



(19)
Bundesrepublik Deutschland
Deutsches Patent- und Markenamt

(10) **DE 697 36 673 T2** 2007.09.13

(12) **Übersetzung der europäischen Patentschrift**

(97) **EP 0 929 681 B1**

(21) Deutsches Aktenzeichen: **697 36 673.1**

(86) PCT-Aktenzeichen: **PCT/EP97/04495**

(96) Europäisches Aktenzeichen: **97 938 923.6**

(87) PCT-Veröffentlichungs-Nr.: **WO 1998/007868**

(86) PCT-Anmeldetag: **18.08.1997**

(87) Veröffentlichungstag
der PCT-Anmeldung: **26.02.1998**

(97) Erstveröffentlichung durch das EPA: **21.07.1999**

(97) Veröffentlichungstag
der Patenterteilung beim EPA: **13.09.2006**

(47) Veröffentlichungstag im Patentblatt: **13.09.2007**

(51) Int Cl.⁸: **C12N 15/52 (2006.01)**

C12P 17/18 (2006.01)

C12P 17/10 (2006.01)

C12N 9/00 (2006.01)

C12N 1/21 (2006.01)

C12N 15/70 (2006.01)

C12Q 1/68 (2006.01)

(30) Unionspriorität:

96810551 20.08.1996 EP

(73) Patentinhaber:

Novartis AG, Basel, CH

(74) Vertreter:

Spott, Weinmiller & Böhm, 80336 München

(84) Benannte Vertragsstaaten:

**AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LI,
LU, MC, NL, PT, SE**

(72) Erfinder:

**SCHUPP, Thomas, CH-4313 Möhlin, CH; TOUPET,
Christiane, F-68200 Mulhouse, FR; ENGEL,
Nathalie, F-68170 Rixheim, FR**

(54) Bezeichnung: **RIFAMYCIN BIOSYNTHESE GENKLUSTER**

Anmerkung: Innerhalb von neun Monaten nach der Bekanntmachung des Hinweises auf die Erteilung des europäischen Patents kann jedermann beim Europäischen Patentamt gegen das erteilte europäische Patent Einspruch einlegen. Der Einspruch ist schriftlich einzureichen und zu begründen. Er gilt erst als eingelegt, wenn die Einspruchsgebühr entrichtet worden ist (Art. 99 (1) Europäisches Patentübereinkommen).

Die Übersetzung ist gemäß Artikel II § 3 Abs. 1 IntPatÜG 1991 vom Patentinhaber eingereicht worden. Sie wurde vom Deutschen Patent- und Markenamt inhaltlich nicht geprüft.

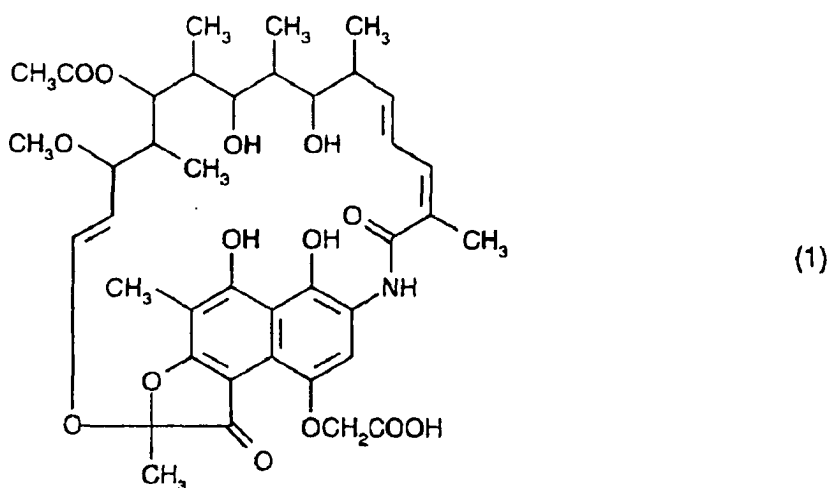
Beschreibung

[0001] Rifamycine bilden eine wichtige Gruppe an makrocyclischen Antibiotika (Wehrli, Topics in Current Chemistry (1971), 72, 21-49). Sie bestehen aus einem Naphthochinonchromophor, das durch eine lange aliphatische Brücke gespannt ist. Rifamycine gehören zur Klasse der Ansamycinantibiotika, die von mehreren Gram-positiven Bodenbakterien der Actinomycetesgruppe und einigen wenigen Pflanzen gebildet werden.

[0002] Ansamycine sind durch einen flachen aromatischen Kern gekennzeichnet, der durch eine lange aliphatische Brücke gespannt ist, welche die gegenüberliegenden Positionen des Kern verbindet. Es können zwei unterschiedliche Gruppen von Ansamycinen durch die Struktur des aromatischen Kerns unterschieden werden. Eine Gruppe besitzt einen naphthoquinoiden Chromophor, wobei typische Vertreter Rifamycin, Streptovaricin, Tolypomycin und Naphthomycin sind. Die zweite Gruppe, die einen benzochinoiden Chromophor aufweist, ist durch Geldanamycin, Maytansine und Ansamitocine gekennzeichnet (Ghisalba et al., Biotechnology of Industrial Antibiotics, E.J. Vandamme, Herausgeber, Decker Inc. New York (1984) 281-327). Im Gegensatz zu Antibiotika vom Makrolidtyp enthalten die Ansamycine im aliphatischen Ringsystem keine Lactonbindung, sondern eine Amidbindung, die die Verbindung an den Chromophor bildet.

[0003] Die Auffindung der durch den Mikroorganismus *Streptomyces mediterranei* (wie der Organismus zu der Zeit genannt wurde, siehe später) gebildeten Rifamycine wurde das erste Mal 1959 beschrieben (Sensi et al., Farmaco Ed. Sci. (1959) 14, 146-147). Die Extraktion mit Ethylacetat der angesäuerten Kulturen von *Streptomyces mediterranei* führt zur Isolierung eines Gemisches aus antibiotisch wirksamen Komponenten, den Rifamycinen A, B, C, D und E. Rifamycin B, die stabilste Komponente, wurde von den anderen Komponenten abgetrennt und aufgrund der stark sauren Eigenschaften und der leichten Salzbildung von den anderen Komponenten abgetrennt.

[0004] Rifamycin B hat die Struktur der Formel (1)



[0005] Rifamycin B ist die Hauptkomponente der Fermentation, wenn Barbiturat zum Fermentationsmedium gegeben wird und/oder verbesserte Produktionsmutanten von *Streptomyces mediterranei* verwendet werden.

[0006] Der Rifamycinproduktionsstamm wurde ursprünglich als *Streptomyces mediterranei* klassifiziert (Sensi et al., Farmaco Ed. Sci. (1959) 14, 146-147). Eine Analyse der Zellwand von *Streptomyces mediterranei* durch Thiermann et al. zeigte später, dass dieser Stamm eine Zellwand aufweist, die für *Nocardia* typisch ist und der Stamm wurde als *Nocardia mediterranei* reklassifiziert (Thiemann et al., Arch. Microbiol. (1969), 67, 147-151). *Nocardia mediterranei* wurde erneut auf der Grundlage von neueren, akuraten morphologischen und biochemischen Kriterien reklassifiziert. Auf der Grundlage der exakten Zusammensetzung der Zellwand, dem Fehlen von Mycolsäure und der Unempfindlichkeit gegenüber Phagen von *Nocardia* und *Rhodococcus* wurde der Stamm der neuen Gattung *Amycolatopsis* als *Amycolatopsis mediterranei* zugeordnet (Lechevalier et al., Int. Syst. Bacteriol. (1986), 36, 29). Lab et al. (Crit. Rev. Microbiol. (1995), 21, 19-30) haben Verfahren zur Verbesserung der Rifamycinbildung durch *Amycolatopsis mediterranei* zusammengefasst.

[0007] Rifamycine haben eine starke antibiotische Aktivität hauptsächlich gegen Gram-positive Bakterien, wie Mycobakterien, Neisseria und Staphylococci. Der bakterizide Effekt von Rifamycinen leitet sich von der spezifischen Hemmung der bakteriellen DNA-abhängigen RNA Polymerase ab, die die RNA Biosynthese unterbricht (Wehrli und Staehlin, Bacteriol. Rev. (1971), 35, 290-309). Das semisynthetische Rifamycin B Derivat

Rifampin (Rifampicin) wird klinisch breit als Antibiotikum gegen den Erreger verwendet, der Tuberkulose verursacht, nämlich *Mycobacterium tuberculosis*.

[0008] Die naphthochinoiden Ansamycine der Streptovaricin- und Tolypomycingruppe zeigen, wie Rifamycin, einen antibakteriellen Effekt durch die Hemmung der bakteriellen RNA Polymerase. Im Gegensatz dazu hat Naphthomycin einen antibakteriellen Effekt ohne die Hemmung der bakteriellen RNA Polymerase. Die Benzo-chinoidansamycine zeigen keine Hemmung der bakteriellen RNA Polymerase und haben daher nur eine relativ schwache antibakterielle Aktivität, falls überhaupt. Andererseits haben einige Vertreter dieser Substanzklasse eine Wirkung auf eukaryontische Zellen. So wurden Antipilz-, Antiprototoen- und Antitumoreigenschaften für Geldanamycin beschrieben. Andererseits werden den Maytansinen antimitotische (Antitubulin), antileukämische und antitumorale Eigenschaften zugeschrieben. Einige Rifamycine zeigen auch eine Antitumor- und Antiviralaktivität, aber nur in hohen Konzentrationen. Der biologische Effekt scheint so unspezifisch zu sein.

