tacgccaccc	cgaccgtccg	ggcgggtggc	cggttcgtcg	52860
agcacggggc	ggtggtggac	gcctcgccac	tcgtcacgct	caaggccacc	ggttcgcggc	52920
cccggctcta	ctgcccgcac	gcgggtctccg	gctcgccgta	ctggtacctc	ccgctgagcc	52980
gggcccctga	tccggagcag	ccggtggacg	gcttcgaggc	gccggggctg	gagggtggca	53040
tcgagccggg	cgacgacctg	gtgacgctcg	ccggcgcgta	cgtcggcgcg	ctgcgcgaac	53100
ggcagccgca	cgggcccgtac	ctgctcgccg	gctggtcgat	ggggggcttc	ctctgcttcg	53160
agatggccc	acagctcgcc	gccgtgggtg	ccgcaccgga	cctcgtggtg	atgatcgact	53220
ccaatgaacc	cggcccgttg	ccgctgccga	gcgagcgaca	ggtgatggag	accttcgtcg	53280
ccgacctggc	cgggctctcc	gcggcacccg	caccgcgcgt	ggacgcccag	gtggtcgacg	53340
cggccgaccc	gatcgcgcg	ctcacccggg	tcctggctccg	gcacggactg	atcccgctcg	53400
acgtgcgggc	cgacttcgtc	gccaacccgt	acgcggtggt	ccgcgccaac	atgcgggcca	53460
tctacggcta	ccaaccggca	ccgtacgcgc	gccggacggg	gctggtgcag	gcccagcggg	53520
aggccagccg	cgccggcgtg	cgccgggtga	gtccggcgat	ggaggtggtg	acgctgcccg	53580
gcgaccacta	ctcgatgtgg	tccccgcagc	acctgccgcg	gctggcagcc	ctcatcgacg	53640
accagggtgg	acgggtgctc	gacctcggtc	gaccggcacc	ggggcccgcg	cagaccatcg	53700
gctgggcggg	ccccggacga	gcacgtctac	cgccggcgag	cctggtcacc	tcggcgcgga	53760
acgcgctccg	gtcggcgcg	acgactcctc	gaccgggccc	gttgccggct	cgacgtcgac	53820
ctcaccgggg	ggtgcgttgc	ccggcgccgc	cacggcctcg	acctgctcgg	cggcgtcgcc	53880
gctcggcgcc	gtctcagccg	acacccgggt	ggacagccgc	cagggtccgg	aggccagcag	53940
gaatcccagc	atcagcagac	cgccgtcgac	ccagaggctg	cggtccaacc	ccagcgcgga	54000
acggctccgg	tcgttctcgg	cggtcgccag	gtagacgcgc	ccgacgacga	ccacgctcac	54060
cacgctggcc	agcagcgttc	cggtcgagac	cagggccgac	agggcgcgcg	cgtagcgggg	54120
gcccacggtc	gcggccgtct	gtccacagag	cgtgctgaaa	ccggcccgca	tccccgcacc	54180
ggcgagcagc	agcagcgccg	acaccaccgc	gatcggccag	ccgcccggca	ccacgaccgc	54240
gagtgccggc	agggccaggg	cgaacaccag	caggccggcg	accggcagcg	ggcgggtgaa	54300
acgttttcgg	agccgtgtcc	aggtgaggct	gacggccgcg	aaacccacgg	ggtacgggat	54360
gaacatcaac	ccgcctgca	acgggtgtgt	gcccagcccg	ttctgcacgt	gcagggtgag	54420
cggaaacagc	accccggcga	aggcgaagtt	gaggatgcag	caggccagca	gtcccgaact	54480
gacgcctcgc	ggtcgcagcg	catcgatctc	cagcaccggc	tgtccgcccc	gggcgcccag	54540
ccgcagctcg	taccggacga	agaggacgag	cccggccgtg	ccggcgggcg	tgctggccag	54600
tgcccacgcg	ggccagccct	gttcccgcgc	gaagaccagc	ggtgccacca	ccgcgcccac	54660
cgagacggcg	agcagcagga	tgccgacgat	gtcgggcccg	aaccgcgagc	tgggcgtgtg	54720
ccccggcaga	taacgcggcg	ccgcgcggag	cagggccagc	ccgatcggca	cgttgaccag	54780
gaacgcggcc	ctccagccgc	tgcccagcac	gtcggcgctg	acgatcacgc	cgccggcgat	54840
ctgaccggcg	gccactccga	gcgccagcac	catcgagtag	agcccgatgg	cgccggcccg	54900
ggccgacccg	gtgaagacga	tctggatcac	cgagagcacc	tgccggacca	tcattggccc	54960
gaacgcccc	tgacgacggc	ggcgcgcgac	cagcgtgccc	gcggtcgggg	cgatcccgcg	55020
gagcagcgag	ctggcggtga	agccggtcag	cccgatcatg	aacagccgcc	ggtgcccac	55080
gtcgcgcggc	agtcgcgcgc	cggtgaccac	gagcaccgcg	aacatgagcg	tgtatccgga	55140
caggatcagt	tgacgaacgg	cgccggaggt	ctggaggtca	tccttgatgg	tcttggtggc	55200
cacggcgacg	atggagtgtg	ccatcgctcg	catgacctgt	gccaccagga	ggatgagcag	55260
caccagccgg	gcgcggcggt	gtcgaccttc	gtcctcagtg	gccatcgaca	ctctcgctc	55320
ttctctggtg	tcacggccgt	tgccgtggta	gaggcttcgc	tgccgacagt	ggttgcgcaa	55380
ctacgtgtcg	ctgggtgtcg	cgggcggtcg	ggcatcccgc	ggatcgctcg	cgtcgagac	55440
cgccggtcgg	gtcgcgggca	caacgttgcc	ggtcgccgac	cacgccaccg	acgcgggct	55500
ttcgggtgtg	gtggggccgg	tgccgtggcg	gggaccggtc	cgggacaccg	gcacagtgcc	55560
gacgccgcga	cgcgcgagcg	gcccagaccga	ggcaccacac	tccggcccga	acaccagcag	55620
caggacgggt	gcggccaggc	acagcacacc	cagtgcgatg	gccggcgcg	cgagcccggc	55680
gccgagggag	agaatcagca	gcatggcgga	cggtcggagg	atgccaccga	ccttcgtggc	55740
cgccccggcg	aagccggagc	cgggtggccc	catcgcggtc	gggtagatct	cggcgctgta	55800
cggcagcagc	atcgcggtga	ccgcggacaa	cgacagcacc	agcaggccga	tggaggccag	55860

cagcagcacc	gtcccgccgc	tgtegggtgt	ccagagggcc	actccgccga	gcccagaccgc	55920
ggttatcgcg	atgaacacgc	ccagcgcgcg	gcgcccgctc	caggcccgct	agagcagcac	55980
gaccagcgcg	agggccgggg	tgagtagag	cgccgagtcg	gccagcagcc	cgctcgcgac	56040
accgcccggc	agggccgcct	cgcgagctg	cgccggcagc	atcaccagga	agccgaagtt	56100
caccaggccc	cagacgacac	cgaccacgcc	gatcaccgat	gtcacgaccg	gatacgcccg	56160
cagcatcgcc	cccacgcccg	gacgcgtcgc	ggcgggcgca	gtgtcgcgcg	ccggggggcg	56220
ggcaccacag	tgggacccgt	agcgggcgag	caccgcctcc	gcctcggtac	gccgcccggc	56280
gcgcagcagg	tacagcgcg	actcggggat	caacggcgcc	atcagcagca	gcagcagtc	56340
ggagggcagc	ccgatcaacc	agagtgcgcg	ccaggtcagc	gccggctcca	ggtgagctgc	56400
cgctcgctg	gccgccaggt	aaccgcccag	gccaccgatc	ccaccgacgg	tcaccgccac	56460
ccagcccgcg	tgccggcggt	gggtcagctc	ggcgatcagc	gtgaacacca	ggggcaacag	56520
gccaccggcg	gaggtgcccc	tcgtgaagca	catgacgagg	ttccactcga	aggtgggcat	56580
ggccccgcac	gccgaggtcg	cgatgaacag	caggggtggc	aggagcaggg	cggtccgctg	56640
accgtaccgg	tcgcccagca	ggccccagac	gaccgacccg	accgcggtgc	cggtgagggc	56700
caccaggggc	aggatcgaga	ccgattcgag	ggacatgccg	tactcgctgc	tcagtcgggg	56760
catgacgaag	ccgaggggtca	gcggcttcac	gatgtcgacc	accagggcga	tggtcagcac	56820
cgcgacgggt	atccagtagc	cgggctcgag	ccgcccgtgc	tcgatcgact	ccacggtgac	56880
gctcgcgagg	tcgggtcggc	ggacccccgc	gaacagcgcc	cgcccggcca	ccaccaggcc	56940
gacgacgagc	agcagcatgc	cgaggatcat	cgtcggcgct	cacatgtcca	tcggctcgcc	57000
ggggctcatg	ccctccatgc	ccggcatgct	gtccatgtcg	tcctatgtcg	ccatgccgtc	57060
catgccgggc	tcggccgacc	ccatccgggc	atggctcgatc	atcatcttca	cgtcgggcag	57120
ttggaacagc	acgcccggcg	tcaccaacgc	cagcccgagc	aggaatccga	gtgtcccgcg	57180
gcgtttcgcg	gatgctgtgt	cggcagtgcg	ggattcgagc	atgggtgacc	gacctccgag	57240
ctgtcgcgag	acgagaagaa	cgtcggcgaa	ttggcgacca	ccgaaatcta	caagggtgct	57300
tcccgggtga	caaccgtcaa	ttatatgact	catttcggta	agtgggttatg	aattgctgcc	57360
tcgggggtggc	tcacctgttt	tgctaccgat	ttgggtgggt	ctgatcggtc	agtcagtcga	57420
cgagcggttc	ccgatcgccc	cgtagggatgt	taggttcacc	ctaagtgatc	aactggcggt	57480
cgccgggtgct	gtccaatggt	ccgcttacca	cagcttttcg	gcaggtcagt	gccgttatgg	57540
cggatcggtg	gtcgcggggtg	acgcgcaccg	cgctcgtcgga	tcgcccgggc	gggtggcgctg	57600
tcgcgccag	gccggcacgg	cgagcggtc	gacccgactt	cagcaattaa	ccgagcggtat	57660
tcttcagtga	attgctggca	atctctccgt	aacgctccgg	gccgcacgga	aatgctaggg	57720
tcgggagcga	tgcccgcgcg	ccatcgccctt	tcgactgagg	tcgacggcgg	gctggatcgg	57780
gcgtgagctg	cgccgtttct	cgaatcgcta	tcgaccatc	cgtagcgac	cgcgacccc	57840
gcacgcgcg	accggcccgg	atcgatcccc	tcgggtcactg	tcgggtgccg	agatgtgccc	57900
catgtggacc	ggtcgggccc	atccggccc	tggtgtgcca	gcattggacta	catgaccatc	57960
cagtcgcacc	cgatggcgct	gacggacgag	gtacgcgggtg	agccggccga	cgagttcgct	58020
ggcgccctcg	tctcgctgag	ccgggtcatg	gtggccgctc	ccatgcggac	gctggcgcg	58080
atcgacgcg	acgtgacgct	ccagcagttc	cgagcgctcg	tggtgctggt	gcccggccc	58140
cagcggtga	tcgacctggc	gaagctgctc	ggcgctcagt	cgtagaccgc	caccgggatg	58200
tgcgaccgac	tggtccgcaa	ggggctggtg	gcacggcacg	agcggtggc	ggaccgtcg	58260
gctgcctggg	tcgggctgac	cgtagccggc	aaggaaactcg	tcggcctggg	catgcgagcc	58320
cgacaacagg	aactggagcg	gatgctggcc	gacctgcgac	ccgacgacac	cgagctcacc	58380
accggcgcg	tggtggcggt	actgcgcgcc	accggtgagc	tcaccgacgg	tgccctgggtg	58440
cggcagtggt	cggtatcggt	ggtaggagtg	cgcaacgcgc	tcgggtcccga	gtggccggcg	58500
cagccggggc	cgcccgctcg	cgccgctggg	gtcagacggt	gacgcggctg	gccgcacggc	58560
ggcgggccat	gggcatcgac	ggcacctcgg	cgcccgagcg	cggtaggtcg	gacggtgtca	58620
cgtagaccgg	cgacacgcgc	gccgagacgg	cgctcctgcga	cgcccggttc	cggagttgat	58680
cgagggcggc	gtcgaggtcg	tcgaagatcg	ggaaaacccc	gtcgagcttc	atcgtgtgca	58740
gcacggtgcg	gatgaacctc	gacggcgcg	cgaggcacag	ccgcccgcga	cgctcgcgta	58800
cgtagcgatg	cgcccgagcg	agcatgccc	gcccagagcga	gtcgatgacc	tgcacctggg	58860
tcaggtccac	caccacccgtc	ccggccacgt	cgaacgcgc	gcgcagcgcg	gcccgcagcg	58920
cgtagccgggt	gtcctggtct	ccgcccgtcg	cgggcaccac	caccgcccgc	tcggcgaggc	58980
ccgatccagc	ccggctcgcc	tcaccgagg	agctgcccgg	cggcagcgcg	ctgcaccggg	59040
ggcagcggtg	cgggccgggtg	gcgaacggcg	accgctcca	cccggtggcg	gagaccagcg	59100
tcagacgac	ctcgcgctcg	ggcagcacgc	aggcggtgcc	ggtaggtggtg	tcaccacaga	59160
cgtagcgagat	cagggtcatc	aggttgtcgt	ccggtacgac	ggtagctcctg	cttccctctcc	59220
ggtgcccgcg	ccggctcgcc	gggtgcggg	actcgggggc	ggggcggggtg	aggcacccctc	59280
gccgtacgga	ctcctcacag	ggagtccgag	cgttccttgc	cggtgggtggc	gaccacgatac	59340
accaccgcga	ccgcccgcag	cccggcctgc	acgaccaggc	ccaggatgct	gaacccctga	59400
acgcccggcg	gccaggcggt	gatggcaccg	agcaacgagg	cgacggcacc	gatgggtcatg	59460
gtgagccaca	gcggcgcgag	cggacggccg	ggcaggacca	gtcgggcgag	cgccggcgag	59520

gcgaggccga	cgccgagggc	cgtgacgac	ccggcgctgg	tcatgctgct	cctcgtgctc	59580
gcggtcagat	gcgtcggctg	cgggtgtgtca	ggtgcgacag	gtgacggcgg	cgccggccgta	59640
cgccggcagcg	cgccgggaccg	tacccgggtgc	cgggtgtcccg	tcggccggcg	ggacaccggg	59700
tgcggcggtg	accgcctcgg	tgaccggggc	cgaggcgggcg	tcgggtgtgca	ccggcaggat	59760
ctggtccagc	cgggcggtct	gcaggatgcg	ggtgatgcgc	gggttcgggt	tgccggacggc	59820
cagcacgccc	ccggcacgga	ccatccggcg	gtgcacgtcg	agcagtagac	cgatggcggc	59880
ggcgtcgatg	tgccggcaac	ccgccaggtc	gatcacgacc	tgaccgggac	gcaggggccag	59940
gagctgatcg	aggaccgcgc	cgggtctccg	caggcagctc	agatcgaact	cgggtgatgga	60000
cacctccacc	agccggcaccg	aacactcggg	gcgtcgtggg	gcggtcacgg	ccaaggcccc	60060
ctctcctgcg	gtctccgagt	ggctctcacc	ctgccacagcg	gggttggcgg	ccaccgcgcg	60120
gacggatgac	agtttgcgtga	caaaaccccc	ggttgtcgat	cgtggttggc	ccggcggtca	60180
gggcttgggg	atgattccgg	tatgacagcg	gtactggtga	tcgaggacga	cgaccggatc	60240
cggttggcgt	tgtctgctgc	tctggaggac	gagggttacg	aggcccgggg	agccggccacc	60300
gccgaggagg	ggctgcgtac	gcaacgcggg	gatccgcggc	actacgtcct	ggtcgacctg	60360
atgctgcccg	gcacgcagcg	cttcgagggc	atccgccagt	tgccgcgcga	cgacgacgtg	60420
ccgatcgctg	tggtcagcgc	ccgcgacgac	actcacgaca	tcgtcgccgc	cctcgaggcc	60480
ggcgccgacg	actacgtggt	caagccgggtg	gcgatcaagg	aactgaccgc	tcggctgcgc	60540
gccctgcgcc	gacgcggggcg	caccacgacc	gcgtcggacg	gcccgcagcc	ggtgcgggtg	60600
ctcgccttcg	gtgaccttga	ggtgagcccc	gaggcggggg	aggtgcgccg	cgccgggtgag	60660
ccggtcgcgg	tcacccgcac	cgagttccgg	ctgctgtgcg	agctggccga	acacgccggc	60720
cgggtgctct	cccggcagca	gctgctcagc	cgggtctggg	ggtacgacag	cgccgacgaa	60780
cggcttgctg	acgtgcacgt	cggtcggctg	cgccagaaga	tcgaagccga	tcgggcaaac	60840
ccccgacatt	tggtgacgct	gcgcgggctg	ggttacaaac	tgcaacgatg	agacgcctcg	60900
gactgcgtgc	ccgggtcacc	gccgcgttcg	cggtcggtgc	gltgctgctg	tcgcctcga	60960
tggccctggg	gtcgtacgag	ctgaccgcc	gcaccctgct	cgacgagcgg	gaacgcaccg	61020
tgctgcgggg	ggcgtaactac	gacgcggcgg	tggtccgcgc	cggtctgcac	accgacaacc	61080
cggaactcgc	ggaggcggtg	cgggcglttg	acaccggcac	cggaacggcg	ccctggtgct	61140
tgctcgacgg	cgagtgggtac	gcccgcagcg	ccgacaacgg	gctcaccacg	atcgttccgg	61200
ccgggcttca	ggacctgggtc	gcggcggggg	agccggcggt	gcagcggggtg	cgcgtaacg	61260
acacccccgt	gatgctgggtc	ggtgtgcgcg	tgtccgagtc	gaccgccttc	tacgagatcg	61320
cctcgtgctg	ggaactggaa	cagaccttcc	aggtgctggc	gctggcgctc	accgcggctg	61380
cgatcatgat	cgccgggtcc	ggggcgggcc	tcggctggta	cgccaccggg	cacgggctgc	61440
gtccgctgac	cgccgtcgcc	gacgcgcggc	agccgatcgc	ggccggcgac	ttcaccaccc	61500
gcctcgcccc	cgacaccgac	ccggacctga	cccgctctgc	cacctccttc	aaccggatgg	61560
tcgacgagtt	ggccggggcg	atcgaccggg	accggcggtt	cgccgcggac	gtcagtcacg	61620
aactgcggtc	gccgttgacg	accctcgccg	ccgcgcgag	tgtgctgtcc	cgccgcgggg	61680
agaaccagga	cgagcggacc	gccaccgcgg	ccgggtcgtg	cgccgacgag	atcgaccgtt	61740
tccagcaact	ggtgaacgac	ctgctggacc	tggcccggtg	tgaccagccg	gtcgaccggg	61800
agtcgggtga	cctgcccag	ctggcccggc	agggctgccg	ggaccgggac	ctgccggagt	61860
cgatgggtga	gltggcaacg	gggacaccgc	ccgggtggcg	ggtggaccgt	cgccgggtcg	61920
cccaggtgtt	ggccaacctg	ctggacaacg	ccgagcggtg	cggtgcggga	ccggctcgcg	61980
tgccggtgtc	ccacgacgcg	gacacggggc	tgatcgaggt	cgacgacgag	ggcccgggag	62040
tgccgatgga	ggaccgcgag	gcgatcttcg	accgtttcgt	ccggggggcg	gcgcccaact	62100
accgtggtgg	cgccgacggc	accggcctgg	gcctggccct	ggtggcccg	cacgcgcggc	62160
cgccggcgcg	agacgcgtcg	gtcagcgacc	gcccggaggg	cggtgcccgg	ttccgggtca	62220
ccctgcgggg	gagcctgtca	tgaccgcgcc	cgccaccggc	cgggccgcgc	ggacggtcgg	62280
ggtcgcggcg	ctcgtcaccg	tggttcgccct	gctcggcgcg	tgccgggtgc	ccgccgagga	62340
tcgaccccg	gcggtgacgc	cgccgcgggt	gccgttcccg	gcgaccgctt	cgccgcgcgc	62400
gacggcgggc	gagaccggcg	cggtgaccga	ggtgctctac	ttcgcccgcg	acgaacgctt	62460
ggtgcccggg	atccgcgggg	ccgaccaggt	gcccgcactc	gacgcgcagc	tgccggcgct	62520
gctcgcgggg	ccgaccccg	ccgagcgtga	cgacgggctg	accagtgcct	tgccggggagc	62580
cttcaccagc	gcggtgggtc	agctgggtga	cgggttggcc	cggtcaccg	tcgggtcgac	62640
ggcggtggac	accggccgca	gcgacggccg	gctggcggtac	ggtcagatcg	tctgcagcgt	62700
caccgcccgc	accgatgtca	cgccgggtgct	gttcctggaa	ggcggcacgc	cgtgacgcgt	62760
ccccggggcg	gacggctcgc	tgctcgtcggg	accgctcacc	gctgccgact	acgcggtgtt	62820
gatcaatcct	cggtgacgag	ggccggacca	ccgcacgac	gcgcagcgcg	acgactcgta	62880
tcttttccgc	gaaccgtcgt	gccgacgacc	ggaaaggact	tggtgtgaac	gacggggccg	62940
agacgttgga	gttcaggcc	gaagcgcgtc	agctgctccg	gctgatggtc	caactcgatct	63000
actcgaacaa	ggacgtcttc	ctccgcgagt	tgatctcgaa	cgccctcagac	gcgctggaca	63060
agctgcgcct	ggcgtcgatg	gtcgacaagg	acctgcgcgt	cgacaccggc	gacctgcaca	63120
tcgccatcga	ggtcgaccgg	gacgcccgc	ccctcaccgt	gcgggacaac	ggcatcgga	63180

```

tgagccgcga cgaggtggtc tccgtcatcg gcaccatcgc caagtccggc accgcggagc 63240
tgctgcggca gctgcgggag gccaaggacg ccggcgccctc acaggagctg atcggtcagt 63300
tcggcgctcg cttctacgcc gcgttcatgg tcgccgagcg ggtcgagctg gtcacccgac 63360
gggcccggca gaccggtggt acccgctggg agtccaccgg cgagggcacc tacaccatca 63420
ccgagctgga cgacgcccc cagggcacgg cggtgaccct gcacctcaag cccgcccagc 63480
ccgaggacaa cctgcacgac tacaccgccc aatggacgat ccggggggatc gtcaagcggc 63540
actccgactt catcgccac cogatccgga tgaccgtcga gcggcccggc accgacgacg 63600
cccccgccac caccgaggcg gtgacgctca actcgatgaa ggcgctgtgg gcccgcccgc 63660
gcggtgaggt cgagcccgcg gaggtaaccg agttctacaa gcacgtcagc caccgactggg 63720
cggaccogct ggagaccatc cacatgcggg gggaggggac cttcgagtac gaggcgctgc 63780
tgttcatccc caccacgcg cgcgtggacc tgttctcccc ccaggggccgc cgaggggtgc 63840
agctctacgt caagcgcggtg ttcattcatg acgactgcga cgcgctgatg cccaactacc 63900
tgcggttcgt caaggggtgtg gtggacgccc acgacctgtc gctgaacatc tcccgggaga 63960
tcctccagca ggaccgccag atccgcgcgg tccgcgctcg cctggtcaag aagatcctcg 64020
ccacggtcaa ggatctcaag accggccagc cggagcgcta ccgcaccttc tggaccgagt 64080
tcgggacggt ggtcaaggaa ggtctgctgg aggacaccga caaccgggac accctgctgg 64140
agatcctctc ggtggcctcc accaacgacc cggcogagca gaccgacctg gccggctacg 64200
tggaacgcat gcgcgagggc cagcaggaca tctactacgc caccggcgag aaccggagca 64260
ccatcgagaa ctccccgcac attgaggcgt tccgcgccc aaaggctacgag gtgctgctgc 64320
taaccgaccc ggtcgacgag gtttgggtgg agcggttcgg cagctacgcg ggcaagacgc 64380
tgcgctcggc cgccaagggg caggctcgacc tggacaccga gcaggagcgc accgaggccg 64440
aagccgaacg ggagcggcag cgcaccgagt acgcccacct gctcggtcgg atgggcgagg 64500
tgttggccga caacgtcaag gaggtccggc tgtcgtcgcg gctgaccacc tcgccggcct 64560
gcgtggtcgg cgacgcgcac gacctcacgc ccacctgga gaagatgtac cgcgcgatgg 64620
ggcaggaggt gccgaaggtc aagcggatcc