[0009] Trotz der großen strukturellen Varietät der Ansamycine scheint ihre Biosynthese durch einen metabolischen Weg stattzufinden, der viele gemeinsame Elemente enthält (Ghisalba et al. *Biotechnology of Industrial Antibiotics*, E. J. Vandamme, Herausgeber, Decker Inc. New York (1984), 281-327). Der aromatische Kern für alle Ansamycine wird möglicherweise ausgehend von 3-Amino-5-hydroxybenzoesäure aufgebaut. Ausgehend von diesem Molekül, das vermutlich als Coenzym A aktiviert wird, wird die gesamte aliphatische Brücke durch eine multifunktionale Polyketidsynthase synthetisiert. Die Länge der Brücke und die Prozessierung der Keto-gruppen, die anfänglich durch die Kondensationsschritte gebildet werden, werden durch die Polyketidsynthase kontrolliert. Um die vollständige aliphatische Brücke für Rifamycine aufzubauen, sind 10 Kondensationsschritte, 2 mit Acetat und 8 mit Propionat als Bausteine erforderlich. Die Sequenz dieser einzelnen Kondensationsschritte wird ähnlich durch die Polyketidsynthase bestimmt. Strukturvergleiche und Untersuchungen mit dem Einbau von radioaktivem Acetat und Propionat haben gezeigt, dass die Sequenz aus Acetat- und Propionateinbau für die verschiedenen Ansamycine gemäß einem Schema stattfindet, das in den ersten Kondensationsschritten identisch oder sehr ähnlich zu sein scheint. Daher zweigt die Synthese der verschiedenen Ansamycine von einem gemeinsamen Syntheschema der Ansamycinpolyketidsynthasen (dem Rifamycinsyntheschema) früher oder später gemäß ihrer Strukturunterschiede von der Rifamycinstruktur in die Seitenäste der Synthese ab (Ghisalba et al., *Biotechnology of Industrial Antibiotics*, E. J. Vandamme, Dekker Inc. New York (1984), 281-327).

[0010] Aufgrund der großen strukturellen Vielzahl der Rifamycine und ihrem spezifischen und interessanten biologischen Effekt besteht ein großes Interesse, die genetische Basis ihrer Synthese zu verstehen, um die Möglichkeit zu generieren, diese spezifisch zu beeinflussen. Dies ist insbesondere erwünscht, da, wie oben erklärt, viele Gemeinsamkeiten zwischen der Synthese der Rifamycine und der von anderen Ansamycinen bestehen. Diese Ähnlichkeit in der Biosynthese, die sich möglicherweise von einem gemeinsamen evolutorischen Ursprung dieses Biosynthesewegs ableitet, hat naturgemäß eine genetische Basis.

[0011] Die genetische Basis der Biosynthese der Sekundärmetaboliten existiert im wesentlichen in den Genen, die für die einzelnen Biosynthesenzyme kodieren und in den regulatorischen Elementen, die die Expression der Biosynthesegene kontrollieren. Die Synthesegene der sekundären Metaboliten der Actinomyceten wurden in allen untersuchten Systemen als Cluster von benachbarten Genen gefunden. Die Größe solcher antibiotischer Gencluster erstreckt sich von etwa 10 Kilobasen (kb) bis mehr als 100 kb. Die Cluster enthalten oft spezifische Regulatorgene und Gene für die Resistenz des Produktionsorganismus gegenüber dem eigenen Antibiotikum (Chater, Ciba Found. Symp. (1992), 171, 144-162).

[0012] Die hierin beschriebene Erfindung hat es nun durch die Identifizierung und Klonierung von Genen der Rifamycinbiosynthese geschafft, die genetische Basis zur Synthese von Rifamycinanaloga oder neuen Ansamycinen durch genetische Verfahren zu schaffen, die Strukturelemente von Rifamycin mit anderen Ansamycinen kombiniert. Dies schafft auch die Basis zur Herstellung von neuen Substanzsammlungen auf der Basis der Rifamycinbiosynthesegencluster durch kombinatorische Biosynthese.

[0013] Es war in einem ersten Schritt möglich, ein DNA Fragment aus dem Genom von *A. mediterranei* zu identifizieren und zu klonieren, das eine Homologie mit bekannten Polyketidsynthasegenen zeigt. Nach dem Erhalt der Sequenzinformation aus dem DNA Fragment, die eine typische Sequenz für Polyketidsynthasen bestätigt, war es möglich, eine Cosmidgenbank von *A. mediterranei* mit spezifischen DNA Sonden, die sich von diesem Fragment ableiten, in einem Screeningprogramm für weitere DNA Fragmente zu screenen, die im Rifamycincluster beteiligt sind. Als Ergebnis wird das vollständige Rifamycinpolyketidsynthasegencluster identifiziert und einer Sequenzbestimmung unterzogen (siehe SEQ ID Nr. 3). Das Gencluster umfasst 6 offene Leserahmen, die hierin als ORF A, B, C, D, E und F bezeichnet werden und die für die in SEQ ID Nr. 4 bis 9

gezeigten Proteine und Polypeptide kodieren.

[0014] Das auf diese Weise isolierte und charakterisierte Gencluster stellt beispielsweise die Basis für eine gezielte Optimierung der Produktion von Rifamycin, Ansamycinen oder Analoga hiervon dar. Beispiele für Techniken und mögliche Anwendungsgebiete, die in diesem Zusammenhang erhältlich sind, sind folgende:

- Überexpression von einzelnen Genen in Produktionsstämmen mit Plasmidvektoren oder durch den Einbau in das Chromosom
- Untersuchung der Expression und Transkriptionsregulation des Genclusters während der Fermentation mit verschiedenen Produktionsstämmen und die Optimierung hiervon durch physiologische Parameter und geeignete Fermentationsbedingungen
- Identifizierung der regulatorischen Gene und der DNA Bindungsstellen der entsprechenden regulatorischen Proteine im Gencluster. Die Charakterisierung des Effekts dieser regulatorischen Elemente bei der Produktion von Rifamycinen oder Ansamycinen und deren Beeinflussung durch spezifische Mutation dieser Gene oder der DNA Bindungsstellen
- Duplikation des kompletten Genclusters oder von Teilen hiervon in Produktionsstämmen. Neben diesen Anwendungen des Genclusters zur Verbesserung der Produktion durch Fermentation, wie dies oben beschrieben ist, kann es ähnlich zur biosynthetischen Herstellung von neuen Rifamycinanaloga oder neuen Ansamycinen oder Ansamycin-ähnlichen Verbindungen verwendet werden, worin die aliphatische Brücke nur an einem Ende des aromatischen Kerns gebunden ist. Die folgenden Möglichkeiten kommen hierfür in Betracht, wie beispielsweise:
 - Inaktivierung der einzelnen Schritte der Biosynthese, beispielsweise durch Genzerstörung
 - Mutation von einzelnen Schritten in der Biosynthese, beispielsweise durch Genaustausch. Verwendung des Clusters oder von Fragmenten hiervon als DNA Sonde, um andere natürliche Mikroorganismen zu isolieren, die einzelne Metaboliten produzieren, die zu Rifamycin oder Ansamycinen ähnlich sind.
 - Austausch von einzelnen Elementen in diesem Gencluster durch die von anderen Genclustern.
 - Verwendung von modifizierten Polyketidsynthasen zum Aufbau von Genbanken für verschiedene Rifamycinanaloga oder Ansamycine, die dann auf ihre Aktivität getestet werden (Jackie & Khosala, Chemistry & Biology (1995), 2, 355-362).
 - Konstruktion von mutierten Actinomycetenstämmen, aus denen das natürliche Rifamycin- oder Ansamycinbiosynthesegencluster im Chromosom teilweise oder vollständig deletiert wurde und die so zur Expression von genetisch modifizierten Genclustern verwendet werden können.
- Austausch von einzelnen Elementen innerhalb des Genclusters

Detaillierte Beschreibung der Erfindung

[0015] Die Erfindung betrifft ein DNA Fragment aus dem Genom von *Amiclatopsis mediterranei*, das eine DNA Region umfasst, die direkt oder indirekt im Gencluster beteiligt ist, das für die Rifamycinsynthese verantwortlich ist, und die benachbarten DNA Regionen und funktionellen Bestandteile oder Domänen hiervon.

[0016] Die erfindungsgemäßen DNA Fragmente der Erfindung können darüber hinaus regulatorische Sequenzen umfassen, wie Promotoren, Repressor- oder Aktivatorbindungsstellen, Repressor- oder Aktivatorgene, Terminatoren oder Strukturgene. Ebenfalls sind ein Teil der Erfindung alle Kombinationen dieser DNA Fragmente miteinander oder mit anderen DNA Fragmenten, beispielsweise Kombinationen von Promotoren, Repressor- oder Aktivatorbindungsstellen und/oder Repressor oder Aktivatorgenen aus einem Ansamycingencluster, insbesondere aus dem Rifamycingencluster mit fremden Strukturgenen oder Kombinationen von Strukturgenen aus dem Ansamycingencluster, speziell dem Rifamycingencluster mit fremden Promotoren und Kombinationen von Strukturgenen miteinander oder mit Genfragmenten, die für enzymatisch aktive Domänen kodieren und von verschiedenen Ansamycinbiosynthesesystemen stammen. Fremde Strukturgene und fremde Genfragmente, die für enzymatisch aktive Domänen kodieren, kodieren beispielsweise für Proteine, die bei der Biosynthese von anderen Ansamycinen beteiligt sind.

[0017] Ein bevorzugtes DNA Fragment ist eines, das direkt oder indirekt im Gencluster beteiligt ist, das für die Rifamycinsynthese verantwortlich ist.

[0018] Das Gencluster oder die DNA Region, die oben beschrieben sind, enthält beispielsweise die Gene, welche für die einzelnen Enzyme kodieren, die bei der Biosynthese von Ansamycinen beteiligt sind, insbesondere von Rifamycin, und die regulatorischen Elemente, die die Expression der Biosynthesegene kontrollieren. Die Größe solcher antibiotischen Gencluster erstreckt sich von etwa 10 Kilobasen (kb) bis über 100 kb. Die Gencluster umfassen normalerweise spezifische regulatorische Gene und Gene für die Resistenz des Produktionsorganismus gegenüber seinem eigenen Antibiotikum. Beispiele, was durch Enzyme oder enzymatisch ak-

tive Domänen gemeint ist, die bei dieser Biosynthese beteiligt sind, sind die, welche zur Synthese der Ansamycine, wie Rifamycin ausgehend von 3-Amino-5-hydroxybenzoesäure erforderlich sind, beispielsweise Polyketidsynthasen, Acyltransferasen, Dehydratasen, Ketoreduktasen, Acylcarrierproteine oder Ketoacylsynthasen.