```

<210> 2

<211> 178

<212> PRT

<213> *Micromonospora* sp., cepa ML1

<220>

<223> Proteína ORF1

<400> 2

```

Ile Pro Ser Arg Glu Trp Gln Pro His Ser Cys Gly Pro Gly Ala Lys
1           5           10           15

```

```

Gly Pro Arg Glu Tyr Leu Trp Ala Trp Ile Thr Thr Ala Thr Ser Ala
          20           25           30

```

```

Gly Glu His Arg Trp Leu Leu Met Arg Arg Asn Arg Ser Thr Gly Glu
          35           40           45

```

```

Leu Ala Phe Tyr Leu Cys Trp Ser Pro Arg Pro Ala Pro Leu His Thr
          50           55           60

```

```

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

```

```

Ser Gly Lys Gly Gln Val Gly Leu Asp His Tyr Gln Val Arg Gly Trp
          85           90           95

```

```

Thr Gly Trp His Arg His Val Thr Leu Ala Met Leu Ala Leu Ala Val
          100          105          110

```

Leu Thr Ile Leu Ala Ala Arg Gln Pro Asp Pro Asp Pro Glu Ile Ile
 115 120 125

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

Ala Val Thr Leu Pro Pro Ala His Thr Leu His Trp Ser Thr Trp Arg
 145 150 155 160

Arg Thr Ser Gln Ala Arg Ala Arg Arg Ala His Tyr Gln Arg Arg Gln
 165 170 175

Ala Lys

<210> 3
 <211> 46
 <212> PRT
 <213> *Micromonospora sp.*, cepa ML1

<220>
 <223> Proteína ORF2

<400> 3
 Met Arg Ser Cys Gly Arg Arg Leu Ala Arg Val Thr Arg Ile Gly Asn
 5 10 15
 Pro Arg Gln Gly Leu Gln Ala Pro Arg Gly Gln Gln Ser Val Ser Arg
 20 25 30
 Ala Gly Arg Gly Met Leu Thr Gly Tyr Val Asp Gln Arg Arg
 35 40 45

<210> 4
 <211> 205
 <212> PRT
 <213> *Micromonospora sp.*, cepa ML1

<220>
 <223> Proteína Tio3

<400> 4
 Met Ser Val Ser Gly Ser Arg Arg Ser Glu Gly Ser Tyr Arg Arg Ser
 5 10 15
 Gly Gly Cys Ala Gln Gly Arg Ser Pro Arg Pro Ala Arg Glu Asp Val
 20 25 30
 Lys Phe Val Arg Trp Pro Ala Glu Arg His Leu Arg Asp Leu Cys Arg
 35 40 45
 Gln Glu Arg Ile Pro Cys Leu Leu Val Val Glu Gly Gly Ala Glu Pro
 50 55 60

Pro Val Cys Ser Asp Pro Arg Glu Asp Trp Val Arg Thr Pro Ile Ser
 65 70 75 80
 Pro Ser Asp Leu Asp Ala Arg Val Thr Ala Leu Arg Gln Arg Ser His
 85 90 95
 Asp Arg Arg Thr Pro Val Leu Asp Pro Ala Gly Thr Leu Ser Phe Arg
 100 105 110
 Thr Ala Ser Val Val Met Ser Ser Thr His Leu Glu Leu Met Asp Leu
 115 120 125
 Phe Val Glu Arg Phe Gly Gln Val Val Tyr Arg Trp Glu Leu Met Thr
 130 135 140
 Ile Leu Thr Lys Arg Asp Ser Thr Ser Thr Arg Asn Ala Leu Asp Leu
 145 150 155 160
 His Ile Met Arg Met Arg Lys Arg Ile Ala Pro Leu Asp Leu Asp Ile
 165 170 175
 Val Thr Ala Trp Gly Arg Gly Tyr Ala Leu Glu Glu Lys Ser Ser Trp
 180 185 190
 Ala Ala Val Ala Ser Lys Pro Ala Ala Val Asp His Cys
 195 200 205

<210> 5
 <211> 223
 <212> PRT
 <213> *Micromonospora sp.*, cepa ML1

<220>
 <223> Proteína Tio4

<400> 5
 Met Ser Ala Gly Ser Arg Ser Ser Arg Asn Ala Val Ala Ala Leu Arg
 5 10 15
 Val Val Pro Ala Ala Thr Ala Gly Val Pro Lys Gln Arg Pro Ala Glu
 20 25 30
 Pro Ala Gln Val Met Leu Thr Arg Val Gly Leu Arg Leu Pro Glu Arg
 35 40 45
 Met Ser Phe Glu His Trp Glu Arg Thr Gly Ala Gln Leu Ser Gly Ile
 50 55 60
 Val Asp Ser Ser Ser Trp Cys Leu Gly Asp Trp Leu Val Tyr Gly Lys
 65 70 75 80
 Glu Gln Tyr Ala Asp Arg Tyr Arg His Val Val Gln Ala Val Gly Leu
 85 90 95

Asp Tyr Gln Thr Leu Arg Asn Tyr Ala Trp Val Ala Gly Arg Phe Pro
 100 105 110
 Leu Ala Arg Arg Arg Ser Gly Leu Ser Phe Gln His His Ala Glu Val
 115 120 125
 Ala Ala Met Ser Pro Asp Asp Gln Asp Arg Trp Leu Asp Glu Ala Glu
 130 135 140
 Thr Gln Met Trp Ser Val Lys Gln Leu Arg Ala Ala Ile Lys Ala Thr
 145 150 155 160
 Arg Ser Asp Arg Thr Thr Glu Thr Ile Lys Thr Ala Leu Met Pro Arg
 165 170 175
 Ile Pro Ile Pro Asp Arg Arg Leu Thr Arg Trp Arg Gln Ala Ala His
 180 185 190
 Gln Ser Ser Val Glu Leu Asp Gln Trp Ile Val Ser Thr Leu Asp Arg
 195 200 205
 Ala Ala Ala Glu Thr Leu Asp Lys Thr Thr His Asp Glu Thr Glu
 210 215 220

<210> 6
 <211> 274
 <212> PRT
 <213> *Micromonospora* sp., cepa ML1

<220>
 <223> Proteína Tio5

<400> 6
 Met Val Ser Thr Ser Leu Lys Pro Arg Leu Asp Ala Ala Ala Ala Gln
 5 10 15
 Gln Gln Thr Gly Leu Ser Trp Trp Leu Ser Asp Ala Trp Gln Met Ala
 20 25 30
 Leu Arg Asn Leu Arg His Leu Thr Arg Ser Pro Glu Leu Val Met Phe
 35 40 45
 Ser Leu Val Gln Pro Val Met Phe Ile Leu Leu Phe Ala Tyr Val Phe
 50 55 60
 Gly Gly Ala Ile Asn Val Glu Gly Gly Asn Tyr Pro Gln Phe Leu Leu
 65 70 75 80

Pro Gly Ile Leu Val Gln Met Val Leu Phe Gly Ser Val Ala Gly Thr
 85 90 95
 Thr Ile Gly Ile Ser Thr Asp Met Asp Arg Gly Leu Met Asp Arg Phe
 100 105 110
 Arg Ser Met Pro Met Ser Arg Ser Ala Val Leu Val Gly Arg Thr Leu
 115 120 125
 Ser Glu Ile Leu Arg Asn Val Val Ser Met Leu Val Val Val Leu Met
 130 135 140
 Gly Leu Leu Ile Gly Phe Arg Phe Gln Gly Gly Val Leu Ser Ala Leu
 145 150 155 160
 Gly Gly Leu Ala Leu Leu Leu Leu Phe Gly Tyr Ala Leu Thr Trp Val
 165 170 175
 Gly Ala Met Ile Gly Leu Ser Val Pro Asn Pro Glu Ala Ala Gln Ser
 180 185 190
 Ala Gly Leu Ile Trp Ile Phe Pro Phe Ser Leu Ile Ser Ser Ala Phe
 195 200 205
 Val Pro Thr Glu Ser Met Pro Thr Trp Leu Arg Val Tyr Ala Asp Ala
 210 215 220
 Ser Pro Met Thr Ala Thr Val Asn Ala Leu Arg Asp Leu Phe Asn Gly
 225 230 235 240
 Asn Pro Pro Gly Leu Asp Val Leu Leu Thr Val Ala Trp Ser Val Leu
 245 250 255
 Leu Ile Thr Ile Phe Ala Pro Leu Ala Thr Arg Lys Tyr Ser Leu Arg
 260 265 270
 Ser Arg

<210> 7
 <211> 328
 <212> PRT
 <213> *Micromonospora sp.*, cepa ML1

<220>
 <223> Proteína Tio6

<400> 7
 Met Pro Ala Ala Ile Glu Val Ala Gly Leu Cys Lys Arg Phe Gly Ser
 5 10 15
 Val Thr Ala Leu Asp His Leu Glu Leu Thr Val Pro Glu Gly Lys Val
 20 25 30

His Gly Leu Leu Gly Pro Asn Gly Ala Gly Lys Thr Thr Thr Val Arg
 35 40 45
 Val Leu Thr Thr Leu Leu Arg Pro Asp Glu Gly Lys Ala Glu Ile Phe
 50 55 60
 Gly Phe Asp Val Arg Arg Asn Ala Leu Gln Val Arg Ser Ile Ile Gly
 65 70 75 80
 Leu Thr Gly Gln Tyr Thr Thr Val Asp Asp Met Leu Thr Gly Arg Glu
 85 90 95
 Asn Leu His Met Val Gly Lys Leu Phe Arg Met Ser Thr Pro Asp Ile
 100 105 110
 Lys Val Arg Ala Thr Glu Leu Leu Glu Arg Phe Asp Leu Thr Glu Ala
 115 120 125
 Ala Gly Arg Gln Ala Lys Thr Tyr Ser Gly Gly Met Arg Arg Arg Leu
 130 135 140
 Asp Ile Ala Ala Thr Leu Met Ser Arg Pro Gln Leu Ile Phe Leu Asp
 145 150 155 160
 Glu Pro Thr Ala Ala Leu Asp Pro Arg Ser Arg Met Asp Val Trp Arg
 165 170 175
 Ile Val Asn Glu Leu Val Glu Glu Gly Thr Thr Val Leu Leu Thr Thr
 180 185 190
 Gln Tyr Leu Glu Glu Ala Asp His Leu Ala Asp Ser Leu Ser Val Ile
 195 200 205
 Asp Ser Gly Lys Val Ile Ala Ser Gly Ala Pro Asp Glu Leu Lys Ala
 210 215 220
 Glu Val Gly Gln Ala Trp Ala Asp Val Thr Leu Ile Asp Ala Ala Ala
 225 230 235 240
 Val Glu Lys Thr Thr Gln Ile Leu Arg Glu Arg Leu Gly Thr Glu Pro
 245 250 255
 Val Gly Asp Pro Glu Gly Met Gln Val His Ala Pro Leu Pro Glu Gly
 260 265 270
 Met Thr Leu Asn Glu Leu Leu Asn Pro Leu Arg Asp Ser Gly Ile Ala
 275 280 285
 Ile Ala Asp Val Ala Val Arg Arg Pro Thr Leu Asp Asp Val Phe Leu
 290 295 300
 Ala Leu Thr Gly His Thr Ala Thr Ala Arg Ala Ser Glu Asp Val Asn
 305 310 315 320
 Ser Lys Lys Lys Gly Gly Lys Arg
 325

```

<400>      8
Met Tyr Phe Arg Pro Ala Asp Leu Ala His Glu His Gly Leu Ser Pro
           5                      10                      15

Gln Ser Val Arg Asn Tyr Glu Arg Asp Gly Leu Ile Pro Pro Ala Arg
           20                      25                      30

Arg Thr Glu Ser Gly Tyr Arg Arg Tyr Thr Glu Lys His Ala Ala Ala
           35                      40                      45

Leu Arg Ala Tyr Arg Ala Leu Ile Pro Ala His Gly Tyr Ala Glu Ser
           50                      55                      60

Gly Ala Ile Met Arg Asp Ile Thr Ala Gly Arg Leu Asn Glu Ala Leu
           65                      70                      75                      80

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

Leu Asp Ala Val Ala Glu Val Leu Thr His Leu Thr Gly Arg Ser Thr
           100                     105                     110

Gly Thr Trp Arg Ala Pro Lys Ser Ile Gln Pro Tyr Ser Ile Gly Glu
           115                     120                     125

Leu Ala His Arg Leu Gly Val Asn Val Ala Thr Ile Arg Lys Trp Glu
           130                     135                     140

Thr Ser Gly Val Leu Ala Pro Val Arg Arg Pro Glu Thr Gly His Arg
           145                     150                     155                     160

Met Tyr Asp Ala Ser Asp Val Arg Asp Ala Glu Leu Ala His Phe Leu
           165                     170                     175

Arg Arg Gly Arg Tyr Pro Leu Glu Leu Ile Ala Thr Val Val Gln Gln
           180                     185                     190

Val Arg Ala Ala Gly Asp Thr Gln Ala Leu Ala Ala Ala Leu Ala Asp
           195                     200                     205

Trp Arg Gln Arg Val Thr Thr Arg Gly Leu Ala Met Leu Asn Ala Ala
           210                     215                     220

Ala Gln Leu Ala Ser Tyr Leu Ala Met Glu Glu Asn
           225                     230                     235

```

<210> 9
 <211> 242
 <212> PRT
 <213> *Micromonospora sp.*, cepa ML1

<220>
 <223> Proteína Tio8

<400> 9
 Met Thr Thr Leu Thr Tyr Ala Asp Tyr Leu Gln Val Asp Thr Leu Leu
 5 10 15
 Ser Leu Gln Glu Thr Arg Ala Pro Asp Thr Ala Ser Arg Ser Val Val
 20 25 30
 Leu Ala Glu His Phe Phe Ile Val Thr His Gln Ala Cys Glu Leu Trp
 35 40 45
 Leu Lys Gln Val Thr Tyr Asp Leu Asp Ala Ala Ala Asp Ala Met Thr
 50 55 60
 Pro Pro Tyr Gly Pro Glu Asp Glu Glu Leu Ala Leu Glu Phe Leu Gln
 65 70 75 80
 Arg Thr Ala Glu Leu Leu Arg Val Leu His Glu Gln Val Val Ala Leu
 85 90 95
 Glu Arg Leu Pro Leu Arg His Phe Ala Glu Phe Arg Pro Phe Leu Asp
 100 105 110
 Gly Ala Ser Gly Ala Gln Ser Ala Gln Phe Arg Leu Leu Ala Arg Leu
 115 120 125
 Leu Gly Gly Glu Arg Gln Glu Gly Arg Leu Tyr Arg Ala Phe Glu Arg
 130 135 140
 Leu Thr Glu His Asn Gly Ile Ser Val Ala Glu Val Cys Ala Arg Gly
 145 150 155 160
 Pro Gly Ala Gly Val Leu Asn Arg Ile Ala Asp Ser Leu Leu Asp Val
 165 170 175
 Ala Asn Gly Tyr Trp Arg Trp Lys Met Thr His Leu Ala Leu Val Ser
 180 185 190
 Lys Met Val Gly Asn Arg Asn Gly Thr Gly Gly Ser Thr Gly Val Asp
 195 200 205
 Tyr Leu Ala Arg Arg Ile Thr Leu Pro Phe Ala Gln Leu Arg Gln Leu
 210 215 220
 Arg Ala Gln Val His Asp Ala Leu Val Trp Glu Gln Pro Thr Ala His
 225 230 235 240

Ser Thr Ala Ser Leu Glu Trp Asp Val Pro Val Ser Glu Arg Ile Val
 225 230 235 240
 His Ser Ser Tyr Ser Phe Ser Lys Ser Leu Ala Met Thr Gly Tyr Arg
 245 250 255
 Leu Gly Tyr Val Ser Thr Ala Asn Asp Arg Thr Ala His Ala Phe Arg
 260 265 270
 Ile Val Gln Glu Ala Asn Leu Ile Ala Met Ser Thr Pro Val Gln Tyr
 275 280 285
 Ala Gly Leu Glu Ala Leu Arg Leu Gly His Val Val Arg Glu Asn Arg
 290 295 300
 Ala Met Val Leu Ser Asn Arg Asp Ala Val Leu Pro Ala Val Arg Asp
 305 310 315 320
 Ala Gly Leu Leu Pro Glu Leu Pro Ala Gly Gly Trp Tyr Ala Val Leu
 325 330 335
 Asp Val Ala Pro Thr Gly Leu Thr Ala Glu Glu Phe Ala Thr Gly Leu
 340 345 350
 Leu Glu Thr Arg Asn Val Ala Val Val Pro Gly Thr Gly Phe Ala Leu
 355 360 365
 Met Pro Arg Leu Asp Ala Arg Gly Arg Val Ile Thr Val Glu Pro Ala
 370 375 380
 Pro Trp Ser Arg His Leu Ile Arg Ile Ala Met Cys Val Ala Pro Ser
 385 390 395 400
 Val Leu Arg Ser Gly Val Arg Glu Leu Leu Glu Phe Ala Arg Glu Arg
 405 410 415
 Pro Gly Thr Ala Arg
 420
 <210> 11
 <211> 264
 <212> PRT
 <213> *Micromonospora sp.*, cepa ML1
 <220>
 <223> Proteína Tio10
 <400> 11
 Val Ser Pro Thr Glu Ala Glu Ala Trp Pro Ala Val Ala Glu Arg Gly
 5 10 15
 Asp Gly Leu Ala Gly Arg Thr Val Ile Val Thr Gly Ala Gly Thr Gly
 20 25 30

Leu Gly Arg Ala Ile Gly Glu Arg Leu Leu Ala Arg Asp Ala Asn Val
 35 40 45
 Val Ala Val Gly Arg Arg Glu Ala Ala Leu Arg Glu Leu Thr Gly Asp
 50 55 60
 Gly Gly Gly Gln Val Ala Thr Val Ala Ala Asp Ile Thr Glu Pro Gly
 65 70 75 80
 Ala Ala Glu His Val Val Ala Ala Ala Thr Thr Arg Phe Gly Thr Val
 85 90 95
 His Gly Leu Val Asn Asn Ala Gly Leu Ala Arg Phe Ala Thr Leu Asp
 100 105 110
 Thr Ala Ser Asp Glu Asp Phe Asp Arg Leu Leu Ala Thr Asn Val Arg
 115 120 125
 Ala Pro Ala Ala Leu Ile Arg Ala Ala Leu Pro His Leu Arg Glu His
 130 135 140
 Arg Gly Ala Val Val Asn Val Ser Ser Val Gly Ala Val Leu Ala Met
 145 150 155 160
 Pro Gly Arg Ser Phe Tyr Gly Ala Ser Lys Ala Ala Leu Asn Ser Leu
 165 170 175
 Thr Arg Ser Leu Ala Arg Glu Leu Ala Pro Asp Val Arg Val Asn Ala
 180 185 190
 Val Leu Pro Gly Pro Val Asp Thr Pro Met Trp Asp Asp Met Gly Leu
 195 200 205
 Asp Ser Val Lys Ser Ala Leu Leu Arg Ser Ser Leu Leu Ala Ala Thr
 210 215 220
 Pro Leu Gly Arg Phe Gly Gln Pro Ala Glu Ile Ala Gly Trp Val Cys
 225 230 235 240
 Gly Leu Leu Asp Pro Glu Phe Ser Gly Trp Ile Thr Gly Ala Leu Ile
 245 250 255
 Pro Val Asp Gly Gly Arg Thr Ser
 260

<210> 12
 <211> 404
 <212> PRT
 <213> *Micromonospora sp.*, cepa ML1

<220>
 <223> Proteína Tio11

<400> 12

Gly Gln Gln Val Arg Ala Gly Asp Ile Val Thr Leu Trp Asn Leu Ser
305 310 315 320

Ala Asn Phe Asp Glu Arg Glu Phe Asp Arg Pro Tyr Arg Phe Glu Val
325 330 335

Gly Arg Thr Pro Asn Lys His Leu Ser Phe Gly His Gly Pro His Phe
340 345 350

Cys Leu Gly Ala Tyr Leu Gly Arg Ala Glu Leu Gln Ala Leu Leu Thr
355 360 365

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

Arg Val Tyr Ser Asn Phe Leu Asn Gly His Ser Ser Leu Pro Val Ala
385 390 395 400

Phe Thr Gly Arg

<210> 13

<211> 523

<212> PRT

<213> *Micromonospora* sp., cepa ML1

<220>

<223> Proteína Tio12

<400> 13

Val Pro Trp Pro Glu Glu Leu Ala Ala Leu Tyr Arg Arg Glu Gly Tyr
5 10 15

Trp Gln Gly Arg Pro Leu Gly Asp Leu Leu Thr Glu Ala Cys Ala Arg
20 25 30

His Ala Asp Lys Thr Ala Leu Val Cys Gly Glu Arg Arg Met Thr Tyr
35 40 45

Ala Glu Leu Ser Arg Ala Ser Ala Asp Val Ala Gly Gly Leu Leu Asp
50 55 60

Val Gly Ile Glu Pro Leu Asp Arg Val Val Val Gln Leu Pro Asn Val
65 70 75 80

Pro Glu Phe Val Val Val Val Tyr Ala Leu Leu Arg Ile Gly Ala Ile
85 90 95

Pro Val Met Ala Leu Pro Gly His Arg Lys Val Glu Leu Thr His Leu
100 105 110

Cys Ala His Ser Gln Ala Val Ala Leu Val Val Val Asp Gln Val Lys
115 120 125

Lys Val Pro Ala Glu Glu Ile Glu Glu His Leu Leu Ala His Ala Ala
435 440 445

Ile Arg Asp Ala Ala Val Val Gly Val Ala Asp Pro Val Leu Gly Glu
450 455 460

Lys Thr Val Ala Phe Val Ile Leu Arg Asp Glu Pro Ile Ser Pro Ala
465 470 475 480

Ala Val Lys Ser Phe Leu Arg Glu Arg Gly Leu Ala Thr Tyr Lys Ile
485 490 495

Pro Asp Arg Val Val Pro Ile Glu Glu Phe Pro Arg Thr Ala Val Gly
500 505 510

Lys Val Asp Lys Val Thr Leu Arg Thr Ser Gly
515 520

<210> 14

<211> 596

<212> PRT

<213> *Micromonospora* sp., cepa ML1

<220>

<223> Proteína Tio13

<400> 14

Val Pro Thr Val Ala Glu Leu Phe Glu Ser Val Val Ala Arg Ser Gly
5 10 15

Asp His Val Ala Val Ile Phe Gly Asp Arg Leu Val Ser Tyr Ala Glu
20 25 30

Leu Asp Ala Arg Ala Asn Arg Leu Ala Arg Val Leu Val Glu Arg Gly
35 40 45

Val Gly Pro Asp Thr Thr Val Ala Val Ala Met Pro Lys Cys Asp Glu
50 55 60

Leu Met Val Val Leu Leu Ala Val Leu Lys Ala Gly Gly Ala Tyr Leu
65 70 75 80

Pro Leu Asp Pro Gln Tyr Pro Ala Lys Arg Leu Ala Tyr Met Val Gln
85 90 95

Asp Ala Arg Pro Thr Leu Leu Val Arg Met Ala Gln Val Ser Leu Glu
100 105 110

Leu Gly Val Pro Val Ala Glu Leu Val Val Asp Asp Pro Asp Thr Ala
115 120 125

Arg Glu Thr Glu Ser Arg Pro Asp His Gln Leu Thr Asp Ala Ser Arg
130 135 140

Asn	Ala	Pro	Leu	Arg	Pro	Asp	Asn	Leu	Met	Tyr	Val	Ile	Tyr	Thr	Ser	145	150	155	160
Gly	Ser	Thr	Gly	Thr	Pro	Lys	Gly	Val	Ala	Val	Thr	His	Thr	Gly	Val	165	170	175	
Ala	Asp	Ile	Val	Ala	Ala	Gln	Thr	Ala	Ser	Ile	Ala	Pro	Arg	Pro	Gly	180	185	190	
Asp	Arg	Val	Leu	Gln	Trp	Ala	Ser	Val	Ser	Phe	Asp	Ala	Ala	Phe	Trp	195	200	205	
Asp	Trp	Ser	Ala	Ala	Leu	Leu	Ser	Gly	Ala	Thr	Leu	Ile	Met	Ala	Pro	210	215	220	
Ala	Asp	Glu	Leu	Leu	Pro	Gly	Gln	Pro	Leu	Arg	Asp	Thr	Leu	Arg	Arg	225	230	235	240
Tyr	Ala	Val	Thr	His	Ala	Val	Leu	Pro	Pro	Val	Ala	Leu	Ser	Val	Thr	245	250	255	
Glu	Pro	Asp	Asp	Val	Leu	Val	Gly	Gly	Thr	Leu	Met	Ser	Thr	Gly	Asp	260	265	270	
Ala	Cys	Thr	Arg	Ala	Leu	Val	Ala	Lys	Trp	Ala	Pro	Gly	Arg	Arg	Met	275	280	285	
Tyr	Asn	Gly	Tyr	Gly	Pro	Thr	Glu	Thr	Thr	Val	Gly	Ser	Thr	Ile	Ala	290	295	300	
Gly	Pro	Ile	Ala	Ile	Ser	Asp	Glu	Ile	Thr	Ile	Gly	Thr	Pro	Trp	Ser	305	310	315	320
Gly	Asn	Arg	Val	Tyr	Val	Leu	Asp	Glu	Arg	Leu	Arg	Pro	Val	Pro	Asn	325	330	335	
Gly	Arg	Asp	Gly	Glu	Leu	Tyr	Leu	Ala	Gly	Asn	Gly	Leu	Ala	Arg	Gly	340	345	350	
Tyr	Leu	Asn	Arg	Pro	Gly	Leu	Thr	Ala	Ser	Arg	Phe	Val	Leu	Asp	Pro	355	360	365	
Phe	Gly	Pro	Pro	Gly	Gly	Arg	Met	Tyr	Arg	Ser	Gly	Asp	Arg	Gly	His	370	375	380	
Arg	Arg	Ala	Asp	Gly	Gln	Leu	Val	Phe	Ala	Ser	Arg	Gly	Asp	Asp	Gln	385	390	395	400
Val	Lys	Ile	Arg	Gly	Phe	Arg	Val	Glu	Leu	Gly	Glu	Val	Glu	Ala	Arg	405	410	415	
Leu	Ala	Ala	His	Pro	Ala	Val	Asp	Val	Ala	Ala	Val	Val	Val	Val	Gly	420	425	430	

Asp Leu Ala Gly Ala Arg Leu Ala Ala Phe Val Ser Thr Val Ala Gly
 435 440 445
 Thr Gly Val Ser Ala Ala Glu Leu Ser Ala His Ser Ala Glu Thr Leu
 450 455 460
 Pro Glu His Met Val Pro Ser Ser Val Gln Val Leu Gln Arg Leu Pro
 465 470 475 480
 Val Thr Ala Asn Gly Lys Ile Asp Arg Ala Ala Leu Arg Asp Leu Ala
 485 490 495
 Glu Ser Ala Ala His Pro Asp Ala Ala Ala Ser Ala Gly Pro Pro Gly
 500 505 510
 Glu Asp Pro Thr Ser Val Glu Ser Arg Leu Cys Leu Leu Val Arg Glu
 515 520 525
 Leu Leu Ala Val Pro Asp Val Arg Pro Gly Asp Asn Phe Phe Lys Leu
 530 535 540
 Gly Gly His Ser Val Leu Ala Thr Arg Leu Ala Gly Arg Leu Arg Asp
 545 550 555 560
 Glu Trp Gly Ile Thr Val Pro Ile Arg Ala Val Phe Glu Ala Ala Thr
 565 570 575
 Leu Ser Glu Leu Ala Leu Val Val Gln Ala Ser Lys Ala Ser Ile Pro
 580 585 590
 Asn Gly Arg Pro
 595

<210> 15
 <211> 424
 <212> PRT
 <213> *Micromonospora* sp., cepa ML1

<220>
 <223> Proteína Tio14

<400> 15
 Val Thr Gly Val Arg Thr Leu Thr Val Val Ala Ile Leu Ala Leu Thr
 5 10 15
 Ala Thr Ser His Gln Val Glu Pro Val Ala Ala Ser Gly Asp Val Leu
 20 25 30
 Gly Val Ala Thr Leu Ala Asp Thr Gln Thr Thr His Ala Leu Ile Ala
 35 40 45
 Asn Pro His Leu Arg Gly Thr Gly Ser Thr Ser Ser Gly Pro Gly Gly
 50 55 60

Ala	Pro	Ala	Glu	Ala	Leu	Ser	Arg	Phe	His	Asp	Leu	Gly	Ser	Leu	Ala	65	70	75	80
Gly	Pro	Gly	Thr	Pro	Ser	Glu	Ala	Thr	Asp	Leu	Asn	Asp	Arg	Gly	Asp	85	90	95	
Val	Val	Gly	Val	Thr	Ser	Leu	Pro	Gly	Thr	Gln	Thr	Ala	His	Ala	Phe	100	105	110	
Ile	Ala	Asn	Pro	Arg	His	Lys	Gly	Gly	Arg	Leu	Ile	Glu	Ile	Thr	Pro	115	120	125	
Ser	Gln	Ala	Gly	Ser	Ser	Arg	Ala	Gln	Ala	Ile	Asn	Arg	Phe	Gly	Val	130	135	140	
Val	Ala	Gly	Thr	Leu	Gly	Leu	Gly	Ile	Gly	Pro	Pro	Thr	Pro	Ala	Pro	145	150	155	160
Asp	Gly	Ser	Gln	Pro	Leu	Leu	Gln	Pro	Phe	Val	Trp	Glu	Ala	Arg	Thr	165	170	175	
Gly	Leu	Asp	Val	Leu	Pro	Leu	Pro	Pro	Gly	Ala	Gly	Gly	Ala	Gln	Ala	180	185	190	
Val	Asp	Ile	Asn	Asp	Lys	Gly	Thr	Val	Leu	Val	Val	Gly	Thr	Asn	Ala	195	200	205	
Thr	Gly	Val	Gly	Leu	Pro	Val	Gly	Ser	Tyr	Leu	Trp	Asn	Pro	Ala	Ser	210	215	220	
Arg	Thr	Tyr	Thr	Thr	Leu	Pro	Pro	Pro	Thr	Thr	Ser	Val	Thr	Gly	Pro	225	230	235	240
Val	Ala	Leu	Ala	His	Thr	Leu	Asp	Glu	Arg	Gly	Gly	Val	Ala	Gly	Gly	245	250	255	
Leu	Val	Thr	Gln	Val	Gly	Glu	Gln	Ser	Trp	His	His	Thr	Ala	Val	Val	260	265	270	
Trp	Gln	Pro	Gly	Thr	Leu	Ala	Val	Gln	Gln	Leu	Pro	Thr	Gly	Gly	Ser	275	280	285	
Ala	Asp	Thr	Tyr	Ala	Thr	Asp	Arg	Asn	Glu	Asn	Gly	Leu	Ile	Val	Gly	290	295	300	
Trp	Arg	Ile	Asn	Ala	Thr	Gly	Gln	Ser	Thr	Thr	Ala	Ile	Tyr	Trp	Pro	305	310	315	320
Ser	Ala	Asn	Ala	Gln	Pro	Val	Glu	Leu	Pro	Gly	Arg	Val	Ala	Tyr	Arg	325	330	335	
Val	Asn	Asp	Ala	Gly	Gln	Ile	Val	Gly	Ile	Arg	Asn	Phe	Thr	Gly	Ser	340	345	350	
Ser	Ala	Tyr	Pro	Phe	Ser	Ala	Val	Met	Trp	Glu	Pro	Arg	Arg	Gly	Arg	355	360	365	

Thr Thr Arg Leu Gly Asp Asn Asp Leu Gly Ser Tyr Val Ile Ala Val
 370 375 380
 Asn Ala Ser Gly Arg Ser Ala Gly Phe Ala Val Ala Thr Arg Asn Gly
 385 390 395 400
 Asn Asn His Asn Thr Ala Gly Trp Trp Gly Pro Pro Arg Arg Gly Asp
 405 410 415
 Cys Pro Arg His Cys Ser Pro Arg
 420

<210> 16
 <211> 450
 <212> PRT
 <213> *Micromonospora sp.*, cepa ML1

<220>
 <223> Proteína Tio15

<400> 16
 Met Thr Thr Arg Ser Thr Glu Pro Ser Leu Ser Pro Leu Val Pro Gly
 5 10 15
 Leu Ser Arg Arg Gly Leu Leu Ile Ala Ala Ala Ser Thr Ala Ala Gly
 20 25 30
 Thr Ala Gly Leu Leu Ile Leu Pro Thr Lys Ala Glu Ala Ala Thr Ala
 35 40 45
 Ser Pro Arg His Gly Ser Arg Pro Asp Val Val Gly Pro Trp Phe Asp
 50 55 60
 Glu Met Leu Ala Thr Phe Ala Ala Ala Pro Thr Ala Gly Leu Pro Glu
 65 70 75 80
 Arg Met Trp Ala Met Ser Trp Ile Ala Ala Trp Ala Ala Leu Ser Ala
 85 90 95
 Arg Arg Ala Ala Pro Val Arg Ala Pro Gly Gln Arg Ala Val Phe Glu
 100 105 110
 Asp Ala Ala Thr Ala Thr Ala Val His Gly Val Leu Lys Ala Leu Val
 115 120 125
 Pro Gly Gln Ala Asp Lys Leu Asp Gly Val Leu Ala Asp Ser Leu Ala
 130 135 140
 Gly Leu Ala Ala Ser Glu Ala Lys Lys Ala Gly Ile Ala Ala Gly Arg
 145 150 155 160
 Glu Ala Ala Ala Ala Met Val Ala Val Arg Thr Gly Asp Gly Leu Asp
 165 170 175

Leu Ala Ser Val Asn Val Pro Phe Thr Leu Pro Pro Glu Ala Pro Gly
 180 185 190
 Val Tyr Arg Leu Thr Pro Gly Ala Thr Gln Thr Val Ala Ala Gly Tyr
 195 200 205
 Gly Leu Ala Arg Pro Phe Leu Leu Gly Arg Ala Asp Arg Phe Arg Pro
 210 215 220
 Pro Pro Pro Pro Ala Leu Gly Thr Ala Pro Tyr Arg Arg Ala Leu Asp
 225 230 235 240
 Glu Val Arg Arg Leu Gly Gly Glu Val Ser Glu Arg Thr Glu Ala Gln
 245 250 255
 Ser Asp Leu Ala Trp Leu Asp Pro Leu Ser Gln Tyr Val Pro Val Leu
 260 265 270
 Arg Leu Leu Val Gly Asp Ser Ala Arg Ser Arg Arg Ser Lys Val Arg
 275 280 285
 Leu Leu Ala Ala Leu Gly Ala Ala Ile Val Asp Ser Ser Ile Ala Thr
 290 295 300
 Phe Glu Ala Lys Tyr Ala Leu Leu His Trp Arg Pro Val Thr Ala Ile
 305 310 315 320
 Arg Asp Ala Asp Thr Asp Gly Asp Pro Arg Thr Ala Pro Asp Pro Thr
 325 330 335
 Trp Asn Ser Phe Leu Pro Thr Pro Pro His Pro Glu Tyr Pro Ser Gly
 340 345 350
 His Ala Val Thr Ala Gly Ala Ala Glu Gln Val Leu Ser Gln Leu Val
 355 360 365
 Gly Pro Cys Thr Pro Met Ala Phe Ser Ile Thr Phe Ala Arg Gly Asp
 370 375 380
 Gly Arg Val Val Thr Arg Asn Tyr Arg Arg Gly Thr Pro Trp Ser Thr
 385 390 395 400
 Leu Thr Lys Glu Asn Ile Asp Ala Arg Val Leu Ala Gly Val His Phe
 405 410 415
 Arg Tyr Ser Asp Glu Val Gly Ala Ala Leu Gly Arg Gln Val Ala Arg
 420 425 430
 Tyr Asp Leu Asn Leu Leu Ala Arg Gly Gly Val Gly Pro Val Arg Val
 435 440 445
 Thr Trp
 450

<210> 17
 <211> 768
 <212> PRT
 <213> *Micromonospora sp.*, cepa ML1

<220>
 <223> Proteína Tio16

<400> 17

Val	Pro	Trp	Ser	Val	Trp	Asn	Gln	Thr	Ala	Arg	His	Phe	Arg	Ala	Asp
				5					10					15	
Val	Thr	Ile	Pro	Asp	Leu	Leu	Ala	Glu	Ala	Ala	Glu	Arg	His	Pro	Asp
			20					25					30		
Arg	Pro	Ala	Ile	Val	Thr	Ser	Asp	Lys	Gln	Val	Leu	Thr	His	Arg	Glu
		35					40					45			
Leu	His	Arg	Arg	Ala	Asn	Arg	Leu	Ala	Arg	Leu	Leu	Arg	Asp	Leu	Gly
	50					55					60				
Val	Gly	Arg	Gly	Asp	Thr	Val	Ala	Leu	Phe	Gly	Glu	Arg	Asp	Ala	Pro
65					70					75					80
Ala	Leu	Val	Gly	Leu	Leu	Ala	Val	Leu	Lys	Cys	Gly	Ala	Ala	Tyr	Val
			85						90					95	
Pro	Ile	Asp	Pro	Ser	Trp	Pro	Ala	Arg	Arg	Val	Val	Gly	Leu	Leu	Asp
			100					105					110		
Gln	Leu	Asp	Val	Ala	Ala	Val	Leu	Thr	Asp	Arg	Thr	His	Leu	Arg	Ser
		115					120					125			
Met	Gln	Glu	Leu	Arg	Trp	Gln	Val	Arg	Ser	Leu	Arg	His	Val	Leu	Cys
	130					135					140				
Pro	Asp	Ile	Ala	Glu	Glu	Phe	Ser	Trp	Ser	Ala	Thr	Phe	Arg	Pro	Asp
145					150					155					160
Leu	Val	Val	Asn	Phe	Phe	Asp	Phe	Val	Ser	Ser	Ala	Pro	Asp	Gln	Leu
			165						170					175	
Glu	Ala	Ala	Gly	Phe	Asn	Met	Arg	Arg	Ser	Glu	Arg	Pro	Phe	Asp	Glu
			180					185					190		
Asn	Asp	Val	Gln	Gln	Tyr	Arg	Asp	His	Val	Ala	Ala	Phe	Ala	Leu	Ala
		195					200					205			
Ala	Thr	Ala	Ala	Ala	Arg	Pro	Ser	Ile	Leu	Glu	Ile	Gly	Ser	Gly	Ser
	210					215					220				
Gly	Leu	Ile	Val	Glu	Ala	Leu	Ala	Pro	Arg	Ala	Ser	Arg	Tyr	Val	Ala
225					230					235					240

Ala	Asp	Pro	Ser	Pro	Val	Ser	Val	Thr	Arg	Ser	Gln	Glu	Gly	Gly	Phe	245	250	255
Arg	Ala	Gly	Gly	Glu	Val	Asp	Gly	Val	Val	Cys	Phe	Ala	His	Glu	Val	260	265	270
Ala	Glu	Arg	Val	Gln	Gly	Pro	Phe	Asp	Val	Ile	Leu	Leu	Ala	Ser	Thr	275	280	285
Val	Gln	Phe	Leu	Pro	Asp	Leu	Asp	Tyr	Leu	Leu	Ser	Val	Leu	Gly	Ser	290	295	300
Leu	Leu	Gly	Leu	Leu	Ala	Asp	Asp	Gly	Val	Ile	Val	Leu	Ala	Asp	Leu	305	310	315
Ile	Asp	Pro	Asp	Leu	Ser	Asp	Tyr	Ser	Gly	Ile	Arg	Val	Ala	Pro	His	325	330	335
Phe	Ile	Glu	Arg	Leu	Pro	Glu	Phe	Leu	Pro	Gly	Val	Thr	Gly	Gly	Ser	340	345	350
Val	His	Arg	Arg	Gln	Val	Glu	Thr	Phe	Thr	Gly	Glu	Leu	Ala	Gln	Arg	355	360	365
Tyr	Asp	Ala	Val	Ile	Arg	Val	Ala	Pro	Ser	Gly	Asp	Arg	Pro	Ala	Ser	370	375	380
Gly	Arg	Glu	Pro	Leu	Trp	Thr	Gly	His	His	Ile	Glu	Leu	Arg	Pro	Ser	385	390	395
Glu	Pro	Leu	Ala	Thr	Gly	Pro	Thr	Ala	Asp	Asp	Val	Cys	Tyr	Thr	Ile	405	410	415
Phe	Thr	Ser	Gly	Ser	Thr	Gly	Ala	Pro	Lys	Gly	Val	Val	Val	Ser	His	420	425	430
Arg	Ser	Ala	Val	Asn	Leu	Ile	Asp	Trp	Val	Asn	Arg	Ser	Tyr	Arg	Val	435	440	445
Gly	Pro	Thr	Asp	Gln	Ile	Leu	Phe	Val	Thr	Ser	Phe	Cys	Phe	Asp	Leu	450	455	460
Ser	Val	Tyr	Asp	Ile	Phe	Gly	Val	Leu	Ala	Ala	Gly	Gly	Ser	Val	Arg	465	470	475
Leu	Ala	Ser	Ala	Ala	Glu	Leu	Ala	Glu	Pro	Asp	Thr	Leu	Val	Asp	Leu	485	490	495
Leu	Glu	Gln	Glu	Glu	Ile	Thr	Ile	Trp	Asp	Ser	Ala	Pro	Ala	Ala	Leu	500	505	510
Ala	Met	Val	Met	Pro	Phe	Val	Gln	Val	Arg	Glu	Pro	Ala	Gly	Arg	Gly	515	520	525
Thr	Leu	Arg	Leu	Val	Leu	Leu	Ser	Gly	Asp	Trp	Ile	Pro	Val	Pro	Leu	530	535	540

Pro Asp Gln Ile Arg Ala Ala Phe Pro Asn Ala Glu Val Val Ala Leu
 545 550 555 560
 Gly Gly Ala Thr Glu Cys Thr Val Trp Ser Asn His Tyr Arg Val Gly
 565 570 575
 Asp Val Asp Pro Gly Trp Pro Ser Ile Pro Tyr Gly Lys Pro Met Gln
 580 585 590
 Asn Ala Arg Tyr Tyr Val Leu Asp His Gly Leu Gly Pro Cys Pro Asn
 595 600 605
 Gly Ala Glu Gly Asp Leu Tyr Ile Ala Gly Glu Cys Val Ala Leu Gly
 610 615 620
 Tyr Ala Arg Ser Pro Gly Leu Thr Ala Ala Lys Phe Leu Pro Asp Pro
 625 630 635 640
 Trp Ala His Arg Thr Gly Glu Arg Met Tyr His Thr Gly Asp Arg Ala
 645 650 655
 Arg Trp Leu Pro Ser Gly Asp Leu Glu Phe Leu Gly Arg Leu Asp Asp
 660 665 670
 Gln Val Lys Val Arg Gly Tyr Arg Ile Glu Leu Gly Glu Val Gln Ser
 675 680 685
 Ala Leu Leu Gln Cys Ala Gly Val Arg Ala Ala Ser Val Val Ala Pro
 690 695 700
 Ser Gly Pro Gly Gly Arg Thr Leu Ala Ala Phe Tyr Val Pro Thr Ala
 705 710 715 720
 Glu Pro Leu Pro Thr Ala Glu Val Val Gly Ala Leu Gln Arg Ile Leu
 725 730 735
 Pro Ser Tyr Met Val Pro Thr Arg Ile Val Ala Leu Asp Glu Leu Pro
 740 745 750
 Leu Thr Asn Thr Gly Lys Val Asp Arg Ala Ala Leu Leu Ser Arg Leu
 755 760 765

<210> 18
 <211> 94
 <212> PRT
 <213> *Micromonospora* sp., cepa ML1

<220>
 <223> Proteína Tio17

<400> 18
 Val Ala Asn Pro Asp Arg Asp Gly Asn Asp Gly Ser Ala Arg Gly Leu
 5 10 15

Leu Arg Ala His Val Arg Ala Gly Trp Ala Ala Ala Leu Gln His Glu
 20 25 30
 Arg Phe Asp Asn Asp Asp Asp Phe Phe Ala Ile Gly Gly His Ser Leu
 35 40 45
 Leu Ile Ala Glu Ile Met Ala Gly Leu Gly Ala Leu Ala Gly Thr Arg
 50 55 60
 Leu Ser Leu Lys Leu Phe Phe Asp His Pro Thr Val Asn Glu Leu Thr
 65 70 75 80
 Asp Ala Leu Ala Ala His Glu Ala Phe Gly Ser Leu Ala Arg
 85 90

<210> 19
 <211> 259
 <212> PRT
 <213> *Micromonospora* sp., cepa ML1

<220>
 <223> Proteína Tio18

<400> 19
 Met Gly Thr Arg Thr Lys Asn Arg Trp Phe Val Glu Val Ser Arg Ser
 5 10 15
 Arg Leu Lys Glu Cys Ala Leu Tyr Ala Ile Pro His Ala Gly Gly Gly
 20 25 30
 Val Ala Thr Val Arg Ser Leu Cys Val Glu Leu Gly Asp Ala Phe Asp
 35 40 45
 Thr Phe Gly Val Arg Leu Pro Gly Arg Glu Ser Arg Leu Gly Asp Pro
 50 55 60
 Val Val Thr Val Met Ser Arg Leu Ala Asp Glu Leu Ala Gly Glu Leu
 65 70 75 80
 Leu Arg His Ala Gln Gly Arg Pro Val Val Leu Tyr Gly His Cys Ser
 85 90 95
 Gly Ser Ala Ile Ala Tyr Glu Val Ala Arg Arg Met Asp Pro Ala Gln
 100 105 110
 Leu Arg Gly Leu Ala Val Ser Ser His Pro Ala Pro Gly Ser Val Arg
 115 120 125
 Arg Ala Pro Thr Trp Glu Leu Pro Arg Gln Glu Phe Leu Lys Gln Val
 130 135 140
 Val Thr Asp Gly Tyr Leu Pro Pro Glu Ile Ile Ala Asn Glu Glu Leu
 145 150 155 160

<400>	20															
Met	Arg	Lys	Lys	Gly	Trp	Ile	Arg	Thr	Leu	Thr	Thr	Val	Glu	Lys	Pro	
				5					10					15		
Thr	Gly	Arg	Leu	Val	Cys	Phe	Pro	His	Ser	Gly	Gly	Ser	Ala	Thr	Ala	
			20					25					30			
Phe	Ser	Asp	Trp	Ala	Asn	Ala	Met	Pro	Pro	Gly	Val	Glu	Leu	Val	Ala	
		35					40					45				
Val	Gln	Tyr	Pro	Gly	Arg	Gly	Asp	Arg	Phe	Ala	Glu	His	Leu	Val	Asp	
	50					55					60					
Asp	Val	Ala	Glu	Met	Ala	Gly	His	Ile	Val	Val	Glu	Leu	Leu	Glu	Leu	
65					70					75					80	
Pro	Cys	Arg	Asp	Gln	Val	Leu	Phe	Gly	His	Ser	Leu	Gly	Ala	Met	Val	
				85					90					95		
Ala	Tyr	Glu	Thr	Ala	Met	Leu	Leu	Arg	Asp	Met	Arg	Val	Glu	Pro	Arg	
			100					105					110			
Arg	Leu	Cys	Ala	Ser	Ala	Cys	Leu	Pro	Pro	Gly	Gln	Met	Lys	Asn	Arg	
		115					120					125				