[0019] Daher ist die vollständige Sequenz des Genclusters, die in SEQ ID Nr. 3 gezeigt ist, wie auch DNA Fragmente, die Sequenzteile enthalten, welche für eine Polyketidsynthase oder eine enzymatisch aktive Domäne hiervon kodieren, besonders bevorzugt. Beispiele für solche bevorzugten DNA Fragmente sind beispielsweise jene, die für ein oder mehrere Proteine und Polypeptide kodieren, die in den SEQ ID Nr. 4, 5, 6, 7, 8 und 9 gezeigt sind oder funktionelle Derivate hiervon, die beispielsweise 15 oder mehr aufeinanderfolgende Nukleotide umfassen. Andere bevorzugte Ausführungsformen umfassen DNA Regionen des Genclusters gemäß der Erfindung oder von Fragmenten hiervon, wie jene, die in den hinterlegten Klonen pNE95, pRi44-2 und pNE112 vorkommen oder hiervon abgeleitet sind. Weitere bevorzugte DNA Fragmente sind jene, die Sequenzteile umfassen, die Homologien mit den Sequenzen zeigen, die in den Klonen pNE95, pRi44-2 und/oder pNE112 enthalten sind oder mit SEQ ID Nr. 1 und/oder 3 und können daher als Hybridisierungssonde innerhalb einer genomischen Genbank von Ansamycin-, insbesondere Rifamycin-bildendem Organismus verwendet werden, um Bestandteile des entsprechenden Genclusters aufzufinden. Das DNA Fragment kann darüberhinaus beispielsweise ausschließlich genomische DNA umfassen. Ein besonders bevorzugtes DNA Fragment ist eines, das die in SEQ ID Nr. 1 oder 3 gezeigte Nukleotidsequenz oder partielle Sequenzen hiervon umfasst, das aufgrund der Homologien als Struktur- oder Funktionsäquivalent zu dieser Sequenz oder der partiellen Sequenz hiervon betrachtet werden kann und das daher zur Hybridisierung mit dieser Sequenz fähig ist.

[0020] Die erfindungsgemäßen DNA Fragmente umfassen beispielsweise Sequenzteile, die Homologien mit dem oben beschriebenen Enzymen, Enzymdomänen oder Fragmenten hiervon aufweisen.

[0021] Der Ausdruck Homologien und Struktur- und/oder Funktionsäquivalente bezieht sich primär auf DNA und Aminosäuresequenzen mit wenigen oder minimalen Differenzen zwischen den relevanten Sequenzen. Diese Unterschiede können sehr unterschiedliche Ursachen haben. Dies kann beispielsweise Mutationen oder Stamm-spezifische Unterschiede umfassen, die natürlich vorkommen oder künstlich induziert sind. Oder die aus der gegenüber der anfänglichen Sequenz beobachteten Unterschiede stammen von einer gezielten Modifikation, die beispielsweise während einer chemischen Synthese eingeführt werden kann.

[0022] Funktionale Unterschiede können als minimal betrachtet werden, falls beispielsweise die Nukleotidsequenz, die für ein Polypeptid oder eine Proteinsequenz kodiert, im wesentlichen dieselben charakteristischen Eigenschaften aufweist, wie die anfängliche Sequenz, ob in Bezug auf die enzymatische Aktivität, die immunologische Reaktivität oder im Fall einer Nukleotidsequenz, in Bezug auf die Genregulation.

[0023] Strukturunterschiede können als minimal betrachtet werden, solange eine signifikante Überlappung oder Ähnlichkeit zwischen den verschiedenen Sequenzen besteht oder sie haben zumindest ähnliche physikalische Eigenschaften. Die letzteren umfassen beispielsweise die elektrophoretische Mobilität, chromatographische Ähnlichkeiten, Sedimentationskoeffizienten, spektrophotometrische Eigenschaften usw.

[0024] Im Fall der Nukleotidsequenzen sollte die Übereinstimmung mindestens 70 %, aber vorzugsweise 80 % und vor allem 90 % oder mehr betragen. Im Fall der Aminosäuresequenz sollten die entsprechenden Zahlen zumindest 50 %, aber vorzugsweise 60 % und vor allem 70 % sein. 90 % Übereinstimmung ist äußerst bevorzugt.

[0025] Die Erfindung betrifft ferner ein Verfahren zur Identifizierung, Isolierung und Klonierung eines der oben beschriebenen DNA Fragmente. Ein bevorzugtes Verfahren umfasst beispielsweise die folgenden Schritte:

- a) Aufbau einer genomischen Genbank,
- b) Screenen dieser Genbank mit Hilfe der erfindungsgemäßen DNA Sequenzen, und
- c) Isolierung der als positiv identifizierten Klone.

[0026] Ein allgemeines Verfahren zur Identifizierung von DNA Fragmenten, die bei der Biosynthese von Ansamycinen beteiligt sind, umfasst beispielsweise die folgenden Schritte:

- 1) Klonierung eines DNA Fragments, das eine Homologie mit bekannten Polyketidsynthasegenen zeigt
 - a) Das Vorkommen von DNA Fragmenten, die eine Homologie mit den erfindungsgemäßen Polyketidsynthasegenen aufweisen, wird in den zu untersuchenden Stämmen durch ein Southernexperiment mit der chromosomalen DNA dieses Stamms detektiert. Die Größe solcher homologer DNA Fragmente kann durch den Verdau der DNA mit einem geeigneten Restriktionsenzym bestimmt werden.

b) Herstellung einer Plasmidgenbank, die die oben verdauten chromosomalen Fragmente umfasst. Normalerweise werden einzelne Klone dieser Genbank erneut auf eine Homologie mit den Polyketidsynthasegenen gemäß der Erfindung getestet. Die Klone mit rekombinanten Plasmiden, die Fragmente umfassen, welche eine Homologie mit der Polyketidsonde aufweisen, werden dann normalerweise auf der Grundlage dieser Homologie isoliert.

2) Analyse der klonierten Region

a) Restriktionsanalyse der isolierten rekombinanten Plasmide und Überprüfung der Identität dieser klonierten Fragmente miteinander.

b) Durch einen chromosomalen Southern Blot mit DNA des ursprünglichen Mikroorganismus und des isolierten DNA Fragments als Sonde kann gezeigt werden, dass das klonierte Fragment ein ursprüngliches chromosomales DNA Fragment aus dem ursprünglichen Mikroorganismus ist.

c) Es ist als Option möglich, eine signifikante Homologie des klonierten DNA Fragments mit einer chromosomalen DNA aus anderen Ansamycinproduzenten zu zeigen (Streptovaricin, Tolypomycin, Geldanamycin, Ansamitocin). Dies würde bestätigen, dass die klonierte DNA für die Gencluster der Ansamycinbiosynthese und so auch der Rifamycinbiosynthese typisch ist.

d) DNA Sequenzierung als internes Restriktionsfragment und Demonstration durch vergleichende Sequenzanalyse, dass die klonierte Region eine typische DNA Sequenz von Polyketidsynthasen ist, die für die Biosynthese von Polyketidantibiotika aus Actinomyceten kodiert.

3) Isolierung und Charakterisierung von benachbarten DNA Regionen

a) Konstruktion einer Kosmidgenbank aus dem ursprünglichen Mikroorganismus und Analyse hiervon auf Homologie mit den isolierten Fragmenten. Isolierung von Cosmiden, die eine Homologie mit diesem Fragment aufweisen.

b) Demonstration durch Restriktionsanalyse, dass die isolierten Cosmidklone eine DNA Region des ursprünglichen Mikroorganismus umfassen, die mit dem ursprünglichen Fragment überlappt.

[0027] Wie oben beschrieben, ist der erste Schritt bei der Isolierung der erfindungsgemäßen DNA Fragmente normalerweise das Erstellen von genomischen Genbanken aus dem Organismus von Interesse, der das erforderliche Ansamycin, speziell Rifamycin synthetisiert.

[0028] Die genomische DNA kann aus einem Wirtsorganismus auf verschiedenen Wegen erhalten werden, beispielsweise durch Extraktion aus der Kernfraktion und Reinigung der extrahierten DNA durch bekannte Verfahren.

[0029] Die Fragmentierung, die zur Erstellung einer repräsentativen Genbank erforderlich ist, der zu klonierenden genomischen DNA auf eine Größe, die zur Insertion in einem Klonierungsvektor geeignet ist, kann entweder durch mechanische Scherkräfte oder sonstiges, vorzugsweise durch Schneiden mit geeigneten Restriktionsenzymen stattfinden.

[0030] Geeignete Klonierungsvektoren, die sich bereits in einer Routineverwendung zur Herstellung von genomischen Genbanken befinden, sind beispielsweise Cosmidvektoren, Plasmidvektoren oder Phagenvektoren.

[0031] Es ist dann in einem Screeningprogramm möglich, geeignete Klone zu erhalten, die die erforderlichen Gene oder Genfragmente aus den auf diese Weise hergestellten Genbanken umfassen.

[0032] Eine Möglichkeit zur Identifizierung der erforderlichen DNA Region besteht beispielsweise durch die Verwendung der oben beschriebenen Genbank zur Transformation von Stämmen, die aufgrund eines blockierten Synthesewegs nicht mehr zur Bildung von Ansamycinen fähig sind, und der Identifizierung der Klone, die nach der Transformation erneut zur Bildung von Ansamycin fähig sind (Revertanden). Die Vektoren, die zu Revertanden führen, umfassen ein DNA Fragment, das für eine Ansamycinsynthese erforderlich ist.

[0033] Eine weitere Möglichkeit zur Identifizierung der erforderlichen DNA Region basiert beispielsweise auf der Verwendung von geeigneten Sondenmolekülen (DNA Sonde), die beispielsweise wie oben beschrieben erhalten werden. Es sind verschiedene Standardverfahren zur Identifizierung geeigneter Klone erhältlich, wie Differentialkoloniehybridisierung oder Plaquehybridisierung.

[0034] Es ist möglich, als Sondenmolekül ein vorher isoliertes DNA Fragment aus demselben oder einem strukturell verwandten Gen oder Gencluster zu verwenden, das aufgrund der vorhandenen Homologien zur Hybridisierung mit dem entsprechenden Sequenzabschnitt innerhalb des erforderlichen Gens oder Genclusters fähig ist, das identifiziert werden soll. Vorzugsweise wird als Sondenmolekül für die Zwecke der vorliegen-

den Erfindung ein DNA Fragment verwendet, das aus einem Gen oder einer DNA Sequenz erhältlich ist, die bei der Synthese von Polyketiden beteiligt ist, wie Ansamycinen oder Soraphenen.

[0035] Falls die Nukleotidsequenz des zu isolierenden Gens oder zumindest Teile dieser Sequenz bekannt sind, ist es in einer alternativen Ausführungsform möglich, aufgrund dieser Sequenzinformation eine entsprechend synthetisierte DNA Sequenz für die Hybridisierungen oder PCR Amplifikationen zu verwenden.