Asp Val His Leu Lys Pro Asp Asp Glu Phe Trp Ala Ala Ile Cys Ala
 130 135 140
 Leu Gly Gly Ile Asp Pro Arg Val Ala Glu Ile Ala Glu Leu Arg Glu
 145 150 155 160
 Leu Phe Met Pro Thr Leu Arg Ser Asp Leu Arg Ala His Ala Asn Tyr
 165 170 175
 Arg Pro Gly Ser Gly Thr Thr Pro Leu Ser Cys Pro Val Ser Cys Tyr
 180 185 190
 His Gly Val Gly Asp Pro Leu Val Asp Glu Ser Arg Leu Thr Glu Trp
 195 200 205
 Ala Asp Val Thr Thr Gly Glu Phe Arg Leu Thr Val Arg Pro Gly Gly
 210 215 220
 His Phe His Val Ser Ile Asp Val Asp Gly Leu Val Ser Asp Ile Met
 225 230 235 240
 Gly Arg Glu Trp Ala
 245

<210> 21
 <211> 2594
 <212> PRT
 <213> *Micromonospora sp.*, cepa ML1

<220>
 <223> Proteína Tio20

<400> 21
 Met Gly Tyr Gly Gly Val Ser Ser Ala Gln Glu Ala Met Trp Leu Glu
 1 5 10 15
 Gln Asn Leu Ala Pro Glu Val Pro Asn Ser Thr Thr Thr Ile Trp Asp
 20 25 30
 Val Arg Gly Glu Leu His Leu Pro Asp Leu Glu Asp Ala Leu Arg Thr
 35 40 45
 Ala Leu Gly Glu Thr Ser Ala Leu Leu Val Asn Phe Arg Arg Glu Ser
 50 55 60
 Asp Glu Leu Arg Ile Val Pro Arg Lys Leu Asp Ala Trp Glu Pro Phe
 65 70 75 80
 Arg Ile Asp Ala Ser Asp Ala Ala Asp Pro Ala Ala Ala Ala Arg Glu
 85 90 95
 Phe Leu Thr Glu Gln Ile Arg Arg Pro Phe Asp Leu Gly Ala Asp Ala
 100 105 110
 Leu Phe Arg Ile Gly Thr Val Arg Leu Gly Asp Gly His His Leu Leu
 115 120 125

Cys Leu Ile Met His His Ile Leu Thr Asp Ala Phe Gly Val Phe Thr
 130 135 140
 Leu Leu Ser Gln Arg Val Ala Glu Ile Tyr Arg Ala Leu Arg Ala Asp
 145 150 155 160
 Ser Pro Val Pro Asp Arg Pro Asp Thr Pro Ala Glu Ile Ala Gly Glu
 165 170 175
 Arg Glu Ala Thr Tyr Arg Ser Ser Glu Arg Phe Thr Asp Ala Ala Arg
 180 185 190
 Phe Trp Arg Asp Tyr Val Thr Ala Gly Gly Thr Ala Ala Arg Leu Pro
 195 200 205
 Thr Gly Met Arg Ala Gly Ala Thr Arg Thr Ala Ser Val Gly Phe Trp
 210 215 220
 Asp Gly Leu Thr Ala Pro Leu Gly Val Ala Ser Tyr Thr Ala Ser Val
 225 230 235 240
 Ala Ala Ser Glu Val Ala Arg Trp Arg Ala Ala Ala Asp Ser Thr Glu
 245 250 255
 Thr Ser Val Pro Asp Leu Leu Thr Ser Ala Ala Thr Ala Phe Leu His
 260 265 270
 Arg Leu Cys Asn Thr Pro Arg Pro Met Leu Ser Phe Thr Val Asn Tyr
 275 280 285
 Arg Arg Gly Asp Ile Lys Gln Ala Cys Gly Leu Tyr Ser Asn Ala Leu
 290 295 300
 Pro Leu Val Ile Glu Val Pro Ala Ala Gly Asp Phe Val Asp Leu Ala
 305 310 315 320
 Lys Ser Val Arg Leu Glu Arg Leu Arg Val Leu Arg Tyr Ala Glu Tyr
 325 330 335
 Asn Val Ser Leu Ile Lys Arg Ala Ala Gly His Ala Gly Asp Val Arg
 340 345 350
 Ser Pro Phe Gly Pro Val Val Asn Val Ile Pro Phe Leu Arg Glu Leu
 355 360 365
 Asp Leu Ala Gly Ala Val Gly Arg Phe Ala Gly Gly Thr Phe Gly Leu
 370 375 380
 Pro Asp Glu Val Arg Ile Ser Thr Tyr Val Asp Gly Gly Pro Asp Ser
 385 390 395 400
 Asp Leu Tyr Ile Arg Phe Asp Ala Pro Gly Arg Leu Tyr Thr His Glu
 405 410 415

Asp	Ile	Ala	Gly	Leu	Ala	Glu	Arg	Phe	Val	Thr	Phe	Val	Arg	Ala	Ala	
			420					425					430			
Leu	Thr	Asp	Pro	Arg	Ala	Arg	Val	Arg	Asp	Ile	Ser	Ala	Ile	Gly	Glu	
		435					440					445				
Gln	Glu	Arg	Val	Ala	Leu	Leu	Asp	Arg	Cys	Ala	Gly	Pro	Val	Val	Glu	
	450					455					460					
Leu	Ala	Gly	Val	Thr	Val	Pro	Val	Leu	Leu	Asp	Met	Gln	Ala	Leu	Ala	
465					470					475					480	
Thr	Pro	Asp	Ala	Ala	Ala	Val	Val	His	Glu	Asn	Ala	Val	Leu	Thr	Tyr	
			485						490					495		
Arg	Glu	Leu	Ala	Ala	Arg	Ser	Asp	Ala	Leu	Ala	Arg	Arg	Leu	Ala	Glu	
			500					505					510			
Cys	Gly	Ala	Gly	Pro	Glu	Arg	Leu	Val	Ala	Val	Ala	Val	Pro	Arg	Ser	
		515					520					525				
Ile	Gly	Leu	Ala	Val	Ala	Leu	Cys	Ala	Val	Leu	Lys	Ala	Gly	Ala	Ala	
	530					535					540					
Tyr	Leu	Pro	Ile	Asp	Pro	Asp	Ser	Pro	Ala	Glu	Arg	Val	Glu	Phe	Met	
545					550					555					560	
Leu	Gly	Asp	Ala	Arg	Pro	Thr	Leu	Leu	Leu	Thr	Val	Ala	Gly	Thr	Asp	
				565					570					575		
Gly	Leu	Ala	Ile	Thr	Pro	Leu	His	Gly	Ala	Gln	Pro	Arg	Trp	Glu	Ala	
			580					585					590			
Ser	Val	Ile	Glu	Arg	Thr	Thr	Ala	Ala	Ser	Pro	Ala	Ala	Thr	Ser	Gln	
		595					600					605				
Ala	Gly	Gly	Thr	Gly	Glu	His	Ala	Ala	Tyr	Leu	Arg	Tyr	Arg	Ser	Gly	
	610					615					620					
Pro	Phe	Gly	Pro	Ala	Ala	Gly	Val	Val	Val	Thr	His	Ala	Ala	Leu	Val	
625					630					635					640	
Asn	Glu	Leu	Leu	Trp	Met	Gln	Gln	Arg	Tyr	Arg	Leu	Asn	Asp	Thr	Asp	
				645					650					655		
Arg	Val	Leu	His	Lys	Ala	Pro	Ala	Ser	Leu	Asn	Ala	Ser	Gly	Trp	Glu	
			660					665					670			
Ile	Phe	Trp	Pro	Leu	Ile	Ala	Gly	Ala	Thr	Val	Val	Met	Ala	Thr	Gln	
		675					680					685				
Asp	Val	His	Gly	Asp	Ala	Thr	Ala	Leu	Ala	Ala	Met	Ile	Arg	Arg	Thr	
	690					695					700					
Gly	Val	Thr	Thr	Val	His	Val	Val	Pro	Ser	Val	Leu	Ala	Glu	Leu	Val	
705					710					715					720	

Arg	Asp	Glu	Ala	Val	Val	Thr	Ala	Gly	Lys	Gly	Leu	Arg	His	Val	Met
				725					730					735	
Cys	Ser	Gly	Glu	Val	Leu	Pro	Ala	Asp	Leu	Ala	Ala	Arg	Phe	His	Glu
			740					745					750		
Ser	Cys	Asp	Val	Pro	Val	His	Ser	Leu	Tyr	Gly	Pro	Val	Glu	Ala	Met
		755					760					765			
Pro	Val	Val	Thr	Ser	Cys	Asp	Tyr	Arg	Pro	Gly	Ser	Pro	Thr	Val	Pro
	670					675					680				
Leu	Gly	Arg	Pro	Ile	Ser	Asn	Ala	Gly	Val	Tyr	Val	Leu	Asp	Ser	Gln
685					690					695					700
Leu	Gln	Leu	Thr	Pro	Pro	Gly	Val	Val	Gly	Glu	Val	Tyr	Val	Thr	Gly
				705					710					715	
Val	Gly	Leu	Ala	Arg	Gly	Tyr	Leu	His	Arg	Pro	Ala	Ala	Thr	Ala	Asp
			720					725					730		
His	Phe	Val	Ala	Cys	Pro	Phe	Gly	Pro	Ala	Gly	Thr	Arg	Met	Tyr	Arg
		735					740					745			
Thr	Gly	Asp	Leu	Ala	Arg	Trp	Thr	Asp	Asp	Gly	Glu	Leu	Glu	His	Ala
	750					755					760				
Gly	Ser	Ala	Asp	Gln	Arg	Val	Arg	Ile	Arg	Gly	Phe	Trp	Ala	Ala	Phe
765					770					775					780
Leu	Glu	Ile	Glu	Ala	Val	Leu	Thr	Ala	His	Pro	Gly	Val	Asp	Gln	Ala
				785					790					795	
Ala	Val	Ile	Ala	Arg	Pro	Asp	Gly	Pro	Asp	Arg	Ala	Ala	Gln	Leu	Ile
			800					805					810		
Ala	Tyr	Val	Val	Pro	Ala	Thr	Gly	Gly	Arg	Met	Gly	Ala	Ala	Gly	Leu
		815					820					825			
Asp	Val	Asp	Phe	Ser	Ala	Gly	Leu	Asp	Val	Val	Glu	Leu	Arg	Arg	Phe
	830					835					840				
Val	Ala	Ala	Arg	Leu	Pro	Glu	His	Leu	Val	Pro	Ala	Thr	Ile	Val	Val
845					850					855					860
Leu	Asp	Arg	Leu	Pro	Leu	Thr	Ala	Glu	Gly	Lys	Arg	Asp	His	Ala	Ala
				865					870					875	
Leu	Pro	Glu	Pro	Gln	Ala	Thr	Arg	Arg	Ala	His	Arg	Val	Ala	Arg	Thr
			880					885					890		
Ala	Glu	Glu	Arg	Leu	Leu	Ala	Asp	Ala	Phe	Ala	Glu	Val	Leu	Gly	Ala
		895					900					905			

Gly Arg Ile Gly Leu Asp Asp Asp Phe Phe Ala Leu Gly Gly Asp Ser
 910 915 920
 Ile Ser Ala Met Arg Val Val Ala Leu Ala Arg Ala Gly Gly Leu Ala
 925 930 935 940
 Leu Thr Ala Arg Thr Val Phe Glu Cys Arg Thr Val Ala Ala Leu Ala
 945 950 955
 Ala Val Ala Ala Pro Ala Asp Glu Pro Ala Val Ala Ala Ala Ala Glu
 960 965 970
 Gly Asp Gly Gly Arg Asp Asp Ala Gly Pro Ile Pro Leu Pro Pro Met
 975 980 985
 Ala Arg Leu Phe Ala Glu Arg Gly Pro Gly Leu Asp Arg Leu Ala Gln
 990 995 1000
 Trp Leu Val Leu Thr Leu Pro Ala Glu Met Ser Pro Asp Leu Leu Thr
 1005 1110 1115 1120
 Gly Ala Val Arg Ala Val Leu Asp Arg His Asp Val Leu Arg Ser Arg
 1125 1130 1135
 Leu Ala Asp Gly Gln Leu Leu Ile Gln Pro Pro Gly Ser Val Asp Val
 1140 1145 1150
 Asp Lys Leu Ile Arg Gln Val Pro Cys Ser Gly Asp Trp Ser Gly Pro
 1155 1160 1665
 Ser Trp Pro Ala Leu Leu Ala Asp Glu Ala Ala Lys Ala Val Gly Gln
 1170 1175 1180
 Leu Ser Ala Ser Ala Gly Gln Val Ile Arg Met Val Trp Phe Ala Pro
 1185 1190 1195 1200
 Pro Gly Gly Ala Pro Gly Arg Leu Leu Ile Ala Ala His His Leu Val
 1205 1210 1215
 Val Asp Gly Met Ser Trp Arg Val Leu Val Pro Asp Leu Ala Glu Ala
 1220 1225 1230
 Ala Ala Gln Ile Arg Ser Gly Arg Thr Pro Ala Leu Ala Pro Val Gly
 1235 1240 1245
 Thr Ser Leu Arg Ser Trp Leu Asp Glu Leu Ala Ala Glu Ala Val Arg
 1250 1255 1260
 Pro Glu Arg Val Ala Glu Leu Glu Leu Trp Arg Gly Ile Leu Ala Pro
 1265 1270 1275 1280
 Ser Gly Asp Pro Leu Gly Ala Arg Pro Leu Asp Pro Ala Val Asp Leu
 1285 1290 1295
 Thr Arg Thr Ala Gln Thr Leu Arg Val Glu Val Pro Ala Glu Leu Ser
 1300 1305 1310

Arg Ala Leu Leu Thr Ser Val Pro Ala Thr Phe Arg Cys Gly Val Asp
 1315 1320 1325
 Asp Val Leu Leu Thr Ala Leu Ala Leu Ala Leu Ala Ser Arg Arg His
 1330 1335 1340
 Asn Ser Ala Gly Val Val Arg Leu Glu Gly His Gly Arg Gln Glu Glu
 1345 1350 1355 1360
 Leu Val Pro Gly Thr Asp Leu Thr Arg Thr Val Gly Trp Phe Thr Thr
 1365 1370 1375
 Val His Pro Val Arg Phe Asp Leu Ser Gly Leu Asp Leu Thr Asp Ala
 1380 1385 1390
 Ile Asn Gly Gly Pro Ala Ala Ala Glu Ala Leu Arg Arg Val Lys Glu
 1395 1400 1405
 Ser Arg Arg Val Leu Pro Asp Arg Gly Ile Gly Tyr Ser Leu Leu Arg
 1410 1415 1420
 Tyr Leu Asn Gln Ala Thr Gly Ala Met Leu Arg Asp Leu Pro Ala Asp
 1425 1430 1435 1440
 Glu Val Gly Phe Asn Tyr Leu Gly Arg Phe Ser Phe Asp Arg Ser Arg
 1445 1450 1455
 Thr Gly Arg Glu Met Ala Trp Thr Ser Ala Pro Glu Ser Ala Ala Leu
 1460 1465 1470
 Val Ala Ala Pro Asp Pro Asp Met Pro Leu Ile Ser Ala Leu Glu Leu
 1475 1480 1485
 Asn Ser Leu Val Val Asp Thr Asp His Gly Glu Gln Leu Gln Ala Leu
 1490 1495 1500
 Phe Thr Phe Ala Thr Gly Val Leu Ala Thr Asp Glu Val Arg His Leu
 1505 1510 1515 1520
 Ala Asp Asp Trp Gln Ala Ala Leu Arg Gly Ile Arg Ala Gln Ala Ala
 1525 1530 1535
 Thr Gly Ala Ser Gly Leu Thr Pro Ser Asp Val Pro Leu Val Arg Val
 1540 1545 1550
 Arg Gln Gln Asp Leu Asp Asp Trp Gln Ser Arg Phe Gly Arg Val Asp
 1555 1560 1565
 Asp Val Trp Pro Leu Thr Pro Leu Gln His Gly Leu Leu Tyr His Ser
 1570 1575 1580
 Leu Leu Gly Asp Ser Ala Ala Ala Ser Tyr Gln Thr Gln Phe Val Leu
 1585 1590 1595 1600

Arg Ile Asn Gly Pro Val Asp Pro Ala Arg Leu Arg Gly Ala Ala Gln
 1605 1610 1615
 Thr Leu Leu Asn Arg His Pro Asn Leu Arg Val Ala Phe Leu Thr Lys
 1620 1625 1630
 Ser Thr Gly Glu Pro Val Gln Val Val Val Asp Gly Val Glu Val Pro
 1635 1640 1645
 Phe Ala His Val Asp Leu Ala Gly Gln Pro Gln Ala Glu Glu Arg Leu
 1650 1655 1660
 Ala Arg Ile Leu Ala Glu Glu Arg Glu Ala Gln Phe Arg His Asp Ser
 1665 1670 1675 1680
 Ala Pro Leu Leu Arg Phe Ser Leu Ile Thr Leu Thr Pro Thr Gln Ala
 1685 1690 1695
 Glu Leu Ala Leu Thr Ala His His Val Leu Phe Asp Gly Trp Ser Leu
 1700 1705 1710
 Pro Leu Ile Glu Gln Gln Leu Met His Leu Tyr Ala Gly Gly Ala Asp
 1715 1720 1725
 Ala Leu Glu Leu Val Glu His Gly Tyr Arg Glu Phe Leu Arg Trp His
 1730 1735 1740
 Ser Arg Gln Asn Ala Asp Glu Ala Leu Ala Ala Trp Thr Glu Glu Leu
 1745 1750 1755 1760
 Ala Asp Ala Thr Gln Pro Thr Leu Leu Ala Ser Gly Phe Ala Pro Ala
 1765 1770 1775
 Glu Gln Ala Gly Thr Ser Thr Glu Ile Gly Gln Val Asp Val Leu Leu
 1780 1785 1790
 Thr Arg Gln Glu Ala Ala Ser Ala Val Arg Arg Ala Ala Asp Cys Gly
 1795 1800 1805
 Ile Thr Pro Asn Val Leu Val Gln Gly Ala Trp Ala Val Val Leu Ala
 1810 1815 1820
 Gln Leu Thr Gly Asn Ala Asp Val Leu Phe Gly Ser Ser Val Thr Val
 1825 1830 1835 1840
 Arg Pro Pro Glu Leu Pro Gly Ala His Thr Val Val Gly Met Phe Thr
 1845 1850 1855
 Asn Thr Val Pro Val Arg Val Arg Cys Ala Pro Ala Asp Ser Leu Glu
 1860 1865 1870
 Gly Met Leu Gln Ala Val Gln Asp Ala Gln Ala Arg Thr Leu Asp His
 1875 1880 1885
 Tyr His Val Gly Leu Gly Asp Ile Gln Arg Ala Thr Gly Leu Thr Thr
 1890 1895 1900

Leu Phe Asp Thr Ile Val Ile Phe Glu Ser Phe Pro Val Asp Arg Ala
 1905 1910 1915 1920
 Ala Leu Ser Lys Ala Ser Ala Ser Ala Glu Leu Thr Val Thr Gly Ile
 1925 1930 1935
 Arg Pro Phe Ala Pro Thr His Tyr Pro Leu Thr Val Leu Ala Ala Ala
 1940 1945 1950
 Asp Pro Leu Leu Ser Leu Thr Leu Gln Tyr His Pro Asp Ala Leu Ala
 1955 1960 1965
 Pro Ser Thr Val Gln Ala Ile Ala Asp Arg Phe Ala Arg Val Leu Arg
 1970 1975 1980
 Gln Phe Ala Ala Asp Pro Arg Thr Arg Ile Ala Ala Ile Asp Val Leu
 1985 1990 1995 2000
 Ser Glu Ala Glu Arg Gln Leu Val Val His Asp Trp Asn Ala Thr Ala
 2005 2010 2015
 Leu Glu Val Pro Pro Glu Phe Val Pro Glu Gln Phe Ala Arg Gln Ala
 2020 2025 2030
 Ala Ala Thr Pro Asp Ala Val Ala Val Val Val Gly Glu Gln His Leu
 2035 2040 2045
 Thr Tyr Ala Glu Leu Ser Glu Arg Ala Asn Arg Met Ala His Trp Leu
 2050 2055 2060
 Ile Ala Gln Gly Ile Gly Ala Glu Ser Arg Val Val Val Leu Leu Pro
 2065 2070 2075 2080
 Arg Ser Ala Asp Leu Val Ala Ala Leu Leu Ala Val Trp Gln Ala Gly
 2085 2090 2095
 Gly Cys Tyr Val Pro Val Asp Pro Asp Tyr Pro Ala Ala Arg Val Arg
 2100 2105 2110
 Ser Val Ile Asp Asp Cys Ala Ala Thr Leu Val Leu Asp Glu Asn Leu
 2115 2120 2125
 Leu Asp Arg Thr Asp Leu Ser Arg Tyr Pro Ala His Ala Pro Ala Ile
 2130 2135 2140
 Ala Val Ala Pro Gly Gln Ala Ala Tyr Thr Ile Tyr Thr Ser Gly Ser
 2145 2150 2155 2160
 Thr Gly Lys Pro Lys Gly Val Val Ile Gly His Asp Ala Leu Thr Asn
 2165 2170 2175
 Phe Leu Ala Ser Met Gln Arg Thr Leu Arg Leu Ser Ala Gln Asp Arg
 2180 2185 2190

Phe Ala Ala Val Thr Thr Ile Ala Phe Asp Ile Ala Ala Leu Glu Leu
 2195 2200 2205
 Phe Leu Pro Leu Thr Thr Gly Ala Arg Val Val Leu Ala Asn Arg Glu
 2210 2215 2220
 His Val Ala Asn Pro Pro Ala Met Leu Asp Leu Val Glu Gln Thr Gly
 2225 2230 2235 2240
 Val Thr Val Met Gln Ala Thr Pro Ala Leu Trp Gln Met Leu Ala Thr
 2245 2250 2255
 His Asp Val Arg Arg Leu Ser Gly Val Arg Val Leu Thr Gly Gly Glu
 2260 2265 2270
 Ala Leu Pro Arg Pro Leu Ala Glu Gly Leu Cys Ala His Ala Ala Gln
 2275 2280 2285
 Val Ile Asn Leu Tyr Gly Pro Thr Glu Thr Thr Val Trp Ser Thr Leu
 2290 2295 2300
 Ala Asp Val Ala Thr Cys Lys Pro Pro Ser Ile Gly Thr Pro Ile Ala
 2305 2310 2315 2320
 Asn Thr Gln Val Leu Ile Leu Asp Ser Ala Leu Arg Pro Val Pro Pro
 2325 2330 2335
 Gly Val Gly Gly Asp Leu Trp Ile Ala Gly Asp Gly Leu Ala Arg Gly
 2340 2345 2350
 Tyr His Arg Gln Pro Gly Leu Thr Ala Ser Arg Phe Val Ala Asn Pro
 2355 2360 2365
 Phe Gly Pro Gln Gly Ala Arg Met Tyr Arg Thr Gly Asp Leu Ala Arg
 2370 2375 2380
 Trp Ser Glu Ser Gly Arg Ile Glu Phe Leu Gly Arg Ala Asp Phe Gln
 2385 2390 2395 2400
 Ile Lys Leu Arg Gly Phe Arg Ile Glu Pro Gly Asp Val Glu His Ala
 2405 2410 2415
 Leu Thr Arg His Pro Ala Val Arg Glu Ala Val Val Thr Ile Arg Glu
 2420 2425 2430
 Asp Gln Pro Asp Asp Lys Arg Leu Val Gly Tyr Val Val Val Ala Glu
 2435 2440 2445
 Asp Ala Glu Thr Pro Pro Thr Arg Glu Leu Gln Gln Phe Val Arg Glu
 2450 2455 2460
 Gln Leu Pro Asp Tyr Met Val Pro Ser Ala Val Val Thr Leu Pro Ala
 2465 2470 2475 2480
 Ile Pro Leu Thr Ala Asn Gly Lys Leu Asp Arg Ser Arg Leu Pro Arg
 2485 2490 2495

Pro Glu Ala Ser Ser Thr Lys Tyr Arg Ala Pro Glu Ser Ala Arg Glu
 2500 2505 2510

Val Ala Leu Cys Gly Ile Phe Ala Glu Val Leu Gly Ala Glu Arg Val
 2515 2520 2525

Gly Val Asp Asp Asp Phe Phe Ala Leu Gly Gly His Ser Leu Leu Ala
 2530 2535 2540

Thr Arg Leu Ala Ala Arg Leu Arg Ser Asp Leu Gly Val Glu Val Pro
 2545 2550 2555 2560

Ile Arg Thr Ile Phe Ala Ala Pro Thr Val Ala Glu Leu Ala Arg Arg
 2565 2570 2575

Trp Ser Gly Leu Ser Thr Ser Val Arg Lys Pro Leu Arg Arg Met Thr
 2580 2585 2590

Glu Arg

<210> 22
 <211> 3140
 <212> PRT
 <213> *Micromonospora* sp., cepa ML1

<220>
 <223> Proteína Tio21

<400> 22
 Met Ala Pro Val Ser Tyr Ala Gln Gln Arg Met Trp Phe Ile Asn Arg
 1 5 10 15

Phe Glu Gly Gln Thr Ala Thr Tyr Asn Ile Pro Ile Leu Ile Arg Met
 20 25 30

Arg Gly Gln Ile Asp Arg Asp Glu Leu Arg Ala Ala Phe Met Asp Val
 35 50 55

Leu Ser Arg His Glu Ile Leu Arg Thr Val Tyr Asp Glu Leu Asp Gly
 50 55 60

Gln Pro Leu Gln Arg Ile Ile Glu Pro Ala Glu Leu Ser Leu Pro Trp
 65 70 75 80

Thr Asp Trp Gly Trp Val Ser Pro Gln Gln Val Asp Ser Ile Val Ala
 85 90 95

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

Ala Tyr Leu Leu Arg Pro Glu Pro Asp Val Ala Leu Leu Ala Val Val
 115 120 125

Ile	His	His	Ile	Ala	Gly	Asp	Gly	Gly	Ser	Leu	Gly	Pro	Phe	Thr	Arg	130	135	140
Asp	Leu	Ser	Ala	Ala	Tyr	Arg	Ala	Arg	Leu	Ala	Gly	Val	Ala	Pro	Thr	145	150	155
Trp	Asp	Glu	Leu	Pro	Val	Gln	Tyr	Ser	Asp	Tyr	Ala	Leu	Trp	Gln	Arg	165	170	175
Glu	Leu	Leu	Gly	Asp	Glu	Ser	Asp	Pro	Asp	Ser	Val	Phe	Ser	Thr	Gln	180	185	190
Leu	Arg	Tyr	Trp	Arg	Gln	Glu	Leu	Asp	Gly	Val	Ser	Arg	Pro	Val	Ala	195	200	205
Leu	Pro	Val	Asp	Arg	Pro	Arg	Pro	Ala	Arg	Ala	Ser	Tyr	Arg	Gly	Asp	210	215	220
Asn	His	Gly	Phe	Glu	Ile	Pro	Ala	Ala	Leu	Phe	Glu	Lys	Val	Glu	Arg	225	230	235
Leu	Ala	Gln	Ser	Tyr	Asn	Val	Thr	Val	Pro	Met	Val	Phe	Gln	Thr	Ala	245	250	255
Leu	Val	Val	Leu	Leu	His	Arg	Leu	Gly	Gly	Asp	Arg	Asp	Ile	Thr	Leu	260	265	270
Gly	Ser	Pro	Ile	Ala	Gly	Arg	Thr	Asp	Gln	Ala	Leu	Glu	Asn	Leu	Val	275	280	285
Gly	Phe	Phe	Val	Asn	Thr	Trp	Val	Leu	Arg	Val	Gly	Val	Cys	Thr	Gly	290	295	300
Asp	Ser	Phe	Ala	Gln	Val	Leu	Glu	Gln	Val	Arg	Gln	Lys	Ala	Ile	Ala	305	310	315
Ala	Tyr	Asp	His	Gln	Asp	Leu	Pro	Phe	Glu	Arg	Leu	Val	Glu	Leu	Leu	325	330	335
Arg	Pro	Gly	Arg	Ser	Val	Ala	His	His	Pro	Leu	Phe	Gln	Val	Thr	Leu	340	345	350
Ala	Trp	Gln	Asn	Ser	Val	Arg	Gln	Glu	Leu	Asp	Leu	Pro	Gly	Val	Thr	355	360	365
Ala	Ser	Met	Glu	Pro	Val	Ala	Thr	Gly	Thr	Ala	Lys	Phe	Asp	Leu	Phe	370	375	380
Val	Asn	Leu	Met	Pro	Asn	His	Ser	Arg	Arg	Ser	Val	Thr	Val	Asp	Leu	385	390	395
Glu	Tyr	Ala	Thr	Asp	Leu	Phe	Asp	Pro	Ala	Thr	Val	Glu	Ala	Ile	Ala	405	410	415
Ala	Arg	Tyr	Leu	Arg	Val	Leu	Glu	Gln	Val	Val	Glu	Ser	Pro	Asp	Leu	420	425	430

Gln	Leu	Ala	Pro	Leu	Leu	Pro	Gln	Thr	Ala	Gly	Ser	Pro	Gly	Val	Gln	
				725					730					735		
Asp	Leu	Arg	Leu	Val	Phe	Phe	Ser	Gly	Asp	Tyr	Thr	Pro	Leu	Ser	Leu	
			740					745					750			
Pro	Asp	Gln	Val	Arg	Ala	Val	Phe	Pro	His	Ala	Glu	Ile	Ile	Ser	Leu	
		755					760					765				
Gly	Gly	Ala	Thr	Glu	Ala	Thr	Val	Trp	Ser	Asn	Phe	Phe	Pro	Val	Gly	
	770					775					780					
Ala	Ile	Asp	Pro	Gln	Trp	Arg	Ser	Ile	Pro	Tyr	Gly	Arg	Pro	Ile	Asp	
785					790					795					800	
Asn	Ala	Arg	Tyr	Tyr	Val	Leu	Asp	Glu	Lys	Leu	Glu	Pro	Cys	Ala	Val	
				805					810					815		
Gly	Val	Glu	Gly	Asp	Leu	Tyr	Ile	Gly	Gly	Pro	Val	Ile	Ala	Leu	Gly	
			820					825					830			
Tyr	Val	Asn	Arg	Pro	Asp	Leu	Thr	Ala	Glu	Arg	Phe	Ile	Ala	Asp	Pro	
		835					840					845				
Phe	Gly	Gly	Val	Pro	Gly	Gln	Arg	Ile	Tyr	Arg	Thr	Gly	Asp	Arg	Ala	
	850					855					860					
Ser	Phe	Phe	Pro	Asp	Gly	Asn	Ile	Cys	Phe	Leu	Gly	Arg	Ala	Asp	Gly	
865					870					875					880	
Gln	Val	Lys	Val	Arg	Gly	Phe	Arg	Ile	Glu	Leu	Gly	Glu	Ile	Glu	Tyr	
				885					890					895		
Ala	Leu	Gly	Arg	His	Pro	Ala	Val	Arg	Gln	Gly	Ile	Ala	Ile	Thr	Arg	
			900					905					910			
Arg	Asp	Ser	Val	Gly	Asp	Leu	Arg	Leu	Val	Ser	Tyr	Val	Leu	Pro	Asp	
		915					920					925				
Leu	Asp	Thr	Val	Glu	Ser	Val	Val	Ala	Ala	Asp	Glu	Gln	Val	Arg	Glu	
	930					935					940					
Trp	Gln	Glu	Ile	Tyr	Asp	Gln	Gly	Tyr	Leu	Glu	Val	Thr	Asp	Gln	Asp	
945					950					955					960	
Phe	Gly	Asp	Asp	Phe	Asn	Leu	Trp	Val	Ser	Ser	Tyr	Thr	Gly	Glu	Pro	
				965					970					975		
Ile	Pro	Val	Gly	Gln	Met	Arg	Glu	Trp	Gln	Asp	Ala	Ala	Val	Asp	Arg	
			980					985					990			
Ile	Leu	Ser	Phe	Thr	Pro	Arg	Arg	Val	Leu	Glu	Val	Gly	Ala	Gly	Thr	
		995					1000					1005				
Gly	Leu	Leu	Leu	Ala	Arg	Val	Ala	Gly	Ser	Val	Glu	Ala	Tyr	Trp	Ala	
	1010					1015					1020					

Thr Asp Phe Ser Glu Pro Val Ile Glu Arg Leu Gly Arg Gln Val Thr
 1025 1030 1035 1040
 Glu Ala Gly Trp Ala Glu Arg Val Arg Leu Leu Cys Arg Arg Ala Asp
 1045 1050 1055
 Asp Leu Asp Gly Ile Pro Arg Ile Phe Asp Thr Val Val Leu Asn Ser
 1060 1065 1070
 Val Val Gln Tyr Phe Pro Asn Glu Arg Tyr Leu Glu Gln Val Leu Asp
 1075 1080 1085
 Gly Val Trp Ala Met Leu Glu Pro Gly Gly Arg Leu Val Leu Gly Asp
 1090 1095 1100
 Ile Arg Arg Ala Arg Ser Leu Arg Ala Phe Gln Val Ala Val Gln Gln
 1105 1110 1115 1120
 Ala Lys His Gly Asn Leu Pro Pro Ala Gln Leu Arg Ser Ala Val Glu
 1125 1130 1135
 Gln Gly Leu Leu Leu Glu Lys Glu Leu Val Ile Asp Pro Glu Trp Phe
 1140 1145 1150
 Gln Arg Trp Ala Glu Arg Ala Gly Ala Ala Gly Val Asp Val Arg Leu
 1155 1160 1165
 Lys Glu Gly Ala Phe Gln Asn Glu Leu Thr Arg His Arg Tyr Glu Ile
 1170 1175 1180
 Val Val His Lys Pro Gly Thr Thr Glu Ala Gly Arg Pro Tyr Ala Val
 1185 1190 1195 1200
 Asp Thr Val Pro Arg Leu Glu Trp Asn Gly Asp Leu Asp Gly Leu Ala
 1205 1210 1215
 Glu Arg Ile Arg Thr Leu Gly Gly Pro Val Val Arg Ile Ala Gly Ile
 1220 1225 1230
 Pro Asn Ala Arg Val Ala Gln Glu Val Ala Ala Ala Arg Asp Leu Gly
 1235 1240 1245
 Leu Glu Glu Ser Glu Pro Pro Thr Ser Val Pro Val Asp Pro His Glu
 1250 1255 1260
 Met Ala Thr Trp Ala Ala Arg Gln Gly Trp Ser Ile Ala Leu Thr Trp
 1265 1270 1275 1280
 Ser Gly Ser Ala Val Asp Gln Phe Glu Ala Val Leu Phe Thr Asp Ala
 1285 1290 1295
 Thr Thr Glu His Arg Ala Leu Ser Gly Thr Tyr Leu Pro Val Ala Ala
 1300 1305 1310

Ala Gly Thr Trp Ile Asn Asp Pro Ala Ala Ala Ala Gly Ile Thr Ala
 1315 1320 1325
 Leu Pro Ala Val Leu Arg Glu Tyr Leu Ser Lys Thr Leu Pro Glu Tyr
 1330 1335 1340
 Met Val Pro Ser Ala Ile Val Pro Ile Gly Arg Val Pro Leu Thr Pro
 1345 1350 1355 1360
 Asn Gly Lys Leu Asp Arg Ala Ala Leu Pro Val Pro Glu Tyr Ala Ala
 1365 1370 1375
 Ala Ser Arg Gly Arg Pro Pro Cys Thr Ser Glu Glu Glu Ser Leu Cys
 1380 1385 1390
 Glu Ile Phe Ala Gly Val Leu Gly Leu Asp Arg Val Gly Val Asp Asp
 1395 1400 1405
 Asn Phe Phe Thr Ile Gly Gly His Ser Leu Leu Ala Thr Arg Val Val
 1410 1415 1420
 Ser Arg Ile Arg Ala Ala His Gly Val Glu Ile Pro Ile Arg Thr Ile
 1425 1430 1435 1440
 Phe Glu Ala Pro Thr Val Ala Glu Leu Val Ala His Leu Thr Gly Gln
 1445 1450 1455
 Arg Thr Val Arg Pro Pro Leu Arg Pro Gln Pro Arg Pro Asp Arg Leu
 1460 1465 1470
 Pro Val Ser Phe Ala Gln Gln Arg Leu Trp Phe Ile His Arg Phe Glu
 1475 1480 1485
 Gly Pro Ser Pro Thr Tyr Ile Ile Pro Leu Ser Met Arg Leu Arg Gly
 1490 1495 1500
 Lys Leu Asp Val Pro Ala Leu Arg Gln Ala Val His Asp Val Val Val
 1505 1510 1515 1520
 Arg His Glu Ser Leu Arg Thr Val Phe Asp Glu Val Asp Gly Ala Pro
 1525 1530 1535
 Tyr Gln Arg Val Leu Asp Pro Asp Gly Leu Ala Ile Pro Trp Gln Glu
 1540 1545 1550
 His Glu Val Thr Glu Ser Ala Leu Gln Glu Ala Leu Arg Leu Glu Ala
 1555 1560 1565
 Arg Arg Pro Phe Asp Leu Ala Ala Glu Ile Pro Val Arg Ala Ala Leu
 1570 1575 1580
 Phe Arg Val Gly Glu His Asp Ala Val Leu Leu Leu Pro Leu His His
 1585 1590 1595 1600
 Ile Ala Ala Asp Gly Trp Ser Leu Gly Pro Leu Ala Glu Asp Leu Thr
 1605 1610 1615

Ala Ala Tyr Thr Ala Arg Cys Asp Gly Gln Arg Pro Gln Trp Ala Ala
 1620 1625 1630
 Leu Pro Val Gln Tyr Ala Asp Tyr Thr Leu Trp Gln Arg Asp Leu Leu
 1635 1640 1645
 Gly Gly Asp Asp Pro His Ser Leu Tyr Arg Arg Gln Leu Asp Tyr Trp
 1650 1655 1660
 Thr Gly Gln Leu Ala Gly Leu Pro Glu Ser Leu Gln Leu Pro Thr Asp
 1665 1670 1675 1680
 Arg Ala Arg Pro Ala Val Ala Ser Phe Ala Gly Glu Val Leu Pro Val
 1685 1690 1695
 Glu Leu Asp Glu Ser Leu Thr Glu Asp Leu Arg Lys Leu Ala Leu Arg
 1700 1705 1710
 Thr Gly Thr Thr Val Ser Met Val Leu Gln Ala Gly Leu Ala Gly Leu
 1715 1720 1725
 Leu Ser Lys Leu Gly Ala Gly Asn Asp Ile Pro Ile Gly Ser Pro Ile
 1730 1735 1740
 Ala Gly Arg Thr Asp Glu Ala Leu Asn Arg Leu Val Gly Phe Phe Val
 1745 1750 1755 1760
 Asn Thr Trp Val Ala Arg Val Asp Thr Ser Gly Ser Pro Ser Leu Ile
 1765 1770 1775
 Glu Leu Val Asp Arg Val Arg Glu Val Ser Leu Ala Ala Tyr Asp His
 1780 1785 1790
 Gln Asp Ile Pro Phe Glu His Leu Val Glu Ala Leu Asn Pro Val Arg
 1795 1800 1805
 Ser Thr Ala His His Pro Leu Phe Gln Val Cys Leu Ala Leu Gln Asn
 1810 1815 1820
 Asn Val Gln Pro Ala Phe Asp Leu Pro Gly Leu Lys Val Ser His Glu
 1825 1830 1835 1840
 Pro Val Asp Met Gly Val Ser Arg Phe Asp Leu Phe Leu Asn Leu Met
 1845 1850 1855
 Glu Glu Thr Thr Gly Glu Gly Thr Ser Arg Ile Ala Gly Val Val Glu
 1860 1865 1870
 Tyr Ala Thr Glu Leu Phe Asp Arg Asp Thr Val Thr Gly Leu Leu Glu
 1875 1880 1885
 Arg Trp Arg Arg Leu Leu Arg Gln Trp Val Gly Ala Pro Asp Ala Pro
 1890 1895 1900

Ile Asp Ala Val Glu Ile Leu Asn Arg Asp Asp Leu Ala Ala Leu Glu
 1905 1910 1915 1920
 Arg Trp Thr Gly Arg Ala Arg Gly Thr Asp Arg Val Val Gly Thr Ile
 1925 1930 1935
 Pro Glu Arg Phe Ala Ala Val Val Ala Glu Gln Pro Glu Ala Val Ala
 1940 1945 1950
 Leu Val Ala Ala Asp Gly Glu Glu Ser Trp Thr Tyr Gly Glu Leu Asp
 1955 1960 1965
 Arg Trp Ala Asn Arg Ile Ala His His Leu His Ala Arg Gly Val Gly
 1970 1975 1980
 Arg Gln His Arg Val Ala Leu Val Met Glu Arg Ser Pro Leu Leu Val
 1985 1990 1995 2000
 Ala Ala Val Leu Gly Thr Leu Lys Ala Gly Ala Cys Tyr Val Pro Val
 2005 2010 2015
 Glu Pro Thr Trp Pro Arg Ala Arg Ile Asp Leu Val Leu Ala Asp Leu
 2020 2025 2030
 Asp Pro Ala Leu Val Ile Asp Glu Arg Leu Ala Glu Glu Asp Leu Thr
 2035 2040 2045
 Gly Tyr Pro Thr Arg Pro Leu Asp Thr Ala Asp Val Gly Gly Glu His
 2050 2055 2060
 Leu Ala Tyr Leu Met Tyr Thr Ser Gly Ser Thr Gly Thr Pro Lys Gly
 2065 2070 2075 2080
 Val Glu Val Ser His Arg Asn Val Leu Ser Leu Ala Leu Asp Pro Cys
 2085 2090 2095
 Trp Ala Asp Ala Asp His Gln Arg Val Leu Val His Ala Pro Pro Thr
 2100 2105 2110
 Phe Asp Ala Ser Thr Tyr Glu Met Trp Val Pro Leu Leu His Gly Gly
 2115 2120 2125
 Ala Ala Val Val Ala Pro Pro Gly Lys Leu Asp Ala Ala Arg Leu Ala
 2130 2135 2140
 Thr Leu Ile Ala Glu Arg Gly Val Thr Ala Leu Trp Leu Pro Ala Gly
 2145 2150 2155 2160
 Leu Phe Asp Leu Ile Thr Gln His His Pro Lys Ser Phe Val Gln Val
 2165 2170 2175
 Arg Glu Val Trp Ala Gly Gly Asp Val Leu Ser Pro Ala Ala Val Arg
 2180 2185 2190
 Arg Leu Val Arg Asp Asp Gly Thr Leu Thr Val Val Asn Gly Tyr Gly
 2195 2200 2205

Pro Thr Glu Thr Thr Thr Phe Ala Ala Arg Tyr Arg Met Ser Ala Pro
 2210 2215 2220
 Ala Arg Cys Lys Asp Pro Leu Pro Ile Gly Glu Pro Met Ala Gly Ser
 2225 2230 2235 2240
 Arg Leu Tyr Ala Leu Asp Asp Arg Leu Arg Gln Val Pro Gln Gly Val
 2245 2250 2255
 Ile Gly Glu Leu Tyr Val Gly Gly Asp Gly Val Ala Arg Gly Tyr Ala
 2260 2265 2270
 Asn His Pro Pro Leu Thr Ser Glu Arg Phe Val Ala Asp Pro Phe Gly
 2275 2280 2285
 Arg Pro Gly Glu Arg Met Tyr Arg Thr Gly Asp Leu Val Arg Trp Asn
 2290 2295 2300
 His Asp Gly Gln Leu Glu Phe Leu Gly Arg Val Asp Glu Gln Val Lys
 2305 2310 2315 2320
 Ile Arg Gly Phe Arg Val Glu Pro Gly Glu Ile Arg Ala Ala Leu Arg
 2325 2330 2335
 Lys Arg Asp Gly Val Ala Gln Ala Val Val Val Pro Arg Thr Asp Arg
 2340 2345 2350
 Leu Gly Glu Arg Arg Leu Val Ala Tyr Val Val Pro Glu Val Pro Ala
 2355 2360 2365
 Gly Ala Asp Glu Asp Ser Thr Glu His Val Glu Lys Trp Arg Ala Ile
 2370 2375 2380
 Tyr Asp Ser Met Tyr Asp Glu Thr Glu Ala Asp Ala Thr Glu Ile Gly
 2385 2390 2395 2400
 Asn Asp Phe Thr Gly Trp Lys Ser Ser Tyr Thr Arg Asp Asn Ile Pro
 2405 2410 2415
 Leu Ser Glu Met Arg Arg Trp Arg Asp Ser Val Val Glu Glu Val Arg
 2420 2425 2430
 Gly Leu Arg Ala Arg Arg Ile Leu Glu Ile Gly Val Gly Ser Gly Leu
 2435 2440 2445
 Leu Leu Gly Pro Leu Ala Pro Glu Ala Glu Ala Tyr Trp Gly Thr Asp
 2450 2455 2460
 Phe Ser Leu Pro Val Ile Glu Arg Leu Glu Val Gln Val Gly Thr Asp
 2465 2470 2475 2480
 Pro Cys Leu Lys Glu Lys Val Ser Leu Arg Cys Gln His Ala Asp Val
 2485 2490 2495

Ala Asp Gly Leu Pro Val Lys Tyr Phe Asp Thr Val Ile Leu Asn Ser
 2500 2505 2510
 Val Val Gln Tyr Phe Pro Asp Ala Ala Tyr Leu Ser Arg Val Leu Asp
 2515 2520 2525
 Val Ala Leu Asp Arg Leu Ala Pro Gly Gly Arg Ile Leu Val Gly Asp
 2530 2535 2540
 Val Arg Asn Tyr Gly Thr Leu Arg Glu Phe Leu Thr Ala Val His His
 2545 2550 2555 2560
 Ala Gln His Pro Gln Asp Ser Ala Ser Ala Val Arg Ala Ala Val Glu
 2565 2570 2575
 Arg Ala Val Leu Ala Glu Lys Glu Leu Val Ile Asp Pro Asp Phe Phe
 2580 2585 2590
 Thr Glu Trp Ala Arg Thr Arg Pro Asp Val Val Ala Val Asp Ile Arg
 2595 2600 2605
 Leu Lys Pro Gly Ala Asp Gln Asn Glu Leu Thr Arg His Arg Tyr Glu
 2610 2615 2620
 Val Ile Leu His Lys Gln Pro Ser Gln Pro Leu Arg Leu Ala Asp Val
 2625 2630 2635 2640
 Arg Thr Ala Asn Trp Gly Ser Glu Val Pro Asp Leu Ser Gly Leu Glu
 2645 2650 2655
 Thr Ala Leu Ala Arg His Gly Gly Arg Leu Arg Leu Ala Arg Ile Pro
 2660 2665 2670
 Asn Ala Arg Leu Val Ser Glu Ala Val Gln Cys Gly Val Pro Thr Asn
 2675 2680 2685
 Val Gly Gly Thr Pro Leu Asp Pro His Glu Leu Ala Ser Trp Gly Gly
 2690 2695 2700
 Gln Arg Gly Tyr Ser Val His Cys Thr Trp Ser Ala Glu Ala Pro Gly
 2705 2710 2715 2720
 Trp Phe Glu Ala Val Ile Ile Pro Val Asp Ser Gly His Cys Arg Asp
 2725 2730 2735
 Gly Val Tyr Arg Pro Val Gly Pro Arg Pro Arg Gln Leu Val Asn Leu
 2740 2745 2750
 Pro Ala Ala Ala Arg Arg Val Ser Arg Leu Pro Ser Trp Leu Arg Glu
 2755 2760 2765
 Glu Leu Ala Ala Glu Leu Pro Glu His Leu Val Pro Gly Asp Ile Val
 2770 2775 2780
 Val Met Glu Arg Leu Pro Leu Thr Thr Asn Gly Lys Ile Asp His Ser
 2785 2790 2795 2800