[0036] Um die Detektierbarkeit des erforderlichen Gens oder Teile des erforderlichen Gens zu erleichtern, kann eine der oben beschriebenen DNA Sondenmoleküle mit einer geeigneten, leicht detektierbaren Gruppe markiert werden. Eine detektierbare Gruppe meint für die Zwecke der Erfindung jedes Material, das bestimmte, leicht detektierbare physikalische oder chemische Eigenschaft aufweist.

[0037] An dieser Stelle können enzymatisch aktive Gruppen besonders erwähnt werden, wie Enzyme, Enzymsubstrate, Coenzyme und Enzyminhibitoren, ferner fluoreszente und lumineszente Mittel, Chromophore und Radioisotope, wie ^3H , ^{35}S , ^{32}P , ^{125}I und ^{14}C . Die leichte Detektierbarkeit dieser Marker basiert andererseits auf ihren intrinsischen physikalischen Eigenschaften (beispielsweise Fluoreszenzmarkern, Chromophoren, Radioisotopen) oder andererseits auf ihren Reaktions- und Bindeeigenschaften (beispielsweise Enzyme, Substrate, Coenzyme, Inhibitoren). Materialien dieser Typen werden breit insbesondere in Immuntests verwendet und können in den meisten Fällen in der vorliegenden Erfindung verwendet werden.

[0038] Allgemeine Methoden, die sich auf die DNA Hybridisierung beziehen, sind beispielsweise von T. Maniatis et al., Molecular Cloning, Cold Spring Harbor Laboratory Press (1982) beschrieben.

[0039] Die Klone innerhalb der vorher beschriebenen Genbanken, die zur Hybridisierung mit einem Sondenmolekül fähig sind und die durch eine der oben erwähnten Detektionsmethoden identifiziert werden können, können dann weiter analysiert werden, um das Ausmaß und die Art der kodierenden Sequenz im Detail zu bestimmen.

[0040] Ein alternatives Verfahren zur Identifizierung klonierter Gene basiert auf der Konstruktion einer Genbank, die aus Plasmid- oder Expressionsvektoren besteht. Das heißt in Analogie zu den vorher beschriebenen Verfahren, dass die genomische DNA, die das erforderliche Gen enthält, anfänglich isoliert und dann in einen geeigneten Plasmid- oder Expressionsvektor kloniert wird. Die auf diese Weise hergestellten Genbanken können dann durch geeignete Verfahren gecreent werden, beispielsweise durch die Verwendung von Komplementationstudien und die Klone, die das erforderliche Gen oder zumindest einen Teil dieses Gens als Insert enthalten, können selektiert werden.

[0041] Daher ist es mit Hilfe der oben beschriebenen Verfahren möglich, ein Gen, mehrere Gene oder ein Gencluster zu isolieren, das für ein oder mehrere bestimmte Genprodukte kodiert.

[0042] Für die weitere Charakterisierung werden die DNA Sequenzen auf die oben beschriebene Weise gereinigt und isoliert und einer Restriktionsanalyse und Sequenzanalyse unterzogen.

[0043] Für die Sequenzanalyse werden die vorher isolierten DNA Fragmente zuerst mittels geeigneter Restriktionsenzyme fragmentiert und dann in geeignete Klonierungsvektoren kloniert. Um Fehler bei der Sequenzierung zu vermeiden, ist es vorteilhaft, beide DNA Stränge vollständig zu sequenzieren.

[0044] Es sind verschiedene Alternativen zur Analyse des klonierten DNA Fragments in Bezug auf dessen Funktion innerhalb der Ansamycinbiosynthese erhältlich.

[0045] So ist es beispielsweise möglich in Komplementationsexperimenten mit defekten Mutanten nicht nur die prinzipielle Beteiligung eines Gens oder Genfragments bei der Biosynthese sekundärer Metaboliten zu zeigen, sondern auch spezifisch den Syntheseschritt zu verifizieren, in dem dieses DNA Fragment beteiligt ist.

[0046] In einem alternativen Analyseverfahren wird die Evidenz auf genau die entgegengesetzte Weise erhalten. Ein Transfer von Plasmiden, die DNA Abschnitte enthalten, welche Homologien mit geeigneten Abschnitten auf dem Genom enthalten, führt zur Integration der homologen DNA Abschnitte über eine homologe Rekombination. Wenn wie im vorliegenden Fall der homologe DNA Abschnitt eine Region innerhalb eines offenen Leserahmens eines Genclusters ist, führt die Plasmidintegration zu einer Inaktivierung des Gens durch eine sogenannte Genzerstörung und daher zu einer Unterbrechung der Produktion des sekundären Metaboliten. Es wird derzeit angenommen, dass eine homologe Region, die zumindest 100 bp, aber vorzugsweise

mehr als 1000 bp aufweist, ausreicht, um das erforderliche Rekombinationsereignis herbeizuführen.

[0047] Jedoch ist eine homologe Region bevorzugt, die sich über einen Bereich von 0,3 bis 4 kb, insbesondere über einen Bereich von 1 bis 3 kb erstreckt.

[0048] Um geeignete Plasmide herzustellen, die eine ausreichende Homologie für die Integration über homologe Rekombination aufweisen, geht vorzugsweise ein Subklonierungsschritt voraus, worin die vorher isolierte DNA verdaut wird und Fragmente mit geeigneter Größe isoliert und anschließend in ein geeignetes Plasmid kloniert werden. Beispiele für geeignete Plasmide sind die Plasmide, die allgemein für genetische Manipulationen in Streptomycceten oder E. coli verwendet werden.

[0049] Es ist im Prinzip möglich, zur Herstellung und Vermehrung der vorher beschriebenen Konstrukte alle herkömmlichen Klonierungsvektoren, wie Plasmid- oder Bakteriophagenvektoren zu verwenden, solange sie Replikations- und Kontrollsequenzen enthalten, die von einer Spezies stammen, welche mit der Wirtszelle kompatibel ist.

[0050] Der Klonierungsvektor weist gewöhnlich einen Replikationsursprung plus spezifische Gene auf, die zu phänotypischen Selektionsmerkmalen in der transformierten Wirtszelle führen, insbesondere Resistenzen gegenüber Antibiotika. Die transformierten Vektoren können auf der Grundlage dieser phänotypischen Marker nach einer Transformation in eine Wirtszelle selektiert werden.

[0051] Selektierbare phänotypische Marker, die für die Zwecke der Erfindung verwendet werden können, umfassen beispielsweise ohne Beschränkung des Gegenstands der Erfindung Resistenzen gegenüber Thiostrepton, Ampicillin, Tetracyclin, Chloramphenicol, Hygromycin, G418, Kanamycin, Neomycin und Bleomycin. Ein weiterer Selektionsmarker kann beispielsweise eine Prototrophie insbesondere für Aminosäuren sein.

[0052] Hauptsächlich bevorzugt für die Zwecke der vorliegenden Erfindung sind Streptomycceten und E. coli Plasmide, beispielsweise die Plasmide, die für die Zwecke der vorliegenden Erfindung verwendet werden.

[0053] Wirtszellen, die primär für die vorher beschriebene Klonierung für die Zwecke der Erfindung geeignet sind, sind Prokaryonten, einschließlich bakterielle Wirte, wie Streptomycceten, Actinomyceten, E. coli oder Pseudomonaden.

[0054] E. coli Wirte sind besonderes bevorzugt, beispielsweise der E. coli Stamm HB101 oder X-1 Blue MR® (Stratagene) oder Streptomycceten, wie die plasmidfreien Stämme von Streptomyces lividans TK23 und TK24.

[0055] Kompetente Zellen des E. coli Stamms HB101 werden durch die Verfahren hergestellt, die normalerweise zur Transformation von E. coli verwendet werden. Die Transformationsmethode von Hopwood et al. (Genetic manipulation of streptomyces, A laboratory manual. The John Innes Foundation, Norwich (1985)) wird normalerweise für Streptomyces verwendet.

[0056] Nach der Transformation und anschließenden Inkubation auf einem geeigneten Medium werden die entstehenden Kolonien einem Differentialscreening durch Plattieren auf Selektivmedien unterzogen. Es ist dann möglich, die geeignete Plasmid DNA aus diesen Kolonien zu isolieren, die Plasmide mit hierin klonierten DNA Fragmenten umfassen.

[0057] Das erfindungsgemäße DNA Fragment, das eine DNA Region umfasst, die direkt oder indirekt bei der Biosynthese von Ansamycin beteiligt ist und auf die vorher beschriebene Weise aus dem Ansamycinbiosynthesegencluster erhalten werden kann, kann auch als Ausgangsklon zur Identifizierung und Isolierung von anderen benachbarten Regionen verwendet werden, die hiermit aus diesem Gencluster überlappen.

[0058] Dies kann beispielsweise durch die Ausführung eines so genannten „Chromosomwalking“ innerhalb einer Genbank, die aus DNA Fragmenten mit gemeinsam überlappenden DNA Regionen besteht, unter Verwendung des vorher isolierten DNA Fragments oder ähnlichem insbesondere der an den 5' und 3' Enden liegenden Sequenzen ausgeführt werden. Die Verfahren für ein Chromosomwalking sind dem Fachmann bekannt. Details finden sich beispielsweise in den Publikationen von Smith et al. (Methods Enzymol. (1987), 151, 461-489) und Wahl et al (Proc. Natl. Acad. Sci, USA (1987), 84, 2160-2164).

[0059] Die Voraussetzung für das Chromosomwalking ist das Vorkommen von Klonen mit kohärenten DNA Fragmenten, die so lange wie möglich sind und gemeinsam innerhalb einer Genbank überlappen und ein ge-

eigneter Ausgangsklon, der ein Fragment umfasst, das in der Nähe liegt oder dergleichen, vorzugsweise in der zu analysierenden Region. Falls die exakte Lage des Ausgangsklons unbekannt ist, wird das Walking vorzugsweise in beide Richtungen ausgeführt.