Arg Leu Pro Glu Val Glu Gln Ala Asp Arg Glu Tyr Gly Ser Pro Arg
 2805 2810 2815
 Thr Pro Leu Glu Ala Glu Leu Ala Gly Leu Phe Ala Glu Val Leu Gly
 2820 2825 2830
 Val Asp Arg Val Gly Ile Glu Asp Asp Phe Phe Asp Cys Gly Gly Ser
 2835 2840 2845
 Ser Leu Gln Val Ile Arg Leu Ile Trp Arg Ile Arg Ala Glu Leu Gly
 2850 2855 2860
 Phe Asp Ile Pro Val Arg Thr Ile Phe Gln His Pro Thr Val Ala Glu
 2865 2870 2875 2880
 Val Ala Glu His Leu Ala Ala Gly Arg Glu Asp Val Glu Phe Asp Asp
 2885 2890 2895
 Pro Phe Ser Val Val Leu Pro Ile Arg Thr Asp Gly Asp Arg Ala Pro
 2900 2905 2910
 Ile Trp Leu Val Pro Pro Gly Gly Gly Leu Ser Trp Ala Tyr Leu Gly
 2915 2920 2925
 Phe Ala Gln His Leu Asp Arg Ser Arg Pro Ile Tyr Gly Leu Gln Ala
 2930 2935 2940
 Arg Gly Phe Val Gly Asp Pro Arg Ala Thr Ser Ile Glu Ser Met Val
 2945 2950 2955 2960
 Asp Asp Tyr Val Asp Gln Ile Leu Arg Val Gln Pro Glu Gly Pro Tyr
 2965 2970 2975
 Asn Val Leu Gly Trp Ser Leu Gly Gly Pro Ile Ala Gln Ala Val Ala
 2980 2985 2990
 Ala Glu Leu Gln Arg Arg Glu Cys Glu Val Asp Leu Leu Ala Val Leu
 2995 3000 3005
 Asp Ser Gly Pro Ala Ser Tyr Phe Ala Asp Phe Gln Thr Pro Asp Glu
 3010 3015 3020
 Lys Leu Val Arg Arg Tyr Leu Ala His Tyr Met Gly His Leu Ala Gly
 3025 3030 3035 3040
 Met Ala Glu Phe Glu Ser Leu Val Gln Thr Ser Thr Thr Leu Phe Ile
 3045 3050 3055
 Glu His Thr Asp Leu Met Thr Gln Phe Thr Ser Pro Pro Phe Arg Gly
 3060 3065 3070
 Asp Leu Ile Phe Phe Ala Ala Leu Ile Asp Gln Gly Thr Arg Glu Arg
 3075 3080 3085

Arg Gln Leu Glu Val Glu Leu Asp Val Gln Trp Arg Glu Tyr Thr Glu
 3090 3095 3100

Gly Thr Val Arg Arg Phe Asp Ile Glu Cys Ala His Asn Glu Met Met
 3105 3110 3115 3120

Trp Pro Glu Asn Ala Ala Met Ile Ser Arg Ile Val Asn Glu Ile Ile
 3125 3130 3135

Asp Ser Ala Arg
 3140

<210> 23
 <211> 71
 <212> PRT
 <213> *Micromonospora sp.*, cepa ML1

<220>
 <223> Proteína Tio22

<400> 23
 Met Ser Val Asn Pro Phe Asp Asp Glu Asp Gly Glu Phe Tyr Val Leu
 1 5 10 15

Val Asn Asp Glu Glu Gln His Ser Leu Trp Pro Thr Phe Gly Asp Val
 20 25 30

Pro Asp Gly Trp Arg Ile Val Phe Gly Pro Ala Gly Arg Ala Glu Ser
 35 40 45

Val Ala Tyr Val Glu Glu Asn Trp Thr Asp Met Arg Pro Lys Ser Leu
 50 55 60

Arg Glu Ala Met Ser Ala Ala
 65 70

<210> 24
 <211> 874
 <212> PRT
 <213> *Micromonospora sp.*, cepa ML1

<220>
 <223> Proteína Tio23

<400> 24
 Val Arg Arg Cys Arg Gln His Asp Val Arg Arg Gly Leu Ala Gly Arg
 1 5 10 15

Arg Val Arg Leu Arg Ala His Asp Phe Glu Leu Pro Arg Arg Arg Trp
 20 25 30

Arg Lys Val Arg Thr Glu Val Phe Thr Gly Cys Leu Gly Ile Glu Leu
 35 40 45

Gly Cys Gly Lys Pro Glu Ile Ala Gly Gly Val Ile Ser Pro Gly His
 50 55 60
 Pro Ala Gly Ser Ala Ser Arg Ala Arg His Phe Gly Gly Pro Met Thr
 65 70 75 80
 Val Ala Ser Pro Arg Gln Ser Arg Glu Asp Leu Ser Leu Val Pro Asp
 85 90 95
 Thr His Asp Val Ile Arg Val Val Gly Ala Arg Glu Asn Asn Leu Lys
 100 105 110
 Asp Ile Arg Leu Glu Ile Pro Lys Arg Arg Leu Thr Val Phe Thr Gly
 115 120 125
 Val Ser Gly Ser Gly Lys Ser Ser Leu Val Phe Asp Thr Ile Ala Ala
 130 135 140
 Glu Ser Gln Arg Leu Ile Asn Glu Thr Tyr Ser Ala Phe Val Gln Gly
 145 150 155 160
 Phe Met Pro Thr Leu Ser Arg Pro Asp Val Asp Val Leu Asp Gly Leu
 165 170 175
 Thr Thr Ala Ile Ile Val Ser Gln Glu Arg Ile Gly Asn Asn Pro Arg
 180 185 190
 Ser Thr Val Gly Thr Val Thr Asp Ala Asn Ala Met Leu Arg Ile Leu
 195 200 205
 Phe Ser Arg Leu Gly Glu Pro Arg Leu Gly Ser Pro Arg Ser Tyr Ser
 210 215 220
 Phe Asn Ile Pro Ser Ile Ser Val Ala Asn Ala Val Thr Val Gly Gln
 225 230 235 240
 Gly Ser Arg Ala Val Glu Glu Gln Arg Gly Phe Asp Val Ala Gly Gly
 245 250 255
 Met Cys Pro Arg Cys Glu Gly Thr Gly Ser Val Asn Asp Val Asp Leu
 260 265 270
 Thr Gln Leu Tyr Asp Ala Thr Lys Ser Leu Asp Glu Gly Ala Ile Arg
 275 280 285
 Val Pro Gly Tyr Gln Val Gly Gly Tyr Asn His Arg Leu Tyr Ala Gly
 290 295 300
 Ser Gly Leu Leu Asp Pro Gly Lys Pro Val Glu Lys Tyr Thr Glu Lys
 305 310 315 320
 Glu Leu Asn Asp Phe Leu His Arg Glu Pro Thr Arg Leu Lys Ile Gly
 325 330 335

Asp	Tyr	His	Thr	Asn	Tyr	Glu	Gly	Leu	Val	Ala	Arg	Val	Gln	Lys	Ser	
			340					345					350			
Leu	Leu	Thr	Lys	Asp	Ile	Glu	Ala	Leu	Gln	Pro	His	Val	Arg	Ala	Phe	
		355					360					365				
Val	Glu	Arg	Ala	Val	Thr	Phe	Thr	Ala	Cys	Pro	Asp	Cys	Gly	Gly	Thr	
	370					375					380					
Arg	Leu	Asn	Glu	Gln	Ala	Arg	Ala	Ser	Arg	Ile	Lys	Gly	Ile	Asn	Ile	
385					390					395					400	
Ala	Asp	Ala	Ser	Ala	Met	Gln	Ile	Asn	Asp	Leu	Ala	Glu	Trp	Val	Arg	
				405					410					415		
Asp	Leu	Ala	Glu	Pro	Ser	Val	Ala	Pro	Leu	Leu	Ala	Thr	Leu	Leu	His	
			420					425					430			
Thr	Leu	Asp	Ser	Phe	Val	Gly	Ile	Gly	Leu	Gly	Tyr	Leu	Ser	Leu	Asp	
		435					440					445				
Arg	Ala	Ser	Gly	Thr	Leu	Ser	Gly	Gly	Glu	Ala	Gln	Arg	Val	Lys	Met	
	450					455					460					
Ile	Arg	His	Leu	Gly	Ser	Pro	Leu	Thr	Asp	Val	Thr	Tyr	Val	Phe	Asp	
465					470					475					480	
Glu	Pro	Thr	Ile	Gly	Leu	His	Pro	His	Asp	Ile	Gln	Arg	Met	Asn	Lys	
				485					490					495		
Leu	Leu	Leu	Arg	Leu	Arg	Asp	Lys	Gly	Asn	Thr	Val	Leu	Val	Val	Glu	
			500					505					510			
His	Lys	Pro	Glu	Thr	Ile	Ala	Ile	Ala	Asp	Arg	Val	Val	Asp	Leu	Gly	
		515					520					525				
Pro	Gly	Ala	Gly	Thr	Ala	Gly	Gly	Thr	Ile	Cys	Phe	Glu	Gly	Thr	Val	
	530					535					540					
Ala	Glu	Leu	Arg	Ser	Ala	Asp	Thr	Leu	Thr	Gly	Arg	His	Leu	Asp	Asp	
545					550					555					560	
Lys	Ala	Ser	Leu	Lys	Asp	Arg	Val	Arg	Thr	Pro	Ser	Gly	Ser	Leu	Glu	
				565					570					575		
Val	Arg	Gly	Ala	Ser	Lys	His	Asn	Leu	Arg	Asp	Val	Asp	Val	Asp	Ile	
			580					585					590			
Pro	Leu	Gly	Val	Leu	Val	Val	Val	Thr	Gly	Val	Ala	Gly	Ser	Gly	Lys	
		595						600				605				
Ser	Ser	Leu	Ile	Glu	Gly	Ser	Val	Ser	Gly	Arg	Asp	Gly	Val	Val	Thr	
	610					615					620					
Ile	Asp	Gln	Gly	Ser	Ile	Lys	Gly	Ser	Arg	Arg	Ser	Asn	Pro	Ala	Thr	
625					630					635					640	

Tyr Thr Gly Leu Leu Asp Pro Ile Arg Lys Ala Phe Ala Lys Ala Asn
 645 650 655
 Gly Val Lys Pro Ala Leu Phe Ser Ser Asn Ser Glu Gly Ala Cys Ala
 660 665 670
 Thr Cys Lys Gly Ala Gly Val Ile Tyr Thr Asp Leu Ala Ile Met Ala
 675 680 685
 Gly Ala Ala Thr Thr Cys Asp Asp Cys Glu Gly Arg Arg Phe Gln Asp
 690 695 700
 Phe Val Leu Gln Tyr Arg Leu Gly Gly Arg Asp Ile Ser Glu Val Leu
 705 710 715 720
 Ala Met Ser Val Ala Glu Ala Gly Arg Phe Phe Ala Ala Gly Gln Ala
 725 730 735
 Arg Thr Pro Ala Ala Val Val Ile Leu Glu Arg Leu Ala Asp Val Gly
 740 745 750
 Leu Gly Tyr Leu Ser Leu Gly Gln Pro Leu Pro Thr Leu Ser Gly Gly
 755 760 765
 Glu Arg Gln Arg Leu Lys Leu Ala Ala His Leu Ala Asp Glu Gly Gly
 770 775 780
 Ile Val Val Leu Asp Glu Pro Thr Thr Gly Leu His Leu Ala Asp Val
 785 790 795 800
 Glu Gln Leu Leu Gly Leu Leu Asp Arg Leu Val Asp Ala Gly Arg Ser
 805 810 815
 Val Met Val Ile Glu His His Gln Ala Val Met Ala His Ala Asp Trp
 820 825 830
 Ile Ile Asp Leu Gly Pro Ala Ala Gly His Asp Gly Gly Arg Val Val
 835 840 845
 Phe Glu Gly Cys Pro Ala Asp Leu Val Ala Ser Arg Ala Thr Leu Thr
 850 855 860
 Gly Glu His Leu Ala Ala Tyr Val Gly Val
 865 870

<210> 25
 <211> 178
 <212> PRT
 <213> *Micromonospora* sp., cepa ML1

<220>
 <223> Proteína Tio24

<400> 25

Val Gly Gly Arg Asn Ser Asp Val Val Phe Ile Arg Trp Pro Ala Gln
 1 5 10 15

Ala Asp Leu Arg Thr Arg Cys Arg Lys Glu Gly Val Pro Arg Leu Leu
 20 25 30

Val Val Glu Gly Gly Ala Gln Pro Pro Ile Ser Asn Asp Pro Val Glu
 35 40 45

Asp Trp Val Arg Pro Pro Val Ser Arg Asn Asp Leu Glu Ala Arg Val
 50 55 60

Lys Ala Leu Gln Gln Arg Phe Asn Ser Arg Arg Leu Pro Thr Ile Asp
 65 70 75 80

Ser Thr Gly Thr Leu Thr Phe Gly Arg His Ser Ile Val Ile Ser Asn
 85 90 95

Thr Gln Ala Glu Leu Met Arg Leu Leu Ile Glu Arg Phe Gly Glu Val
 100 105 110

Val Tyr Arg Ile Glu Phe Val Gln Cys Leu Ala Ile Arg Thr Pro Ala
 115 120 125

Leu Asn Arg Asn Ser Leu Asp Leu His Ile Met Arg Leu Arg Arg Arg
 130 135 140

Leu Leu Pro Ile Asp Leu Ala Ile Arg Thr Val Trp Gly Arg Gly Tyr
 145 150 155 160

Met Leu Ala Tyr Ala Ala Ser Ala Glu Ala Ile Ile Arg Asn Ala Ser
 165 170 175

Gly Met

<210> 26

<211> 287

<212> PRT

<213> *Micromonospora sp.*, cepa ML1

<220>

<223> Proteína Tio25

<400> 26

Val Ala Asp Ala Leu Ser Ser Arg Pro Cys Ser Thr Ala Ala Ile Lys
 1 5 10 15

Arg Asp His Ile Met Leu Ser Ser His Thr Val Ile Val Leu His Ala
 20 25 30

His Pro Asp Asp Glu Ala Ile Phe Thr Gly Leu Thr Met Arg Arg Leu
 35 40 45

```
<210> 27
<211> 134
<212> PRT
<213> Micromonospora sp., cepa ML1
```

<220>
<223> Proteína Tio26

<400> 27

Val Val Val Glu Leu Ala Tyr Thr Asp Pro Asp Arg Ala Val Asp Trp
 1 5 10 15

Leu Val Arg Val Phe Gly Phe Arg Leu Leu Leu Arg Gln Pro Ala Ile
 20 25 30

Gly Thr Ile Arg His Ala Asp Leu Asp Thr Gly Gly Gly Ile Val Met
 35 40 45

Val Arg Arg Thr Gly Glu Pro Tyr Thr Val Ser Cys Ala Gly Gly His
 50 55 60

Thr Cys Lys Gln Val Ile Val Trp Val Ser Asp Val Asp Glu His Phe
 65 70 75 80

Met Arg Ser Thr Ala Ala Gly Ala Asp Ile Val Gln Pro Leu Gln Asp
 85 90 95

Lys Pro Trp Gly Leu Arg Gln Tyr Leu Val Arg Asp Leu Glu Gly His
 100 105 110

Leu Trp Glu Phe Thr Arg His Leu Arg Asp Val Pro Pro Arg Glu Trp
 115 120 125

Gly Ala Val Ala Met Thr
 130

<210> 28

<211> 1804

<212> PRT

<213> *Micromonospora sp.*, cepa ML1

<220>

<223> Proteína Tio27

<400> 28

Val Gly Leu Gln Leu Leu Gly Ser Ala Thr Gln Ser Thr Glu Pro Asn
 1 5 10 15

Leu Pro Gly Leu Thr Val Thr Ser Leu Asp Ala Asn Pro Gly Gly His
 20 25 30

Pro Phe Asp Leu Ser Ile Thr Cys Thr Glu Ser Pro Asp Gln Leu Arg
 35 40 45

Leu Val Ala Asp Tyr Ser Val Asp Leu Phe Asp Glu Pro Arg Met Ala
 50 55 60

Arg Leu Leu Arg His Leu Glu Arg Val Leu Arg Thr Ala Val Asp Ala
 65 70 75 80

Pro Asp Thr Pro Leu Ser Ala Leu Pro Leu Leu Asp Glu Glu Glu Gln
 85 90 95

Thr	Leu	Leu	Leu	His	Thr	Trp	Gln	Gly	Pro	Gln	Gly	Glu	Arg	Ala	Arg		
			100					105					110				
Glu	Pro	Val	His	Val	Gln	Ile	Ala	Arg	Leu	Ala	Ala	Glu	Arg	Pro	Asp		
		115					120					125					
Leu	Val	Ala	Ala	Thr	Leu	His	Gly	Ala	Glu	Leu	Thr	Tyr	Gly	Glu	Leu		
	130					135					140						
Asp	Arg	Arg	Ala	Gly	Val	Leu	Ala	His	Arg	Phe	Arg	Ala	Leu	Gly	Val		
145					150					155					160		
Gly	Arg	Asp	Asp	Ile	Val	Ala	Leu	Leu	Leu	Glu	Arg	Gly	Phe	Asp	Leu		
				165					170					175			
Ile	Val	Gly	Met	Val	Ala	Ala	Gln	Lys	Ala	Gly	Gly	Ala	Phe	Val	Val		
			180					185					190				
Met	Asp	Pro	Gly	His	Pro	Val	Arg	Arg	Ile	Glu	Phe	Ile	Leu	Glu	Asp		
	195						200					205					
Thr	Ala	Ala	Lys	Ala	Val	Val	Thr	Arg	Ser	Ala	Leu	Ala	Asp	Arg	Leu		
	210					215					220						
Pro	Glu	Gly	Ala	Ala	Ala	Thr	Pro	Val	Leu	Val	Asp	Thr	Glu	Trp	Asp		
225					230					235					240		
Glu	Leu	Ser	Ala	Ala	Ala	Ala	Gly	Ala	Ala	Pro	Leu	Asp	Glu	Leu	Ala		
				245					250					255			
Asp	Glu	Asn	Ser	Leu	Ala	Tyr	Val	Leu	Tyr	Thr	Ser	Gly	Ser	Thr	Gly		
			260					265					270				
Lys	Pro	Lys	Gly	Val	Met	Ile	Glu	His	His	Ala	Leu	Asn	Thr	Phe	Leu		
		275					280					285					
Leu	Trp	Leu	Gly	Asn	Ile	Phe	Asp	Phe	Gly	Pro	Gly	Asp	Arg	Leu	Leu		
	290					295					300						
Gln	His	Met	Ala	Pro	Ile	Phe	Asp	Phe	Ala	Glu	Gly	Glu	Ile	Phe	Thr		
305					310					315					320		
Ala	Leu	Thr	Arg	Gly	Val	Thr	Met	Val	Phe	Val	Pro	Glu	Glu	His	Arg		
				325					330					335			
Thr	Asp	Pro	Asp	Val	Ile	Gly	Gln	Leu	Leu	Val	Ser	Glu	Arg	Ile	Thr		
			340					345					350				
Tyr	Ile	Gly	Gly	Pro	Pro	Ala	Ile	Leu	Gly	Arg	Ile	Pro	Pro	Gly	Ser		
		355					360					365					
Tyr	Pro	Asp	Leu	Lys	Tyr	Met	Ile	Ala	Gly	Gly	Glu	Ala	Val	Thr	Gly		
	370					375					380						

Asp	Leu	Ile	Asn	Arg	Trp	Asn	Thr	Pro	Gly	Arg	Arg	Phe	Ile	Asn	Gly	
385					390					395					400	
Tyr	Gly	Pro	Thr	Glu	Ala	Ala	Val	Gly	Cys	Ile	Phe	Tyr	Glu	Cys	Glu	
				405					410					415		
His	Arg	Thr	Trp	Ile	Gly	Gln	Pro	Pro	Ile	Gly	Arg	Ala	Met	Pro	Asn	
			420					425					430			
Arg	Val	Ala	Tyr	Val	Leu	Asp	Arg	Trp	Asn	Asn	Leu	Gln	Pro	Val	Gly	
		435					440					445				
Val	Pro	Gly	Glu	Val	Ile	Thr	Gly	Gly	Glu	Gly	Leu	Ala	Arg	Gly	Tyr	
	450					455					460					
Leu	Asn	Arg	Pro	Glu	Leu	Thr	Ala	Glu	Lys	Phe	Ile	Asp	Asn	Pro	Tyr	
465					470					475					480	
Arg	Pro	Gly	Glu	Arg	Met	Tyr	Arg	Thr	Gly	Asp	Leu	Gly	Val	Trp	Thr	
				485					490					495		
Glu	Asp	Gly	Gln	Ile	Gln	Phe	Leu	Gly	Arg	Ile	Asp	Thr	Gln	Val	Lys	
			500					505					510			
Leu	Asn	Gly	Leu	Arg	Ile	Glu	Leu	Glu	Glu	Ile	Glu	Ser	Thr	Leu	Thr	
		515					520					525				
Ala	His	Pro	Gly	Val	Ala	Glu	Ala	Ala	Val	Ala	Leu	Arg	Glu	Asp	Ser	
	530					535					540					
Pro	Gly	Ser	Lys	Arg	Leu	Val	Gly	Tyr	Val	Val	Pro	Ala	Gly	Ala	Ala	
545					550					555					560	
Ala	Pro	Pro	Thr	Glu	Glu	Leu	Arg	Glu	His	Leu	Leu	Ala	Asp	Leu	Pro	
				565					570					575		
Pro	Tyr	Met	Val	Pro	His	Val	Phe	Val	Ala	Leu	Asp	Val	Leu	Pro	Leu	
			580					585					590			
Thr	Ser	Val	Gly	Lys	Val	Asp	Arg	Asn	Arg	Leu	Pro	Val	Pro	Glu	Glu	
		595					600					605				
Thr	Gly	Ser	Ala	Glu	Val	Glu	Leu	Val	Ala	Pro	Arg	Thr	Pro	Glu	Glu	
	610					615					620					
Glu	Arg	Val	Ala	Glu	Ile	Phe	Ala	Ala	Val	Leu	Gly	Val	Glu	Arg	Val	
625					630					635					640	
Gly	Ala	Asp	Ala	Asn	Phe	Phe	Glu	Leu	Gly	Gly	Asn	Ser	Leu	Met	Ala	
				645					650					655		
Ala	Arg	Val	Leu	Ser	Arg	Ile	Ala	Glu	Leu	Gly	Gly	Ala	Ser	Ala	Thr	
			660					665					670			
Met	Arg	Asp	Phe	Tyr	Thr	Lys	Pro	Arg	Val	Ser	Glu	Ile	Ala	Gln	Leu	
		675					680					685				

Leu	Ala	Gly	Thr	Pro	Ser	Ser	Thr	Asp	Pro	Ala	Thr	Pro	Thr	Asp	Ala	
690						695					700					
Ala	Pro	Val	Ala	Asp	Gln	Thr	Arg	Ala	Arg	Leu	Glu	Gln	Glu	Ile	Ala	
705					710					715					720	
Asp	Leu	Glu	Gln	Arg	Leu	Arg	Ala	Ala	Arg	Ala	Ala	Leu	Ala	Gly	Gln	
				725					730					735		
Asp	Gly	Pro	Val	Ser	Gly	Gly	His	Glu	Gly	Gln	Arg	Leu	Pro	Leu	Ser	
			740					745					750			
Pro	Thr	Gln	Glu	Gln	Leu	Trp	Phe	Leu	Glu	Gln	Leu	Ala	Pro	Gly	Gln	
		755					760					765				
Pro	Thr	Tyr	His	Ile	Pro	Val	Pro	Val	Arg	Leu	Thr	Gly	Pro	Leu	Asp	
		770				775					780					
Val	Asp	Leu	Leu	Gly	Arg	Ala	Leu	Thr	Ala	Val	Val	Ala	Arg	His	Glu	
785					790					795					800	
Ala	Leu	Arg	Thr	Thr	Phe	Gly	Thr	Asp	Asp	Asp	Thr	Pro	Tyr	Gln	Leu	
				805					810					815		
Val	Ser	Pro	Ala	Gly	Glu	Val	Pro	Leu	Pro	Val	His	Glu	Leu	Asp	Ala	
			820					825					830			
Thr	Ala	Ala	Asp	Asp	Arg	Glu	Gln	Ala	Leu	Arg	Val	Glu	Leu	Glu	Arg	
		835					840					845				
Asp	Ala	Thr	Glu	Pro	Phe	Asp	Leu	Gly	Arg	Asp	Leu	Met	Leu	Arg	Ala	
		850				855					860					
Arg	Leu	Ile	Arg	Leu	Asp	Gly	Gln	Glu	His	Val	Leu	Ser	Met	Val	Leu	
865					870					875					880	
His	His	Ile	Cys	Ala	Asp	Gly	Trp	Ser	Ala	Gly	Leu	Leu	Val	Arg	Asp	
				885					890					895		
Leu	Ala	Glu	Leu	Tyr	Asp	Ala	Leu	Arg	His	Asp	Arg	Gln	Pro	Gln	Leu	
			900					905					910			
Pro	Ala	Leu	Arg	Leu	Gln	Tyr	Ala	Asp	His	Ala	Leu	Ala	Gln	Arg	Arg	
		915					920					925				
Gln	Leu	Asp	Ala	Gly	Asp	Leu	Glu	Ser	Gln	Leu	Ala	Tyr	Trp	Ala	Arg	
		930				935					940					
Gln	Leu	Ala	Asp	Ala	Pro	Ile	Leu	Asp	Leu	Pro	Ser	Asp	Arg	Pro	Arg	
945					950					955					960	
Pro	Ala	Thr	Ala	Thr	Asn	Arg	Gly	Ala	Leu	Leu	Val	Arg	Arg	Met	Pro	
				965					970					975		

Ala Asp Val Leu Glu Gly Leu Arg Arg Val Ala Gln Gln Glu Gln Ala
980 985 990

Ser Thr Tyr Ala Val Leu Leu Ala Gly Phe Lys Thr Val Leu Met Arg
995 1000 1005

Tyr Thr Gly Val Asp Asp Val Val Val Gly Thr Ala Gly Ala Gly Arg
1010 1015 1020

Gly Arg Pro Glu Leu Glu Asp Leu Val Gly Phe Phe Val Asn Met Leu
1025 1030 1035 1040

Ala Leu Arg Thr Asp Leu Ser Gly Asp Pro Ser Phe Gly Glu Ala Val
1045 1050 1055

Arg Arg Thr Met Gly Thr Leu Leu Ala Ala Trp Asp His Gln Asp Ala
1060 1065 1070

Pro Phe Glu Arg Val Val Gly Arg Leu Gly Arg Arg Arg Asp Ala Ser
1075 1080 1085

Arg Asn Pro Val Phe Gln Val Ser Ile Asp Leu Leu Ser Gly Asp Leu
1090 1095 1100

Val Asp Phe Arg Phe Pro Gly Leu Ala Ala Glu Phe Leu Asp Val Asp
1105 1110 1115 1120

Pro Gly Val Ser Arg Phe Asp Met Ala Ile Asn Ala Tyr Glu Glu Thr
1125 1130 1135

Asp Gly Ile Thr Phe Arg Ile Glu Tyr Ala Thr Asp Leu Phe Asp Thr
1140 1145 1150

Ala Arg Ile Glu Ala Met Phe Gly His Leu Glu Arg Val Leu Arg Ala
1155 1160 1165

Ala Val Ala Ala Pro Gln Thr Pro Leu Ser Ala Leu Pro Leu Leu Asp
1170 1175 1180

Asp Thr Glu Arg Thr Glu Leu Leu Ala Leu Ala Gln Gly Lys Arg Arg
1185 1190 1195 1200

Pro His Arg Pro Ala Thr Leu Pro Glu Leu Phe Ala Ala Ser Val Ala
1205 1210 1215

Arg Thr Pro Gln Ala Thr Ala Val Ile Ala Ala Glu Gly Thr Leu Ser
1220 1225 1230

Tyr Ala Glu Leu Asp Glu Arg Ser Arg Arg Leu Ala Ala Arg Leu Arg
1235 1240 1245

Ala Ala Gly Val Gly Thr Gly Asp Val Val Pro Val Leu Leu Glu Arg
1250 1255 1260

Ser Leu Ala Glu Val Val Ser Val Leu Ala Val Leu Arg Ala Gly Ala
1265 1270 1275 1280

Ala Tyr Ala Pro Met Asp Pro Ala Ala Pro Pro Ser Arg Leu Ala Phe
 1285 1290 1295
 Gln Val Arg Asp Thr Gly Ala Arg Val Val Val Gly Asp Pro Thr Arg
 1300 1305 1310
 Leu Asp Thr Val Val Gly Val Thr Ile Val Asp Pro Ala Ala Glu Leu
 1315 1320 1325
 Pro Glu Ala Ala Asp Glu Pro Trp Leu Ala Glu Leu Thr Pro Ala Ser
 1330 1335 1340
 Leu Ala Tyr Val Leu Tyr Thr Ser Gly Ser Thr Gly Glu Pro Lys Gly
 1345 1350 1355 1360
 Val Leu Ile Glu His Gly Met Val Ala Gly Tyr Leu Glu Trp Met Val
 1365 1370 1375
 Pro Asp Phe Gly Val Glu Pro Arg Ser Arg Met Met His Cys Cys Ala
 1380 1385 1390
 Pro Val Phe Asp Leu Ala Val Gly Glu Leu Phe Gly Ala Leu Thr Ala
 1395 1400 1405
 Gly Ala Thr Val Val Val Ala Gly His Asp Asp Val His Arg Pro Gly
 1410 1415 1420
 Ala Leu Thr Ala Leu Ala Arg Glu Thr Gly Ala Thr His Ile Phe Thr
 1425 1430 1435 1440
 Thr Pro Thr Val Leu Gly Leu Leu Asp Pro Ala Glu Leu Pro Gln Leu
 1445 1450 1455
 Arg Ser Val Met Met Ala Gly Glu Pro Ala Pro Pro Ser Leu Val Ala
 1460 1465 1470
 Ala Trp His Val Asp Gly Arg Arg Val Leu Asn Leu Tyr Gly Pro Ala
 1475 1480 1485
 Glu Thr Thr Val Gly Cys Thr Trp Tyr Glu Cys Pro Ala Gly Thr Ala
 1490 1495 1500
 Pro Ala Ala Thr Val Pro Ile Gly Arg Pro Met Pro Asn Arg Arg Val
 1505 1510 1515 1520
 Tyr Val Val Asp Asp Ser Gly Asp Leu Ala Pro Val Gly Val Pro Gly
 1525 1530 1535
 Glu Leu Val Val Ala Gly Asn Gly Val Ala Arg Gly Tyr Leu Asp Arg
 1540 1545 1550
 Pro Glu Leu Thr Ala Ser Ala Phe Gly Ala Asp Glu Arg Ala Gly Glu
 1555 1560 1565

Gly Arg Ala Tyr Arg Thr Gly Asp Leu Ala Arg Trp Asn Arg Asp Gly
 1570 1575 1580
 Gln Leu Glu Phe Leu Gly Arg Arg Asp Gly Gln Val Lys Leu Arg Gly
 1585 1590 1595 1600
 Leu Arg Val Glu Leu Gly Glu Val Glu His Ala Leu Arg Thr Ala Ala
 1605 1610 1615
 Gly Val Ser Gly Ala Ala Ala Ala Val His Gly Thr Gly Ala Glu Ala
 1620 1625 1630
 Val Leu Val Gly Tyr Val Val Gly Glu Leu Ser Asp Asp Ala Glu
 1635 1640 1645
 Ala Leu Arg Thr Thr Leu Ala Gly Lys Leu Pro Gly Tyr Met Val Pro
 1650 1655 1660
 Asn Val Val Leu Arg Leu Pro Gln Leu Pro Leu Thr Arg Thr Gly Lys
 1665 1670 1675 1680
 Val Asp Arg Lys Arg Leu Pro Ala Pro Glu Ser Gly Ala Gly Gln Ala
 1685 1690 1695
 Ala Tyr Val Pro Pro Gly Thr Asp Asp Glu Arg Leu Ile Ala Gly Val
 1700 1705 1710
 Phe Ala Asp Leu Leu Gly Leu Glu Arg Ala Gly Arg Glu Asp Glu Phe
 1715 1720 1725
 Phe Leu Ser Gly Gly Asn Ser Leu Gln Ala Ile Arg Ala Val Gly Arg
 1730 1735 1740
 Leu Asn Asp Ala Leu Gly Val Pro Leu Thr Val Arg Asp Phe Tyr Ala
 1745 1750 1755 1760
 Ala Pro Thr Val Ala Ala Leu Ala Glu Leu Ala Gly Arg Arg Arg Ala
 1765 1770 1775
 Ala Glu Gln Ala Ala Gln Trp Gln Leu Leu Asp Glu Ile Glu Gln Leu
 1780 1785 1790
 Ser Asp Asp Glu Val Glu Arg Leu Leu Ala Ala Asp
 1795 1800

<210> 29
 <211> 1345
 <212> PRT
 <213> *Micromonospora* sp., cepa ML1

<220>
 <223> Proteína Tio28

<400> 29

```

Met Ser Asp Glu Thr Phe Asp Gly Arg Val Ala Arg Leu Asp Glu Arg
 1           5           10           15

Arg Arg Leu Leu Leu Glu Arg Met Arg Gln Gln Arg Arg Gln Pro Pro
          20           25           30

Ala Thr Ile Gly Ala Arg Pro Ala Gly Asp Glu Arg Val Pro Leu Ser
          35           40           45

Arg Ala Gln Glu Gln Leu Trp Phe Leu Ala Gln Leu Ala Pro Ala Glu
          50           55           60

Pro Thr Tyr Asn Leu Val Gln Ala Ser Tyr Leu Thr Gly Glu Val Asp
 65           70           75           80

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

Ala Leu Arg Thr Val Val Glu Ser Val Asp Gly Val Ala Tyr Gln Arg
          100          105          110

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

Leu Pro Glu Glu Gln Arg Arg Pro Ala Ala Leu Thr Leu Leu Gln Gln
          130          135          140

Gln Glu Val His Arg Pro Phe Asp Leu Ala Thr Gly Pro Leu Phe Arg
          145          150          155          160

Ala Arg Leu Val Arg Leu Gly Pro Thr Glu His Ala Leu Ala Val Ala
          165          170          175

Ala His His Ile Val Val Asp Gly Trp Ser Met Gly Leu Ile Leu Arg
          180          185          190

Glu Leu Ala Ala Ala Tyr Asp Ala Asp Arg His Gly Thr Arg Ala Arg
          195          200          205

Leu Asp Pro Pro Ala Leu Gln Tyr Pro Asp Phe Ala Arg Trp Gln His
          210          215          220

Gln Gln Leu Ala Thr Ala Glu Leu Asp Asn His Leu Arg Tyr Trp Ala
          225          230          235          240

Gln Arg Leu Ala Asp Leu Pro Thr Leu Glu Leu Pro Ala Asp Arg Pro
          245          250          255

Arg Pro Thr Thr Ala Ser Phe Arg Gly Ala Leu Leu Glu Gln Pro Leu
          260          265          270

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

```

Ala	Ser	Pro	Phe	Met	Val	Leu	Val	Ala	Ala	Phe	Ala	Ala	Met	Leu	Ala	290	295	300
Arg	Tyr	Thr	Gly	Thr	Asp	Asp	Leu	Ala	Val	Gly	Thr	Thr	Phe	Ser	Gly	305	310	315
Arg	Ser	Arg	Pro	Glu	Leu	Glu	Ser	Val	Val	Gly	Phe	Phe	Ala	Asn	Met	325	330	335
Gly	Val	Leu	Arg	Leu	Asp	Thr	Ser	Gly	Asp	Pro	Ser	Phe	Ala	Thr	Leu	340	345	350
Val	Asp	Arg	Ala	Arg	Asp	Thr	Cys	Leu	Gly	Ala	Trp	Glu	His	Gln	Asp	355	360	365
Ala	Pro	Phe	Glu	Arg	Val	Val	Gln	Arg	Leu	Ala	Pro	Pro	Arg	Asp	Pro	370	375	380
Gly	Arg	Asn	Pro	Leu	Phe	Gln	Val	Ala	Val	Gln	Leu	Leu	Asp	Gly	Ser	385	390	395
Thr	Trp	Gly	Gly	Pro	Thr	Leu	Pro	Gly	Thr	Asp	Ser	Thr	Pro	Val	Asp	405	410	415
Leu	Arg	Leu	Asp	Arg	Ser	Arg	Phe	Asp	Met	Met	Val	Ser	Leu	Val	Asp	420	425	430
His	Gly	Asp	Arg	Tyr	Ser	Val	Leu	Thr	Glu	Tyr	Ser	Thr	Asp	Leu	Phe	435	440	445
Asp	Ala	Arg	Arg	Val	Arg	Ala	Met	Leu	Asp	His	Val	Ala	Arg	Leu	Leu	450	455	560
Asp	Ala	Gly	Leu	Ala	Asp	Pro	Ala	Thr	Thr	Leu	Ser	Arg	Leu	Pro	Met	465	470	475
Leu	Asp	Pro	Ala	Glu	Arg	Glu	Arg	Val	Leu	Ala	Leu	Gly	Ala	Gly	Glu	485	490	495
Arg	Arg	Asp	Val	Ala	Pro	Ala	Thr	Val	His	Gln	Leu	Val	Arg	Ala	Gln	500	505	510
Ala	Glu	Arg	Asn	Pro	Asp	Ala	Val	Ala	Val	Arg	His	Gly	Asp	His	Glu	515	520	525
Leu	Thr	Tyr	Arg	Glu	Leu	Leu	Asp	Arg	Ala	Ala	Arg	Arg	Ala	Ala	Gly	530	535	540
Leu	Arg	Asp	Ala	Gly	Leu	Arg	Pro	Gly	Asp	Arg	Val	Pro	Ala	Leu	Leu	545	550	555
Asp	Arg	Gly	Val	Asp	Glu	Val	Leu	Thr	Pro	Met	Ala	Ile	Trp	His	Ala	565	570	575
Gly	Gly	Val	Tyr	Val	Pro	Leu	Asp	Thr	Ala	Ala	Pro	Pro	Asn	Arg	Leu	580	585	590

Arg	Arg	Ile	Leu	Thr	Asn	Cys	Ala	Ala	Arg	Leu	Val	Leu	Thr	Arg	Thr	595	600	605	
Glu	His	Ala	Ala	Gly	Val	Pro	Glu	Gly	Pro	Trp	Arg	Thr	Leu	Leu	Leu	610	615	620	
Asp	Thr	Gly	Asp	Leu	Asp	Ser	Ala	Ala	Gly	Leu	Ala	Pro	Glu	Pro	Ser	625	630	635	640
Thr	Gly	Pro	Asp	Asp	Leu	Ala	Tyr	Val	Leu	His	Thr	Ser	Gly	Thr	Thr	645	650	655	
Gly	Asp	Pro	Lys	Gly	Val	Gln	Leu	Pro	His	Arg	Gly	Phe	Val	Thr	Tyr	660	665	670	
Leu	Asp	Trp	Met	Arg	Asp	Gln	Trp	Arg	Cys	Gly	Pro	Gly	Asp	Arg	Val	675	680	685	
Leu	His	Gly	Cys	Ala	Pro	Val	Phe	Asp	Leu	Ala	Ala	Gly	Glu	Val	Leu	690	695	700	
Ala	Ala	Leu	Thr	Ser	Gly	Ala	Thr	Val	Val	Val	Ile	Gly	Lys	Asp	Arg	705	710	715	720
Leu	Leu	Ser	Pro	Gly	Gly	Leu	Ala	Glu	Val	Leu	Ala	Ala	Glu	Arg	Ile	725	730	735	
Thr	His	Leu	Phe	Leu	Thr	Pro	Thr	Thr	Leu	Gly	Leu	Val	Glu	Asp	Asp	740	745	750	
Thr	Asp	Arg	Phe	Asp	Ala	Leu	Arg	Glu	Val	Leu	Val	Gly	Gly	Glu	Val	755	760	765	
Cys	Pro	Ala	Asp	Leu	Val	Ala	Arg	Trp	Ala	Arg	Pro	Gly	Arg	Arg	Arg	770	775	780	
Phe	Val	Asn	Leu	Tyr	Gly	Pro	Thr	Glu	Ala	Thr	Val	Gly	Cys	Leu	Ala	785	790	795	800
His	Asp	Cys	Thr	Asp	Trp	Ser	Ala	Ala	Thr	Pro	Pro	Pro	Ile	Gly	Ala	805	810	815	
Thr	Met	Pro	Asn	Arg	Arg	Thr	Tyr	Leu	Val	Asp	Arg	Trp	Asp	Asn	Pro	820	825	830	
Val	Pro	Val	Gly	Val	Pro	Gly	Glu	Ile	Leu	Val	Gly	Gly	Ala	Ala	Leu	835	840	845	
Ser	Thr	Gly	Tyr	Leu	Asn	Asp	Asp	Glu	Leu	Thr	Gly	Arg	Arg	Phe	Thr	850	855	860	
Ala	Asp	Pro	Phe	Asp	Pro	Gln	Gly	Arg	Val	Tyr	Arg	Thr	Gly	Asp	Arg	865	870	875	880

Gly His Trp Asp Ala Gln Gly Arg Leu His Phe Gly Gly Arg Leu Asp
 885 890 895
 Gly Gln Val Lys Leu Arg Gly Leu Arg Val Glu Leu Asp Glu Ile Glu
 900 905 910
 Ala Leu Leu Ser Arg His Ser Gly Val Ala Arg Ala Ala Val Thr Val
 915 920 925
 Val Arg Asp Gly Asn Gly Val Gln Arg Leu Val Gly Tyr Val Val Asp
 930 935 940
 Ala Ala Glu Pro Ala Ser Asp Ala Asp Leu Arg Thr His Leu Ser Thr
 945 950 955 960
 Glu Leu Pro Ala His Leu Val Pro Gly Arg Phe Val Arg Leu Pro Ala
 965 970 975
 Leu Pro Leu Thr Thr Ser Gly Lys Val Asn Arg Arg Ala Leu Pro Ala
 980 985 990
 Pro Glu Ala Ala Asp Thr Asp Asp Thr Gly Ala Ala Pro Thr Thr Gly
 995 1000 1005
 Ala Glu Arg Gln Val Ala Glu Ile Phe Ala Glu Val Leu Gly Val Ala
 1010 1015 1020
 Arg Val Gly Ala Glu Thr Ser Phe Phe Asp Leu Gly Gly His Ser Leu
 1025 1030 1035 1040
 Gln Ala Ala His Val Leu Ala Arg Ile Ala Gly Arg Ile Gly Thr Ser
 1045 1050 1055
 Val Ser Leu Lys Gln Phe Tyr Ala Thr Pro Thr Val Arg Ala Val Ala
 1060 1065 1070
 Arg Phe Val Glu His Gly Pro Val Val Asp Ala Ser Pro Leu Val Thr
 1075 1080 1085
 Leu Lys Ala Thr Gly Ser Arg Pro Arg Leu Tyr Cys Pro His Ala Val
 1090 1095 1100
 Ser Gly Ser Pro Tyr Trp Tyr Leu Pro Leu Ser Arg Ala Leu His Pro
 1105 1110 1115 1120
 Glu Gln Pro Val Asp Gly Phe Glu Ala Pro Gly Leu Glu Gly Gly Ile
 1125 1130 1135
 Glu Pro Val Asp Asp Leu Val Thr Leu Ala Gly Arg Tyr Val Gly Ala
 1140 1145 1150
 Leu Arg Glu Arg Gln Pro His Gly Pro Tyr Leu Leu Ala Gly Trp Ser
 1155 1160 1165
 Met Gly Gly Phe Leu Cys Phe Glu Met Ala Arg Gln Leu Ala Ala Val
 1170 1175 1180

Gly Ala Ala Pro Asp Leu Val Val Met Ile Asp Ser Asn Glu Pro Gly
 1185 1190 1195 1200
 Pro Leu Pro Leu Pro Ser Glu Arg Gln Val Met Glu Thr Phe Val Ala
 1205 1210 1215
 Asp Leu Ala Gly Leu Ser Ala Ala Pro Ala Pro Pro Leu Asp Ala Glu
 1220 1225 1230
 Val Val Asp Ala Ala Asp Pro Ile Ala Ala Leu Thr Gly Phe Leu Val
 1235 1240 1245
 Arg His Gly Leu Ile Pro Ser Asp Val Arg Ala Asp Phe Val Ala Asn
 1250 1255 1260
 Arg Tyr Ala Val Phe Arg Ala Asn Met Arg Ala Ile Tyr Gly Tyr Gln
 1265 1270 1275 1280
 Pro Ala Pro Tyr Ala Gly Arg Thr Val Leu Val Gln Ala Glu Arg Glu
 1285 1290 1295
 Ala Ser Arg Ala Ala Trp Arg Arg Trp Ser Pro Ala Met Glu Val Val
 1300 1305 1310
 Thr Leu Pro Gly Asp His Tyr Ser Met Trp Ser Pro Gln His Leu Pro
 1315 1320 1325
 Ala Leu Ala Ala Leu Ile Asp Asp Gln Val Ala Arg Val Leu Asp Leu
 1330 1335 1340

Gly
1345

<210> 30
 <211> 519
 <212> PRT
 <213> *Micromonospora sp.*, cepa ML1

<220>
 <223> Proteína ORF29

<400> 30
 Met Ala Thr Glu Asp Glu Gly Arg Pro Arg Gly Ala Arg Leu Val Leu
 1 5 10 15
 Leu Ile Leu Leu Val Ala Gln Val Met Ala Thr Met Asp Asn Ser Ile
 20 25 30
 Val Ala Val Ala Thr Lys Thr Ile Lys Asp Asp Leu Gln Thr Ser Gly
 35 40 45
 Ala Val Leu Gln Leu Ile Leu Ser Gly Tyr Thr Leu Met Phe Ala Val
 50 55 60

Leu	Val	Val	Thr	Gly	Ala	Arg	Leu	Gly	Gly	Asp	Val	Gly	His	Arg	Arg	
65					70					75					80	
Leu	Phe	Met	Ile	Gly	Leu	Thr	Gly	Phe	Thr	Ala	Ser	Ser	Leu	Leu	Cys	
				85					90					95		
Gly	Ile	Ala	Pro	Thr	Ala	Gly	Thr	Leu	Val	Ala	Ala	Arg	Leu	Leu	Gln	
			100					105					110			
Gly	Ala	Phe	Gly	Ala	Met	Met	Val	Pro	Gln	Val	Leu	Ser	Val	Ile	Gln	
		115					120					125				
Ile	Val	Phe	Thr	Gly	Ser	Ala	Arg	Ala	Arg	Ala	Ile	Gly	Leu	Tyr	Ser	
	130					135					140					
Met	Val	Leu	Ala	Leu	Gly	Val	Ala	Ala	Gly	Gln	Ile	Ala	Gly	Gly	Val	
145					150					155					160	
Ile	Val	Ser	Ala	Asp	Val	Leu	Gly	Ser	Gly	Trp	Arg	Ala	Ala	Phe	Leu	
				165					170					175		
Val	Asn	Val	Pro	Ile	Gly	Leu	Ala	Leu	Leu	Ala	Ala	Ala	Pro	Arg	Tyr	
			180					185					190			
Leu	Pro	Gly	His	Thr	Pro	Ser	Ser	Arg	Phe	Arg	Pro	Asp	Ile	Val	Gly	
		195					200					205				
Ile	Leu	Leu	Leu	Ala	Val	Ser	Met	Gly	Ala	Val	Val	Ala	Pro	Leu	Val	
	210					215					220					
Phe	Gly	Arg	Glu	Gln	Gly	Trp	Pro	Ala	Trp	Ala	Leu	Ala	Ser	Ile	Ala	
225					230					235					240	
Ala	Gly	Thr	Ala	Gly	Leu	Val	Leu	Phe	Val	Arg	Tyr	Glu	Leu	Arg	Leu	
				245					250					255		
Gly	Ala	Arg	Gly	Gly	Gln	Pro	Val	Leu	Glu	Ile	Asp	Ala	Leu	Arg	Pro	
			260					265					270			
Arg	Gly	Val	Lys	Ser	Gly	Leu	Leu	Ala	Cys	Cys	Ile	Leu	Asn	Phe	Ala	
		275					280					285				
Phe	Ala	Gly	Val	Leu	Phe	Pro	Leu	Thr	Leu	His	Val	Gln	Asn	Gly	Leu	
	290					295					300					
Gly	Tyr	Thr	Pro	Leu	Gln	Ala	Gly	Leu	Met	Phe	Ile	Pro	Tyr	Pro	Val	
305					310					315					320	
Gly	Phe	Ala	Ala	Val	Ser	Leu	Thr	Trp	Thr	Arg	Leu	Pro	Lys	Arg	Phe	
				325					330					335		
His	Arg	Pro	Leu	Pro	Val	Ala	Gly	Leu	Leu	Val	Phe	Ala	Leu	Ala	Leu	
			340					345					350			