[0060] Der eigentliche Walkingschritt beginnt durch die Verwendung des identifizierten und isolierten Ausgangsklons als Sonde in einer der vorher beschriebenen Hybridisierungsreaktionen, um benachbarte Klone zu detektieren, die mit dem Ausgangsklon überlappende Regionen aufweisen. Es ist durch Hybridisierungsanalyse möglich, festzustellen, welches Fragment am weitesten über die überlappende Region hinausgeht. Dies wird dann als Ausgangsklon für den 2. Walkingschritt verwendet, wobei ein Fragment etabliert wird, das mit dem 2. Klon in dieselbe Richtung überlappt. Ein kontinuierliches Fortfahren auf diese Weise auf dem Chromosom führt zu einer Sammlung aus überlappenden DNA Klonen, die eine große DNA Region abdecken. Diese kann dann, wo dies geeignet ist, nach einem oder mehreren Subklonierungsschritten durch bekannte Verfahren unter Bildung eines Fragments zusammenligiert werden, das Teile oder dergleichen umfasst, vorzugsweise alle Bestandteile, die für eine Ansamycinbiosynthese essentiell sind.

[0061] Die Hybridisierungsreaktion zur Etablierung von Klonen mit marginal überlappenden Regionen verwendet vorzugsweise nicht das sehr große und fast vollständige Fragment, sondern an dessen Stelle ein partielles Fragment aus der linken oder rechten Randregion, die in einem Subklonierungsschritt erhalten werden kann. Aufgrund der geringeren Größe des partiellen Fragments führt die Hybridisierungsreaktion zu weniger positiven Hybridisierungssignalen, so dass der analytische Aufwand deutlich geringer ist, wie wenn man das gesamte Fragment verwendet. Es ist ferner ratsam, das partielle Fragment im Detail zu analysieren, um auszuschließen, dass es größere Mengen an repetitiven Sequenzen umfasst, die über das gesamte Genom verteilt sein können und dies so eine gezielte Abfolge von Walkingschritten stark überschreiten würde.

[0062] Da das Gencluster, das für die Ansamycinbiosynthese verantwortlich ist, eine relativ große Region des Genoms umfasst, kann es auch vorteilhaft sein, ein sogenanntes Largestep-Walking oder Cosmidwalking auszuführen. Es ist in diesen Fällen unter Verwendung von Cosmidvektoren, die die Klonierung von sehr großen DNA Fragmenten erlauben, möglich, eine sehr große DNA Region abzudecken, die in einem einzigen Walkingschritt bis zu 42 kb umfassen kann.

[0063] In einer möglichen Ausführungsform der vorliegenden Erfindung wird beispielsweise zur Konstruktion einer Cosmidgenbank von Streptomyces oder Actinomycetes die gesamte DNA isoliert, wobei die Größe der DNA Fragmente im Bereich von etwa 100 kb liegt, und anschließend partiell mit geeigneten Restriktionsendonukleasen verdaut.

[0064] Die verdaute DNA wird dann auf herkömmliche Weise extrahiert, um die Endonuklease zu entfernen, die immer noch vorhanden ist und gefällt und schließlich konzentriert. Das entstehende Fragmentkonzentrat wird dann gemäß der Größe der einzelnen Fragmente fraktioniert, beispielsweise durch Dichtegradientenzentrifugation. Nachdem die auf diese Weise erhaltenen Fragmente dialysiert wurden, können sie auf einem Agarosegel analysiert werden. Die Fraktionen, welche Fragmente einer geeigneten Größe enthalten, werden gepoolt und für eine weitere Verarbeitung konzentriert. Die Fragmente, die als teilweise für die Zwecke der Erfindung geeignet betrachtet werden können, haben eine Größe von 30 kb bis 42 kb, aber vorzugsweise 35 kb bis 40 kb.

[0065] Parallel zur oben beschriebenen Fragmentierung oder später wird beispielsweise ein geeigneter Cosmidvektor pWE15[®] (Stratagene) vollständig mit einem geeigneten Restriktionsenzym, beispielsweise BamHI, für die anschließende Ligasereaktion verdaut.

[0066] Die Ligation der Cosmid DNA an die Streptomyces oder Actinomycetes Fragmente, die gemäß ihrer Größe fraktioniert wurden, kann mittels einer T4 DNA Ligase ausgeführt werden. Das auf diese Weise erhältliche Ligationsgemisch wird nach einer ausreichenden Inkubationszeit durch allgemein bekannte Verfahren in λ Phagen gepackt.

[0067] Die entstehenden Phagenpartikel werden dann zur Infektion eines geeigneten Wirtsstamms verwendet. Ein recA⁻ E. coli Stamm ist bevorzugt, wie E. coli/HB 101 oder X-1 Blue[®] (Stratagene). Die Selektion von transfizierten Klonen und die Isolierung der Plasmid DNA kann durch allgemein bekannte Verfahren ausgeführt werden.

[0068] Das Screenen der Genbank auf DNA Fragmente, die bei der Ansamycinbiosynthese beteiligt sind, wird beispielsweise mittels einer spezifischen Hybridisierungssonde ausgeführt, von der angenommen wird

(beispielsweise auf der Grundlage der DNA Sequenz oder DNA Homologie oder Komplementationstests oder Genzerstörung oder der Funktion hiervon in anderen Organismen), dass sie DNA Regionen aus dem Ansamycinclustern umfasst.

[0069] Ein Plasmid, das ein zusätzliches Fragment der erforderlichen Größe aufweist oder auf der Grundlage von Hybridisierungen identifiziert wurde, kann dann aus dem Gel auf die vorher beschriebene Weise isoliert werden. Die Identität dieses zusätzlichen Fragments mit dem erforderlichen Fragment auf dem vorher selektierten Cosmid kann dann durch Southerntransfer und Hybridisierung bestätigt werden.

[0070] Die Funktionsanalyse der auf diese Weise isolierten DNA Fragmente kann in einem Genzerstörungsexperiment ausgeführt werden, wie dies oben beschrieben ist.

[0071] Eine andere mögliche Verwendung der erfindungsgemäßen DNA Fragmente ist, Enzyme oder Domänen, die bei der Ansamycin- und insbesondere Rifamycinbiosynthese beteiligt sind, zu modifizieren oder zu inaktivieren, um Oligonukleotide zu synthetisieren, die dann wiederum zur Auffindung von homologen Sequenzen bei der PCR Amplifizierung verwendet werden können.

[0072] Neben den erfindungsgemäßen DNA Fragmenten als solche, wird auch ihre Verwendung vor allem zur Herstellung von Rifamycin, Rifamycinanaloga oder Vorläufern hiervon und zur biosynthetischen Produktion von neuen Ansamycinen oder Vorläufern hiervon beansprucht. In diesem Zusammenhang enthalten sind jene Moleküle, bei denen die aliphatische Brücke nur an einem Ende des aromatischen Nukleus gebunden ist.

[0073] Die erfindungsgemäßen DNA Fragmente erlauben beispielsweise durch die Kombination von DNA Fragmenten aus anderen Biosynthesewegen oder durch Inaktivierung oder Modifikation hiervon die Biosynthese von neuen Hybridverbindungen, insbesondere von neuen Ansamycin- oder Rifamycinanaloga. Die Schritte, die hierfür erforderlich sind, sind im allgemeinen bekannt und beispielsweise beschrieben in Hopwood, Current Opinion in Biotechnol. (1993), 4, 531-537.

[0074] Die Erfindung betrifft ferner die Verwendung der DNA Fragmente der Erfindung zur Ausführung der neuen Technologie der kombinatorischen Biosynthese zur biosynthetischen Herstellung von Banken aus Polyketidsynthasen auf der Grundlage der Rifamycin- und Ansamycinbiosynthesegene. Falls beispielsweise mehrere Sätze an Modifikationen hergestellt werden, ist es auf diese Weise auch möglich durch Biosynthesen eine Bank aus Polyketiden, beispielsweise Ansamycin- oder Rifamycinanaloga herzustellen, die dann nur auf die Aktivität der auf diese Weise hergestellten Verbindungen getestet werden muss. Die hierfür erforderlichen Schritte sind im allgemeinen bekannt und beispielsweise in Tsoi und Khosla, Chemistry & Biology (1995), 2, 355-362 und WO 95 08 548 A beschrieben.

[0075] Neben dem DNA Fragment als solches wird auch dessen Verwendung zur genetischen Konstruktion von mutierten Actinomycetenstämmen beansprucht, aus denen die natürlichen Rifamycin- oder Ansamycinbiosynthesegengcluster im Chromosom teilweise oder vollständig deletiert wurden und die so zur Expression von genetisch modifizierten Ansamycin- oder Rifamycinbiosynthesegengclustern verwendet werden können.

[0076] Die Erfindung betrifft ferner einen Hybridvektor, der zumindest ein erfindungsgemäßes DNA Fragment umfasst, beispielsweise einen Promotor, eine Repressor- oder Aktivatorbindungsstelle, ein Repressor- oder Aktivatorgen, ein Strukturgen, einen Terminator oder einen funktionellen Teil hiervon.

[0077] Der Hybridvektor umfasst beispielsweise eine Expressionskassette, die ein erfindungsgemäßes DNA Fragment umfasst, das zur Expression von einem oder mehreren Proteinen fähig ist, die bei der Ansamycinbiosynthese beteiligt sind, insbesondere in der Rifamycinbiosynthese oder ein funktionelles Fragment hiervon. Die Erfindung betrifft in ähnlicher Weise einen Wirtsorganismus, der den oben beschriebenen Hybridvektor umfasst.

[0078] Geeignete Vektoren, die den Ausgangspunkt der erfindungsgemäßen Hybridvektoren darstellen, und geeignete Wirtsorganismen, wie Bakterien oder Hefezellen, sind allgemein bekannt.

[0079] Der Wirtsorganismus kann durch im allgemeinen herkömmliche Verfahren transformiert werden, wie durch Protoplasten, Ca^{2+} , Cs^+ , Polyethylenglycol, Elektroporation, Viren, Lipidvesikel oder einer Partikelkanone. Die erfindungsgemäßen DNA Fragmente können dann sowohl als extrachromosomale Bestandteile im Wirtsorganismus vorkommen als auch über geeignete Sequenzabschnitte in das Chromosom des Wirtsorganismus integriert werden.

[0080] Die Erfindung betrifft auch Polyketidsynthasen, die die erfindungsgemäßen DNA Fragmente umfassen, insbesondere die von *Amycolatopsis mediterranei*, die direkt oder indirekt bei der Rifamycinsynthese beteiligt sind, und funktionelle Bestandteile hiervon, beispielsweise enzymatisch aktive Domänen.

[0081] Die Erfindung betrifft ferner eine Hybridisierungssonde, die ein erfindungsgemäßes DNA Fragment umfasst und die Verwendung hiervon, insbesondere zur Identifizierung von DNA Fragmenten, die bei der Biosynthese von Ansamycinen beteiligt sind.

[0082] Um unzweifelhafte Signale bei der Hybridisierung zu erhalten, wird die DNA an das Filter (beispielsweise hergestellt aus Nylon oder Nitrocellulose) gebundene DNA bei 55-65°C in 0,2 × SSC (1 × SSC = 0,15 M Natriumchlorid, 15 mM Natriumcitrat) gewaschen.

Beispiele

Allgemein

[0083] Allgemeine molekulargenetische Techniken, wie DNA Isolierung und Reinigung, Restriktionsverdau von DNA, Agarosegelelektrophorese von DNA Ligation von Restriktionsfragmenten, Kultivierung und Transformation von *E. coli* und Plasmidisolierung aus *E. coli* werden ausgeführt, wie dies in Maniatis et al., Molecular Cloning: A laboratory manual, 1. Ausgabe, Cold Spring Harbor Laboratory Press, Cold Spring Harbor NY (1982) beschrieben ist.

[0084] Die Kulturbedingungen und molekulargenetischen Techniken mit *A. mediterranei* und anderen Actinomyceten werden ausgeführt, wie dies von Hopwood et al. beschrieben ist (Genetic manipulation of streptomycetes a laboratory manual, The John Innes Foundation, Norwich, 1985). Alle flüssigen Kulturen von *A. mediterranei* und anderen Actinomyceten werden in Erlenmeyerkolben bei 28°C in einem Schüttler bei 250 Upm ausgeführt.

Verwendete Medien:

LB: Maniatis et al., Molecular Cloning: A laboratory manual, 1. Ausgabe, Cold Spring Harbor Laboratory Press, Cold Spring Harbor NY (1982)

NL: Schupp + Divers FEMS Microbiology Lett. 36, 159-162 (1986) (NL 148 = NL148G ohne Glycin)

R2YE: Hopwood et al. (Genetic manipulation of streptomycetes a laboratory manual. The John Innes Foundation, Norwich, 1985).

TB: 12 g/l Bacto Trypton

24 g/l Bacto Hefeextrakt

4 ml/l Glycerin

Beispiel 1: Detektion von chromosomalen DNA Fragmenten von *A. mediterranei* mit einer Homologie zu Polyketidsynthasegenen anderer Bakterien

[0085] Um eine genomische DNA als *A. mediterranei* zu erhalten, werden Zellen des Stamms *A. mediterranei* WT3136 (=LBGA 3136, ETH Stammsammlung) in NL 148 Medium für 48 Stunden kultiviert. 1 ml dieser Kultur wird dann in 50 ml NL 148 Medium (+2,5 g/l Glycin) in einem 200 ml Erlenmeyerkolben überführt und die Kultur wird für 48 Stunden inkubiert. Die Zellen werden aus dem Medium durch Zentrifugation bei 3000 × g für 10 Minuten entfernt und in 5 ml SET (75 mM NaCl, 25 mM EDTA, 20 mM Tris, pH 7,5) resuspendiert. Es wird eine hochmolekulare DNA durch das Verfahren von Pospiech und Neumann (Trends in Genetics (1995), 11, 217-218) extrahiert.

[0086] Um einzelne Fragmente aus der isolierten *A. mediterranei* DNA, die eine Homologie mit Polyketidsynthasegenen aufweisen, mittels Southern Blot zu detektieren, wird eine radioaktive DNA Sonde aus einem bekannten Polyketidsynthasegencluster hergestellt. Um dies auszuführen, wird das PvuI Fragment mit einer Größe von 3,8 kb aus dem rekombinanten Plasmid p98/1 (Schupp et al., J. of Bacteriol. (1995), 177, 3673-3679) isoliert, das eine DNA Region mit einer Größe von etwa 32 kb aus der Polyketidsynthase für das Antibiotikum Soraphen A umfasst. Es werden etwa 0,5 µg des isolierten 3,8 kb PvuI DNA Fragments mit ³²P-dCTP durch das Nicktranslationssystem von Gibco/BRL (Basel) gemäß den Anleitungen des Herstellers radioaktiv markiert.

[0087] Für den Southern Blot werden etwa 2 µg der genomischen DNA, die oben aus *A. mediterranei* isoliert

wurde, vollständig mit dem Restriktionsenzym BglII (Boehringer Mannheim) verdaut und die entstehenden Fragmente werden auf einem 0,8 % Agarosegel fraktioniert. Ein Southern Blot mit diesem Agarosegel und der oben isolierten DNA Sonde (3,8 kb PvuI Fragment) detektiert ein BglII geschnittenes DNA Fragment mit einer Größe von etwa 13 kb aus der genomischen DNA von *A. mediterranei*, das eine Homologie mit der verwendeten DNA Sonde aufweist. Es kann auf der Grundlage dieser Homologie geschlossen werden, dass das detektierte DNA Fragment aus *A. mediterranei* eine genetische Region ist, die für eine Polyketidsynthase kodiert und so bei der Synthese eines Polyketidantibiotikums beteiligt ist.

Beispiel 2: Herstellung einer spezifischen rekombinanten Plasmidsammlung, die BglII verdaute chromosomale Fragmente von *A. mediterranei* mit einer Größe von 12-16 kb umfasst

[0088] Der *E. coli* Positivselektionsvektor pIJ4642 (Derivat von pIJ666, Kieser & Melton, Gene (1988), 65, 83-91), der am John Innes Centre (Norwich, UK) entwickelt wurde, wird zur Herstellung einer Plasmidgenbank verwendet. Dieses Plasmid wird zuerst mit BamHI geschnitten und die zwei entstehenden Fragmente werden auf einem Agarosegel fraktioniert. Das kleinere der zwei Fragmente ist das Füllfragment des Vektors und das größere ist der Vektorteil, der bei einer Selbstligation nach der Deletion des Füllfragments aufgrund der flankierenden Terminationssequenzen ein perfektes Palindrom bildet, was bedeutet, dass das Plasmid als solches nicht in *E. coli* gehalten werden kann. Der Vektorteil mit einer Größe von 3,8 kb wird aus dem Agarosegel durch Elektroelution isoliert, wie dies auf Seite 164-165 von Maniatis et al., Molecular Cloning: A laboratory manual, 1. Ausgabe, Cold Spring Harbor Laboratory Press, Cold Spring Harbor NY (1982) beschrieben ist.

[0089] Um die mit BglII geschnittenen DNA Fragmente von *A. mediterranei* herzustellen, wird die hochmolekulare, genomische DNA verwendet, die in Beispiel 1 hergestellt wurde. Etwa 10 µg dieser DNA werden vollständig mit dem Restriktionsenzym BglII verdaut und anschließend auf einem 0,8 % Agarosegel fraktioniert. Die DNA Fragmente mit einer Größe von etwa 12-16 kb werden aus dem Gel ausgeschnitten und aus dem Gelblock durch Elektroelution entfernt (siehe oben). Etwa 1 µg der auf diese Weise isolierten BglII Fragmente wird an etwa 0,1 µg des oben isolierten BamHI Teils des Vektors pIJ4642 ligiert. Das auf diese Weise erhaltene Ligationsgemisch wird dann in den *E. coli* Stamm HB 101 (Stratagene) transformiert. Es werden etwa 150 transformierte Kolonien aus dem Transformationsgemisch auf LB Agar mit 30 µg pro ml Chloramphenicol selektiert. Diese Kolonien enthalten rekombinante Plasmide mit BglII-geschnittenen genomischen DNA Fragmenten von *A. mediterranei* im Größenbereich von 12 bis 16 kb.

Beispiel 3: Klonierung und Charakterisierung der chromosomalen *A. mediterranei* DNA Fragmente mit einer Homologie zu bakteriellen Polyketidsynthasegenen

[0090] 150 der Plasmidklone, die in Beispiel 2 hergestellt wurden, werden durch Koloniehybridisierung mittels eines Nitrocellulosefilters (Schleicher & Schuell) analysiert, wie dies auf den Seiten 318-319 von Maniatis et al., Molecular Cloning: A laboratory manual, 1. Ausgabe, Cold Spring Harbor Laboratory Press, Cold Spring Harbor NY (1982) beschrieben ist. Die verwendete DNA Sonde ist das 3,8 kb PvuI Fragment von Plasmid p98/1, das mit ³²P-dCTP radioaktiv markiert ist und in Beispiel 1 isoliert wurde. Die Plasmide werden aus 5 Plasmidklonen isoliert, die ein Hybridisierungssignal zeigen und werden durch zwei Restriktionsverdaus mit den Enzymen HindIII oder KpnI charakterisiert. HindIII schneidet zweimal im Vektorteil der Klone, 0,3 kb rechts und links der BamHI Spaltstelle, in die die *A. mediterranei* DNA integriert wurde. KpnI schneidet nicht im pIJ4642 Vektorteil. Diese Restriktionsanalyse zeigt, dass die untersuchten Klone sowohl identische HindIII Fragmente mit etwa 14 und 3,1 kb und identische KpnI Fragmente mit einer Größe von etwa 11,4 kb und 5,7 kb aufweisen. Dies zeigt, dass diese Klone dasselbe genomische BglII Fragment von *A. mediterranei* umfassen und dass das letzte eine Größe von etwa 13 kb aufweist. Es kann zusätzlich aus dieser Restriktionsanalyse geschlossen werden, dass das klonierte BglII Fragment keine interne HindIII Schnittstelle aufweist, sondern 2 KpnI Schnittstellen aufweist, die ein internes KpnI Fragment mit einer Größe von 5,7 kb ergeben.

[0091] Die Plasmid DNA der obigen 5 Klone mit identischen Restriktionsfragmenten wird ferner durch einen Southern Blot charakterisiert. Für diesen Zweck werden die Plasmide mit HindIII und KpnI geschnitten und die verwendete DNA Sonde ist das mit ³²P radioaktiv markierte 3,8 kb PvuI Fragment von Plasmid p98/1, das oben verwendet wird. Dieses Experiment bestätigt, dass die 5 Plasmide identische *A. mediterranei* DNA Fragmente enthalten und dass diese eine signifikante Homologie mit der DNA Sonde aufweisen, die für bakterielle Polyketidsynthasegene charakteristisch ist. Zusätzlich zeigt der Southern Blot, dass das interne KpnI Fragment mit einer Größe von 5,7 kb auch eine signifikante Homologie mit der verwendeten DNA Sonde aufweist. Das pRI7-3 genannte Plasmid wird aus den 5 Plasmiden zur weiteren Prozessierung ausgewählt.