Ala Ala Leu Ala Val Val Val Ala Gly Gly Trp Pro Ile Ala Val Val
 355 360 365

Ser Ala Leu Leu Leu Leu Ala Gly Ala Gly Met Ala Ala Gly Phe Ser
 370 375 380

Thr Leu Val Glu Gln Thr Ala Ala Thr Val Gly Pro Arg Tyr Ala Ala
 385 390 395 400

Ala Leu Ser Ala Leu Val Ser Thr Gly Thr Leu Leu Ala Ser Val Val
 405 410 415

Ser Val Val Val Val Gly Gly Val Tyr Leu Ala Thr Ala Glu Asn Asp
 420 425 430

Pro Ser Arg Ser Ala Leu Gly Leu Ser Arg Ser Leu Trp Val Asp Ala
 435 440 445

Gly Leu Leu Met Leu Gly Phe Leu Leu Ala Tyr Arg Thr Trp Arg Leu
 450 455 460

Ser Thr Arg Val Ser Ala Glu Thr Ala Pro Ser Gly Asp Ala Ala Glu
 465 470 475 480

Gln Val Glu Ala Val Ala Ala Pro Gly Asn Ala Pro Pro Gly Glu Val
 485 490 495

Asp Val Glu Pro Ala Thr Gly Pro Val Glu Glu Ser Ser Ala Pro Thr
 500 505 510

Gly Ala Arg Ser Ala Pro Arg
 515

<210> 31
 <211> 613
 <212> PRT
 <213> *Micromonospora sp.*, cepa ML1

<220>
 <223> Proteína ORF30

<400> 31
 Met Ser Glu Ser Arg Thr Ala Asp Thr Ala Ser Ala Lys Arg Arg Gly
 1 5 10 15

Thr Leu Gly Phe Leu Leu Gly Val Ala Leu Val Thr Ala Gly Val Leu
 20 25 30

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

Ser Ala Glu Pro Gly Met Asp Gly Met Asp Asp Met Asp Asp Met Asp
 50 55 60

Ser Met Pro Gly Met Glu Gly Met Ser Pro Gly Glu Pro Met Asp Met
 65 70 75 80
 Trp Thr Pro Thr Met Ile Leu Gly Met Leu Leu Leu Val Val Gly Leu
 85 90 95
 Val Val Ala Gly Arg Ala Leu Phe Ala Gly Val Arg Arg Pro Asp Ser
 100 105 110
 Ala Ser Val Thr Val Glu Ser Ile Glu His Gly Arg Leu Arg Pro Ala
 115 120 125
 Tyr Trp Ile Thr Cys Ala Val Leu Thr Ile Ala Leu Val Val Asp Ile
 130 135 140
 Met Lys Pro Leu Thr Leu Gly Phe Val Met Pro Gly Met Ser Asp Glu
 145 150 155 160
 Tyr Gly Met Ser Leu Glu Ser Val Ser Ile Leu Pro Leu Val Ala Leu
 165 170 175
 Thr Gly Thr Ala Val Gly Ser Val Val Trp Gly Leu Leu Gly Asp Arg
 180 185 190
 Tyr Gly Arg Arg Thr Ala Leu Leu Leu Ala Thr Leu Leu Phe Ile Ala
 195 200 205
 Thr Ser Ala Cys Gly Ala Met Pro Thr Phe Glu Trp Asn Leu Val Met
 210 215 220
 Cys Phe Thr Met Gly Thr Ser Ala Gly Gly Leu Leu Pro Leu Val Phe
 225 230 235 240
 Thr Leu Ile Ala Glu Leu Thr Pro Arg Arg His Arg Gly Trp Val Ala
 245 250 255
 Val Thr Val Gly Gly Ile Gly Gly Leu Gly Gly Tyr Leu Ala Ala Ser
 260 265 270
 Glu Ala Ala Arg His Leu Glu Pro Ala Leu Thr Trp Arg Ala Leu Trp
 275 280 285
 Leu Ile Gly Leu Pro Ser Gly Leu Leu Leu Leu Leu Met Ala Pro Leu
 290 295 300
 Ile Pro Glu Ser Pro Leu Tyr Leu Leu Arg Ala Gly Arg Arg Thr Glu
 305 310 315 320
 Ala Glu Ala Val Leu Arg Arg Tyr Gly Ser His Leu Gly Ala Thr Pro
 325 330 335
 Pro Ala Ala Asp Thr Ala Ala Ala Ala Thr Arg Pro Gly Val Gly Ala
 340 345 350
 Met Leu Arg Ala Tyr Pro Val Val Thr Ser Val Ile Gly Val Val Gly
 355 360 365

Val Val Trp Gly Leu Val Asn Phe Gly Phe Leu Val Met Leu Pro Ala
 370 375 380
 Gln Leu Arg Glu Ala Gly Leu Ala Gly Gly Val Ala Ser Gly Leu Leu
 385 390 395 400
 Ala Asp Ser Ala Leu Tyr Ser Thr Pro Ala Leu Ala Leu Val Val Leu
 405 410 415
 Leu Tyr Ala Ala Trp Ser Gly Arg Arg Ala Leu Ala Leu Phe Ile Ala
 420 425 430
 Ile Thr Ala Val Gly Leu Gly Gly Val Ala Leu Trp Thr Thr Asp Ser
 435 440 445
 Gly Gly Thr Val Leu Leu Leu Ala Ser Ile Gly Leu Leu Val Leu Ser
 450 455 460
 Leu Ser Ala Val Asn Ala Met Leu Leu Pro Tyr Ser Ala Glu Ile Tyr
 465 470 475 480
 Pro Thr Ala Met Arg Ala Thr Gly Ser Gly Phe Ala Gly Ala Ala Thr
 485 490 495
 Lys Val Gly Gly Ile Leu Gly Pro Ser Ala Met Leu Leu Ile Leu Ser
 500 505 510
 Leu Gly Gly Gly Leu Val Ala Pro Ala Ile Ala Leu Gly Val Leu Cys
 515 520 525
 Leu Ala Ala Thr Val Leu Leu Leu Val Phe Gly Pro Glu Leu Gly Ala
 530 535 540
 Ser Val Gly Pro Leu Ala Arg Arg Gly Val Gly Thr Val Pro Val Ser
 545 550 555 560
 Arg Thr Gly Pro Ala Pro Gly Thr Gly Pro Thr Asp Thr Glu Ser Pro
 565 570 575
 Ala Ser Val Ala Trp Ser Ala Thr Gly Asn Val Val Pro Ala Thr Arg
 580 585 590
 Pro Ala Val Ser Asp Asp Asp Asp Pro Arg Asp Ala Arg Thr Pro Ala
 595 600 605
 Thr Pro Ser Arg Thr
 610

<210> 32
 <211> 191
 <212> PRT
 <213> *Micromonospora sp.*, cepa ML1

<220>
 <223> Proteína ORF31

<400> 32
 Val Pro His Val Asp Arg Ser Ala Arg Ser Gly Pro Cys Cys Ala Ser
 1 5 10 15
 Met Asp Tyr Met Thr Ile Gln Ser Asp Pro Met Ala Ser Thr Asp Glu
 20 25 30
 Val Arg Gly Glu Pro Ala Asp Glu Phe Val Gly Ala Leu Val Ser Leu
 35 40 45
 Ser Arg Val Met Val Ala Val Ala Met Arg Thr Leu Ala Arg Ile Asp
 50 55 60
 Ala Asp Val Thr Leu Gln Gln Phe Arg Thr Leu Val Val Leu Val Pro
 65 70 75 80
 Gly Pro Gln Arg Val Ile Asp Leu Ala Lys Leu Leu Gly Val Gln Ser
 85 90 95
 Ser Thr Ala Thr Arg Met Cys Asp Arg Leu Val Arg Lys Gly Leu Val
 100 105 110
 Ala Arg His Glu Arg Val Ala Asp Arg Arg Ala Ala Trp Val Gly Leu
 115 120 125
 Thr Val Ala Gly Lys Glu Leu Val Gly Leu Gly Met Arg Ala Arg Gln
 130 135 140
 Gln Glu Leu Glu Arg Met Leu Ala Asp Leu Arg Pro Asp Asp Thr Asp
 145 150 155 160
 Val Thr Thr Gly Ala Leu Val Ala Val Leu Arg Ala Thr Gly Glu Leu
 165 170 175
 Thr Asp Gly Ala Trp Trp Arg Gln Trp Ala Asp Ser Val Val Glu
 180 185 190

<210> 33
 <211> 224
 <212> PRT
 <213> *Micromonospora sp.*, cepa ML1

<220>
 <223> Proteína ORF32

<400> 33

Met Thr Val Val Pro Asp Asp Asn Leu Met Thr Leu Ile Cys Asp Val
 1 5 10 15

Cys Gly Asp Thr Thr Thr Gly Thr Ala Cys Val Leu Pro Asp Ala Glu
 20 25 30

Val Val Trp Thr Leu Val Ser Asp His Gly Trp Ser Gly Ser Pro Phe
 35 40 45

Ala Thr Gly Pro His Arg Cys Pro Arg Cys Ser Ala Leu Pro Pro Gly
 50 55 60

Ser Ser Leu Gly Asp Gly Asp Arg Ala Gly Ser Gly Leu Ala Glu Ala
 65 70 75 80

Ala Val Val Val Pro Ala Asp Gly Gly Asp Gln Asp Thr Gly Asp Ala
 85 90 95

Leu Arg Ala Ala Leu Arg Gly Ala Phe Asp Val Ala Gly Thr Val Val
 100 105 110

Val Asp Leu Thr Gln Val Gln Val Ile Asp Ser Leu Gly Leu Gly Met
 115 120 125

Leu Val Arg Ala His Arg Asp Val Arg Glu Arg Gly Gly Arg Leu Cys
 130 135 140

Leu Ala Ala Pro Ser Arg Phe Ile Arg Thr Val Leu His Thr Met Lys
 145 150 155 160

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

Gln Leu Arg Asn Ala Ala Ser Gln Asp Ala Val Ser Ala Gly Val Ser
 180 185 190

Pro Val Asp Val Thr Pro Ser Asp Leu Pro Arg Pro Ala Ala Glu Val
 195 200 205

Pro Ser Met Pro Met Ala Arg Arg Arg Ala Ala Ser Arg Val Thr Val
 210 215 220

<210> 34

<211> 89

<212> PRT

<213> *Micromonospora* sp., cepa ML1

<220>

<223> Proteína ORF33

<400> 34

Met Thr Ser Ala Gly Ile Val Thr Ala Leu Ala Val Gly Leu Ala Val
 1 5 10 15

Gly Ala Leu Gly Arg Leu Val Leu Pro Gly Arg Pro Ala Ala Pro Leu
 20 25 30

Trp Leu Thr Met Thr Ile Gly Ala Val Ala Ser Leu Leu Gly Ala Ile
 35 40 45

Thr Ala Trp Leu Ala Gly Val Gln Gly Phe Ser Ile Leu Gly Leu Val
 50 55 60

Val Gln Ala Gly Leu Ala Ala Val Ala Val Val Ile Val Val Ala Thr
 65 70 75 80

Thr Gly Lys Glu Arg Ser Asp Ser Leu
 85

<210> 35
 <211> 168
 <212> PRT
 <213> *Micromonospora sp.*, cepa ML1

<220>
 <223> Proteína ORF34

<400> 35
 Val Ala Ala Asn Pro Ala Gly Gln Gly Glu Ser His Ser Glu Thr Ala
 1 5 10 15

Gly Glu Gly Ala Leu Ala Val Thr Ala Pro Arg Arg Pro Glu Cys Ser
 20 25 30

Val Pro Leu Val Glu Val Ser Ile Thr Glu Phe Asp Leu Ser Cys Leu
 35 40 45

Pro Glu Thr Gly Ala Val Leu Asp Gln Leu Leu Ala Leu Arg Pro Gly
 50 55 60

Gln Val Val Ile Asp Leu Ala Gly Cys Arg His Ile Asp Ala Ala Ala
 65 70 75 80

Ile Gly Leu Leu Leu Asp Val His Arg Arg Met Val Arg Ala Gly Gly
 85 90 95

Val Leu Ala Val Arg Asn Pro Asn Pro Arg Ile Thr Arg Ile Leu Gln
 100 105 110

Thr Ala Arg Leu Asp Gln Ile Leu Pro Val His Thr Asp Ala Ala Ser
 115 120 125

Ala Pro Val Thr Glu Ala Val Thr Ala Ala Pro Gly Val Pro Pro Ala
 130 135 140

Asp Gly Thr Pro Ala Pro Gly Thr Val Pro Arg Ala Ala Ala Tyr Gly
 145 150 155 160

Arg Ala Ala Val Thr Val Arg Thr
165

<210> 36

<211> 229

<212> PRT

<213> *Micromonospora sp.*, cepa ML1

<220>

<223> Proteína ORF35

<400> 36

Met Thr Ala Val Leu Val Ile Glu Asp Asp Asp Arg Ile Arg Leu Ala
1 5 10 15

Leu Leu Leu Ala Leu Glu Asp Glu Gly Tyr Glu Ala Arg Gly Ala Ala
20 25 30

Thr Ala Glu Glu Gly Leu Arg Thr Gln Arg Arg Asp Pro Ala Asp Tyr
35 40 45

Val Leu Val Asp Leu Met Leu Pro Gly Ile Asp Gly Phe Glu Gly Ile
50 55 60

Arg Gln Leu Arg Arg Asp Asp Asp Val Pro Ile Val Val Val Ser Ala
65 70 75 80

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

Asp Tyr Val Val Lys Pro Val Ala Ile Lys Glu Leu Thr Ala Arg Leu
100 105 110

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

Gln Pro Val Pro Val Leu Ala Phe Gly Asp Leu Glu Val Ser Pro Glu
130 135 140

Ala Gly Glu Val Arg Arg Ala Gly Glu Pro Val Ala Val Thr Arg Thr
145 150 155 160

Glu Phe Arg Leu Leu Cys Glu Leu Ala Glu His Ala Gly Arg Val Leu
165 170 175

Ser Arg Gln Gln Leu Leu Ser Arg Val Trp Gly Tyr Asp Ser Gly Asp
180 185 190

Glu Arg Leu Val Asp Val His Val Gly Arg Leu Arg Gln Lys Ile Glu
195 200 205

Ala Asp Pro Ala Asn Pro Arg His Leu Val Thr Leu Arg Gly Leu Gly
210 215 220

Tyr Lys Leu Gln Arg
225

<210> 37
<211> 427
<212> PRT
<213> *Micromonospora sp.*, cepa ML1

<220>
<223> Proteína ORF36

<400> 37
Met Ala Leu Val Ser Tyr Glu Leu Thr Arg Arg Thr Leu Leu Asp Glu
1 5 10 15
Arg Glu Arg Thr Val Leu Arg Ala Ala Tyr Tyr Asp Ala Ala Val Val
20 25 30
Arg Ala Gly Leu Asp Thr Asp Asn Pro Asp Val Ala Glu Ala Leu Arg
35 40 45
Ala Leu Asp Thr Gly Thr Gly Arg Arg Pro Leu Val Leu Leu Asp Gly
50 55 60
Glu Trp Tyr Ala Arg Ser Ala Asp Asn Gly Leu Thr Thr Ile Val Pro
65 70 75 80
Ala Arg Leu Gln Asp Leu Val Ala Ala Gly Glu Pro Ala Val Gln Arg
85 90 95
Val Arg Val Asn Asp Thr Pro Val Met Leu Val Gly Val Pro Leu Ser
100 105 110
Glu Ser Thr Ala Phe Tyr Glu Ile Ala Ser Leu Arg Glu Leu Glu Gln
115 120 125
Thr Phe Gln Val Leu Ala Leu Ala Leu Thr Ala Val Ala Ile Met Ile
130 135 140
Ala Gly Ser Gly Ala Ala Leu Gly Trp Tyr Ala Thr Arg His Gly Leu
145 150 155 160
Arg Pro Leu Thr Ala Val Ala Asp Ala Ala Glu Arg Ile Ala Ala Gly
165 170 175
Asp Phe Thr Thr Arg Leu Ala Pro Asp Thr Asp Pro Asp Leu Thr Arg
180 185 190
Leu Ser Thr Ser Phe Asn Arg Met Val Asp Glu Leu Ala Gly Arg Ile
195 200 205
Asp Arg Asp Arg Arg Phe Ala Ala Asp Val Ser His Glu Leu Arg Ser
210 215 220

Pro Leu Gln Thr Leu Ala Ala Ala Ala Ser Val Leu Ser Arg Arg Arg
 225 230 235 240
 Glu Asn Gln Asp Glu Arg Thr Ala Thr Ala Ala Arg Leu Val Ala Asp
 245 250 255
 Glu Ile Asp Arg Phe Gln Gln Leu Val Asn Asp Leu Leu Asp Leu Ala
 260 265 270
 Arg Gly Asp Gln Pro Val His Arg Glu Ser Val Asp Leu Pro Glu Leu
 275 280 285
 Ala Arg Gln Ala Cys Arg Asp Arg Asp Leu Pro Glu Ser Met Val Gln
 290 295 300
 Leu Ala Thr Gly Thr Pro Pro Gly Trp Arg Val Asp Arg Arg Arg Val
 305 310 315 320
 Ala Gln Val Leu Ala Asn Leu Leu Asp Asn Ala Glu Arg Tyr Gly Ala
 325 330 335
 Gly Pro Val Ala Val Arg Leu Ser His Asp Ala Asp Thr Gly Val Ile
 340 345 350
 Glu Val Asp Asp Glu Gly Pro Gly Val Pro Met Glu Asp Arg Glu Ala
 355 360 365
 Ile Phe Asp Arg Phe Val Arg Gly Arg Ala Ala Asn Tyr Arg Gly Gly
 370 375 380
 Gly Asp Gly Thr Gly Leu Gly Leu Ala Leu Val Ala Gln His Ala Ala
 385 390 395 400
 Ala His Gly Gly Asp Ala Ser Val Ser Asp Arg Pro Glu Gly Gly Ala
 405 410 415
 Arg Phe Arg Val Thr Leu Pro Gly Ser Leu Ser
 420 425

<210> 38
 <211> 178
 <212> PRT
 <213> *Micromonospora* sp., cepa ML1

<220>
 <223> Proteína ORF37

<400> 38
 Val Phe Ala Leu Leu Gly Ala Cys Gly Val Pro Ala Glu Asp Arg Pro
 1 5 10 15
 Arg Ala Val Thr Pro Pro Pro Val Pro Phe Pro Ala Thr Ala Ser Ala
 20 25 30

Asp	Asn	Gly	Ile	Gly	Met	Ser	Arg	Asp	Glu	Val	Val	Ser	Val	Ile	Gly	
				85					90					95		
Thr	Ile	Ala	Lys	Ser	Gly	Thr	Ala	Glu	Leu	Leu	Arg	Gln	Leu	Arg	Glu	
			100					105					110			
Ala	Lys	Asp	Ala	Gly	Ala	Ser	Gln	Glu	Leu	Ile	Gly	Gln	Phe	Gly	Val	
		115					120					125				
Gly	Phe	Tyr	Ala	Ala	Phe	Met	Val	Ala	Glu	Arg	Val	Glu	Leu	Val	Thr	
	130					135					140					
Arg	Arg	Ala	Gly	Glu	Thr	Gly	Gly	Thr	Arg	Trp	Glu	Ser	Thr	Gly	Glu	
145					150					155					160	
Gly	Thr	Tyr	Thr	Ile	Thr	Glu	Leu	Asp	Asp	Ala	Pro	Gln	Gly	Thr	Ala	
				165					170					175		
Val	Thr	Leu	His	Leu	Lys	Pro	Ala	Asp	Ala	Glu	Asp	Asn	Leu	His	Asp	
			180					185					190			
Tyr	Thr	Ala	Glu	Trp	Thr	Ile	Arg	Gly	Ile	Val	Lys	Arg	Tyr	Ser	Asp	
		195					200					205				
Phe	Ile	Ala	His	Pro	Ile	Arg	Met	Thr	Val	Glu	Arg	Pro	Gly	Thr	Asp	
	210					215					220					
Asp	Ala	Pro	Ala	Thr	Thr	Glu	Ala	Val	Thr	Leu	Asn	Ser	Met	Lys	Ala	
225					230					235					240	
Leu	Trp	Ala	Arg	Pro	Arg	Gly	Glu	Val	Glu	Pro	Ala	Glu	Tyr	His	Glu	
				245					250					255		
Phe	Tyr	Lys	His	Val	Ser	His	Asp	Trp	Ala	Asp	Pro	Leu	Glu	Thr	Ile	
			260					265					270			
His	Met	Arg	Gly	Glu	Gly	Thr	Phe	Glu	Tyr	Glu	Ala	Leu	Leu	Phe	Ile	
		275					280					285				
Pro	Thr	His	Ala	Pro	Leu	Asp	Leu	Phe	Ser	Pro	Gln	Gly	Arg	Arg	Gly	
	290					295						300				
Val	Gln	Leu	Tyr	Val	Lys	Arg	Val	Phe	Ile	Met	Asp	Asp	Cys	Asp	Ala	
305					310					315					320	
Leu	Met	Pro	Asn	Tyr	Leu	Arg	Phe	Val	Lys	Gly	Val	Val	Asp	Ala	His	
			325						330					335		
Asp	Leu	Ser	Leu	Asn	Ile	Ser	Arg	Glu	Ile	Leu	Gln	Gln	Asp	Arg	Gln	
			340					345					350			
Ile	Arg	Ala	Val	Arg	Arg	Arg	Leu	Val	Lys	Lys	Ile	Leu	Ala	Thr	Val	
		355					360					365				
Lys	Asp	Leu	Lys	Thr	Gly	Gln	Pro	Glu	Arg	Tyr	Arg	Thr	Phe	Trp	Thr	
	370					375					380					

Glu	Phe	Gly	Thr	Val	Val	Lys	Glu	Gly	Leu	Leu	Glu	Asp	Thr	Asp	Asn	385	390	395	400
Arg	Asp	Thr	Leu	Leu	Glu	Ile	Leu	Ser	Val	Ala	Ser	Thr	Asn	Asp	Pro	405	410	415	
Ala	Glu	Gln	Thr	Asp	Leu	Ala	Gly	Tyr	Val	Glu	Arg	Met	Arg	Glu	Gly	420	425	430	
Gln	Gln	Asp	Ile	Tyr	Tyr	Ala	Thr	Gly	Glu	Asn	Arg	Ser	Thr	Ile	Glu	435	440	445	
Asn	Ser	Pro	His	Ile	Glu	Ala	Phe	Arg	Ala	Lys	Gly	Tyr	Glu	Val	Leu	450	455	460	
Leu	Leu	Thr	Asp	Pro	Val	Asp	Glu	Val	Trp	Val	Glu	Arg	Val	Gly	Ser	465	470	475	480
Tyr	Ala	Gly	Lys	Thr	Leu	Arg	Ser	Val	Ala	Lys	Gly	Gln	Val	Asp	Leu	485	490	495	
Asp	Thr	Glu	Gln	Glu	Arg	Thr	Glu	Ala	Glu	Ala	Glu	Arg	Glu	Arg	Gln	500	505	510	
Arg	Thr	Glu	Tyr	Ala	Asp	Leu	Leu	Gly	Trp	Met	Gly	Glu	Val	Leu	Ala	515	520	525	
Asp	Asn	Val	Lys	Glu	Val	Arg	Leu	Ser	Ser	Arg	Leu	Thr	Thr	Ser	Pro	530	535	540	
Ala	Cys	Val	Val	Gly	Asp	Ala	His	Asp	Leu	Thr	Pro	Thr	Leu	Glu	Lys	545	550	555	560
Met	Tyr	Arg	Ala	Met	Gly	Gln	Glu	Val	Pro	Lys	Val	Lys	Arg			565	570		