[0092] Um zu demonstrieren, dass das klonierte BglII Fragment mit einer Größe von etwa 13 kb von *A. me-*

diterranei ein ursprüngliches chromosomales DNA Fragment ist, wird ein weiterer Southern Blot aus geführt. Chromosomale DNA aus *A. mediterranei*, die mit BglII, KpnI oder BamHI geschnitten wurde, wird in diesem Blot verwendet. Es werden zwei BamHI Fragmente, die etwa 1,8 und 1,9 kb groß sind und im 5,7 kb KpnI Fragment von pRi7-3 vorkommen, als radioaktiv markierte Sonde verwendet. Dieses Experiment bestätigt, dass das BglII DNA Fragment mit einer Größe von etwa 13 kb, das in das rekombinante Plasmid pRi7-3 kloniert wurde, ein authentisches genomisches DNA Fragment von *A. mediterranei* ist. Zusätzlich bestätigt dieses Experiment, dass das klonierte Fragment ein internes KpnI Fragment mit einer Größe von 5,7 kb und zwei BamHI Fragment mit einer Größe von 1,8 und 1,9 kb umfasst und dass diese DNA Fragmente ebenfalls authentische genomische DNA Fragmente von *A. mediterranei* sind.

Beispiel 4: Demonstration einer signifikanten Homologie mit dem klonierten genomischen 13 kb BglII Fragment von *A. mediterranei* mit chromosomaler DNA von anderen Actinomyceten, die Ansamycine bilden

[0093] Es findet eine Demonstration einer signifikanten Homologie zwischen der klonierten chromosomalen DNA Region von *A. mediterranei* und der chromosomalen DNA von anderen Ansamycin-bildenden Actinomyceten durch ein Southern Blot Experiment statt. Es werden die folgenden Ansamycin-bildenden Stämme für diesen Zweck verwendet (die durch diese Stämme gebildeten Ansamycine sind in Klammern angegeben): *Streptomyces spectabilis* (Streptovaricine), *Streptomyces tolypophorus* (Tolypomycine), *Streptomyces hygroscopicus* (Geldanamycine), *Nocardia species ATCC 31281* (Ansamitocine). Die genomische DNA aus diesen Stämmen wird wie in Beispiel 1 für *A. mediterranei* beschrieben isoliert und mit dem Restriktionsenzym KpnI verdaut und die auf diese Weise erhaltenen Restriktionsfragmente werden auf einem Agarosegel für den Southern Blot fraktioniert. Zwei BamHI Fragmente mit einer Größe von etwa 1,8 und 1,9 kb von *A. mediterranei*, die in Beispiel 3 verwendet und aus dem Plasmid pRi7-3 isoliert wurden, werden als radioaktive Sonde verwendet. Dieses Experiment zeigt, dass diese Ansamycinbildenden Stämme eine signifikante DNA Homologie mit der verwendeten DNA Sonde und somit mit der klonierten chromosomalen Region von *A. mediterranei* aufweisen. Es wird in diesem Zusammenhang beobachtet, dass die Homologie im Fall von Ansamycinbildnern mit einem Naphthochinoidringsystem (Streptovaricin, Tolypomycin) größer ist als bei denen mit einem Benzochinoidringsystem (Geldanamycin, Ansamitocin). Dieses Ergebnis legt nahe, dass die klonierte chromosomale DNA Region von *A. mediterranei* für Ansamycinbiosynthesecenter und speziell von Genclustern von Ansamycinen mit Naphthochinoidringsystemen typisch ist, die den Ringsystemen der Rifamycine entsprechen.

Beispiel 5: DNA Sequenzbestimmung des KpnI Fragments mit einer Größe von 5,7 kb, das innerhalb des klonierten 13 kb BglII Fragments liegt

[0094] Zur Sequenzierung wird das 5,7 kb KpnI Fragment aus dem Plasmid pRi7-3 (DSM 11114) (Maniatis et al. 1992) isoliert und in die KpnI Schnittstelle des Vektors pBRKanf4 subkloniert, der für die DNA Sequenzierung geeignet ist, was die Plasmide pTS004 und pTS005 ergibt. Der Vektor pBRKanf4 (abgeleitet von pBRKanf1, Bhat, Gene (1993) 134, 83-87) ist zur Erführung von sequenziellen Deletionen von Sau3A Fragmenten im klonierten Insertfragment geeignet, da dieser Vektor selbst keine GATC Nukleotidsequenz aufweist. Zusätzlich werden die BamHI Fragmente mit einer Größe von 1,9 und 1,8 kb, die im 5,7 kb KpnI Fragment vorkommen, in die BamHI Schnittstelle von pBRKanf4 subkloniert, was jeweils zu den Plasmiden pTS006 und pTS007 und pTS008 und pTS009 führt.

[0095] Um Subklone herzustellen, die sequenziell durch Sau3A Fragmente für die DNA Sequenzierung verkürzt sind, werden die Plasmide pTS004 bis pTS009 partiell mit Sau3A verdaut und vollständig mit XbaI oder HindIII (einer Schnittstelle in der Mehrfachklonierungsregion des Vektors) verdaut. Die auf diese Weise erhaltene DNA (besteht aus dem linearisierten Vektor mit inserierten DNA Fragmenten, die um Sau3A Fragmente verkürzt sind) wird an den Enden mittels Klenowpolymerase (Fragment der Polymerase I, siehe Maniatis et al., Seiten 113-114) aufgefüllt, mit T4 DNA Ligase selbstligiert und in *E. coli* DH5a transformiert. Die Plasmid DNA, die den Plasmiden pTS004 bis pTS009 entspricht, aber DNA Regionen enthält, die von einer Seite um Sau3A Fragmente verkürzt sind und aus den ursprünglich integrierten Fragmenten von *A. mediterranei* abgeleitet sind, wird aus den auf diese Weise erhaltenen einzeln transformierten Klonen isoliert.

[0096] Die DNA Sequenzierung wird mit den auf diese Weise erhaltenen Plasmiden und mit pTS004 bis pTS009 mittels des Reaktionskits von Perkin-Elmer/Applied Biosystems mit Farbstoff-markierten Terminatorreagenzien (Kit Nr. 402122) und einem Universalprimer oder einem T7 Primer ausgeführt. Ein Standardzyklussequenzierungsprotokoll mit einem Thermocycler (MJ Research DNA Engine Thermocycler, Modell 225) wird verwendet und die Sequenzierungsreaktionen werden durch das automatische DNA Sequenziergerät von Applied Biosystems (Modell 373 oder 377) gemäß den Anleitungen des Herstellers analysiert. Um diese Ergebnisse zu analysieren werden die folgenden Computerprogramme (Software) verwendet: Applied Biosystems DNA

Analysesoftware, Unix Solaris CDE Software, DNA Zusammensetzungs- und Analysepaket GAP, das von R. Staden (Nucleic Acid Research (1995) 23, 1406-1410) und Blast (NCBI) lizenziert ist.

[0097] Die oben beschriebenen Verfahren können zur vollständigen Sequenzierung beider DNA Stränge des 5,7 kb KpnI Fragments von *A. mediterranei* Stamm Wt3136 verwendet werden. Die DNA Sequenz des 5,7 kb Fragments mit einer Länge von 5676 Basenpaaren ist in SEQ ID Nr. 1 gezeigt.

Beispiel 6: Analyse der Protein-kodierenden Regionen (Gene) auf dem 5,7 kb KpnI Fragment von *A. mediterranei*

[0098] Die Nukleotidsequenz des 5,7 kb KpnI Fragments wird mittels des Codonpreference Computerprogramms (Genetics Computer Group, University of Wisconsin, 1994) analysiert. Diese Analyse zeigt, dass dieses Fragment über die gesamte Länge eine Protein-kodierende Region aufweist und so einen größeren Teil des offenen Leserahmens (ORF) bildet. Die verwendeten Codons in diesem ORF sind für Gene von Streptomyces und Actinomycetes typisch. Die Aminosäuresequenz, die von der DNA Sequenz dieses ORF abgeleitet ist, ist in SEQ ID Nr. 2 gezeigt.

[0099] Die Polyketidsynthasen für Makrolidantibiotika (wie Erythromycin, Rapamycin) sind sehr große multifunktionelle Proteine, die mehrere enzymatisch aktive Domänen aufweisen, die jetzt gut charakterisiert sind (Hopwood und Khosala, Ciba Foundation Symposium (1992), 171, 88-112, Donadio und Katz, Gene (1992), 111, 51-60, Schwecke et al., Proc. Natl. Acad. Sci. USA (1995) 92 (17), 7839-7843). Der Vergleich der in SEQ ID Nr. 2 gezeigten Aminosäuresequenz mit der sehr gut charakterisierten Erythromycinpolyketidsynthase eryA ORF1 (Donadio, Science, (1991) 252, 675-679, DNA Sequenz Gen/EMBL Hinterlegungsnummer M63676) ergibt die folgenden Ergebnisse:

Die Region von SEQ ID Nr. 2, Aminosäuren 2-325 ist zu 40 % identisch zu der Acyltransferasedomäne von Modul 2 des eryA Lokus von *Saccharopolyspora erythraea*.

Die Region von SEQ ID Nr. 2, Aminosäuren 325-470 ist zu 43 % identisch zu der Dehydratasedomäne von Modul 4 des eryA Lokus von *Saccharopolyspora erythraea*.

Die Region von SEQ ID Nr. 2, Aminosäuren 762-940 ist zu 48 % identisch zu der Ketoreduktasedomäne von Modul 2 des eryA Lokus von *Saccharopolyspora erythraea*.

Die Region von SEQ ID Nr. 2, Aminosäuren 1024-1109 ist zu 57 % identisch zu der Acycarrierproteindomäne von Modul 2 des eryA Lokus von *Saccharopolyspora erythraea*.

Die Region von SEQ ID Nr. 2, Aminosäuren 1126-1584 ist zu 59 % identisch zu der Ketoacylsynthasedomäne von Modul 2 des eryA Lokus von *Saccharopolyspora erythraea*.

[0100] Die sehr großen Ähnlichkeiten, die in der Aminosäuresequenz und der Größe und Zusammensetzungen der enzymatischen Domänen gefunden werden, legen nahe, dass die klonierte KpnI Region mit einer Größe von 5,7 kb von *A. mediterranei* für einen Teil einer Polyketidsynthase kodiert, die für Polyketide des Makrolidtyps typisch ist.

Beispiel 7: Konstruktion der Cosmidgenbank von *A. mediterranei*

[0101] Der verwendete Cosmidvektor ist das Plasmid pWE15, das gekauft werden kann (Stratagene, La Jolla, CA, USA). pWE15 wird vollständig mit dem Enzym BamHI geschnitten (Maniatis et al., 1989) und mit Ethanol gefällt. Zur Ligation an die Cosmid DNA wird die chromosomale DNA aus *A. mediterranei* wie in Beispiel 1 beschrieben isoliert und partiell mit dem Restriktionsenzym Sau3A (Boehringer Mannheim) unter Bildung von DNA Fragmenten verdaut, von denen die meisten eine Größe von 20 bis 40 kb aufweisen. Die auf diese Weise vorbehandelte DNA wird nach Fragmentgröße durch Zentrifugation ($83\,000 \times g$, 20°C) auf einem 10 % bis 40 % Saccharosedichtegradienten für 18 Stunden fraktioniert. Der Gradient wird in 0,5 ml Aliquots fraktioniert und dialysiert und Proben mit 10 µl werden auf einem 0,3 % Agarosegel mit einem DNA Größenstandard analysiert. Fraktionen mit einer chromosomalen DNA mit einer Größe von 25-40 kb werden vereinigt, mit Ethanol gefällt und in einem kleinen Volumen Wasser resuspendiert.

[0102] Die Ligation der Cosmid DNA an die *A. mediterranei* Sau3A Fragmente, die gemäß ihrer Größe (siehe oben) isoliert wurden, findet mit Hilfe einer T4 DNA Ligase statt. Etwa 3 µg jeder der zwei DNA Ausgangsmaterialien werden in einem Reaktionsvolumen von 20 µl verwendet und die Ligation wird bei 12°C für 15 Stunden ausgeführt. 4 ml dieses Ligationsgemisches werden in Lambdaphagen mittels des in vitro Verpackungskits verpackt (gemäß den Angaben des Herstellers), der von Stratagene (La Jolla, CA, USA) bezogen wird. Die entstehenden Phagen werden durch Infektion in den *E. coli* Stamm X-1 BlueMR® (Stratagene) eingeführt. Die Titration dieses Phagenmaterials ergibt etwa 20 000 Phagenpartikel pro ml, wobei eine Analyse von 12 Cosmid-

klonen zeigt, dass alle Klone Plasmid DNA Insertionen mit einer Größe von 25 bis 40 kb enthalten.

Beispiel 8: Identifizierung, Klonierung und Charakterisierung der chromosomalen *A. mediterranei* DNA Region, die benachbart zum klonierten 5,7 kb KpnI Fragment liegt

[0103] Um die chromosomale *A. mediterranei* DNA Region zu identifizieren und zu klonieren, die benachbart zum oben in den Beispielen 3 und 5 beschriebenen 5,7 kb KpnI Fragment liegt, wird zuerst eine radioaktive DNA Sonde aus diesem 5,7 kb KpnI Fragment hergestellt. Dies wird durch radioaktive Markierung von etwa 0,5 µg des isolierten DNA Fragments mit ³²P-dCTP durch das Nicktranslationssystem von Gibco/BRL (Basel) gemäß den Anleitungen des Herstellers ausgeführt.

[0104] Die Infektion von *E. coli* X-1 Blue MR (Stratagene) mit einem Aliquot der in vitro verpackten Lambda-phagen (siehe Beispiel 7) führt zu mehr als 2000 Klonen auf mehreren LB + Ampicillin (50 µg/ml) Platten. Diese Klone werden durch Koloniehybridisierung auf Nitrocellulosefilter (siehe Beispiel 3 bezüglich des Verfahrens) getestet. Die verwendete DNA Sonde ist das 5,7 kb KpnI DNA Fragment von *A. mediterranei*, die mit ³²P-dCTP radioaktiv markiert ist und oben hergestellt wurde.

[0105] Es werden 5 Cosmidklone gefunden, die signifikantes Signal mit der DNA Sonde zeigen. Die Plasmid DNA dieser Cosmide wird isoliert (Sambrook et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989), mit KpnI verdaut und auf einem Agarosegel analysiert. Eine Analyse zeigt, dass alle 5 Plasmide eine integrierte chromosomale *A. mediterranei* DNA mit einer Größe in der Größenordnung von etwa 25-35 kb aufweisen und alle das 5,7 kb KpnI Fragment enthalten.

[0106] Um die chromosomale *A. mediterranei* DNA Region zu charakterisieren, die benachbart zum klonierten KpnI Fragment liegt, wird die Plasmid DNA eines der 5 Cosmidklone einer Restriktionsanalyse unterzogen. Das ausgewählte Plasmid des Cosmidklons hat die Nummer pNE112 und umfasst ebenfalls das 13 kb BglII Fragment, das in Beispiel 3 beschrieben wurde.

[0107] Der Verdau des Plasmids pNE112 mit den Restriktionsenzymen BamHI, BglII, HindIII (einzeln und in Kombination) erlaubt die Erstellung einer Restriktionskarte der klonierten Region von *A. mediterranei* und dies erlaubt die Charakterisierung dieser etwa 26 kb großen Region im Chromosom von *A. mediterranei*. Diese Region wird durch die folgenden Restriktionsschnittstellen mit dem angegebenen Abstand in kb von einem Ende charakterisiert: BamHI in Position 3,2 kb, HindIII in Position 6,6 kb, BglII in Position 11,5 kb, BamHI in Position 16,6 kb, BamHI in Position 17,3 kb, BamHI in Position 21 kb und BglII in Position 24 kb.

Beispiel 9: Bestimmung der Sequenz der chromosomalen *A. mediterranei* DNA Region, die im Plasmid pNE112 vorkommt und mit dem klonierten 5,7 kb KpnI Fragment überlappt

[0108] Die Plasmid pNE112 DNA wird in die Fragmente mittels eines Aero-Mist Verneblers (CIS-US Inc., Bedford, MA, USA) unter einem Stickstoffdruck von 8-12 Pfund pro Quadratinch gespalten. Diese zufälligen DNA Fragmente werden mit T4 DNA Polymerase, T4 DNA Kinase und *E. coli* DNA Polymerase in Gegenwart der 4 dNTPs behandelt, um stumpfe Enden an den doppelsträngigen DNA Fragmenten zu erzeugen (Sambrook et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor, Laboratory Press, Cold Spring Harbor, NY, 1989). Die Fragmente werden dann in 0,8 % bei niedrigen Temperaturen schmelzender Agarose (FMC Sea Plaque Agarose, Katalognummer 50113) fraktioniert und es werden Fragmente mit einer Größe von 1,5-2 kb durch heiße Phenolextraktion extrahiert (Sambrook et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989). Die auf diese Weise erhaltenen DNA Fragmente werden dann mit Hilfe der T4 DNA Ligase an den Plasmidvektor pBRKanf4 (siehe Beispiel 5) oder pBlueScript KS+ (Stratagene, La Jolla, CA, USA) ligiert, der jeweils mit stumpfen Enden durch einen geeigneten Restriktionsverdau geschnitten wird (SmaI für pBRKanf4 und EcoRV für pBlueScript KS+) und an den Enden durch die Behandlung mit alkalischer Phosphatase dephosphoryliert (Boehringer Mannheim). Das Ligationsgemisch wird dann in *E. coli* DH5a transformiert und die Zellen werden über Nacht auf LB Agar mit dem geeigneten Antibiotikum (Kanamycin 40 µg/ml für pBRKanf4, Ampicillin 100 µg/ml für pBlueScript KS+) inkubiert. Die gewachsenen Kolonien werden einzeln in 1,25 ml flüssiges TB Medium mit Antibiotikum in Platten mit 96 Vertiefungen und Vertiefungen mit einem Volumen von 2 ml überführt und bei 37°C über Nacht inkubiert. Die Matrizen DNA für die Sequenzierung wird direkt aus diesen Kulturen durch alkalische Lyse (Birnboim, Methods in Enzymology (1983) 100, 243-255) hergestellt. Die DNA Sequenzierung findet mittels des Perkin Elmer/Applied Biosystems Reaktionskits mit Farbstoff-markierten Terminatorreagenzien (Kit Nr. 402122) und Universalprimern M13 mp 18/19 oder T3, T7 Primern oder mit Primern statt, die hergestellt wurden, um an interne Sequenzen zu binden. Ein Standardzyklusprotokoll mit 20 Zyklen wird mittels eines Thermocyclers (MJ

Research DNA Engine Thermocycler, Modell 225) verwendet. Die Sequenzierungsreaktionen werden mit Ethanol gefällt, in Formamidauftragspuffer resuspendiert und fraktioniert und analysiert durch eine Elektrophorese unter Verwendung des automatischen Applied Biosystems DNA Sequenziergeräts (Modell 377) gemäß den Anleitungen des Herstellers. Es werden Sequenzdateien mit Hilfe des Applied Biosystems DNA Analyse Software Computerprogramms erstellt und auf einen SUN UltraSpark Computer zur weiteren Analyse überführt. Die folgenden Computerprogramme (Software) werden zur Analyse der Ergebnisse verwendet: DNA Zusammensetzungs- und Analysepaket GAP (Genetics Computer Group, University of Wisconsin, R. Staden, Cambridge University, UK) und die vier Programme: Phred, Cross-Matched, Phrad und Consed (P. Green, University of Washington, B. Ewing und D. Gordon, Washington University in Saint Louis). Nachdem die ursprünglichen Sequenzen miteinander unter Bildung von längeren kohärenten Sequenzen verbunden wurden (Contigs), werden die fehlenden DNA Abschnitte spezifisch mit Hilfe von neuen Primern (Bindung an die sequenzierten Abschnitte) oder durch längeres Sequenzieren oder Sequenzieren des anderen Strangs sequenziert.

[0109] Es ist möglich mit dem oben beschriebenen Verfahren die gesamte chromosomale DNA Region mit einer Größe von 26 kb aus *A. mediterranei* zu sequenzieren, die in pNE112 kloniert ist. Die DNA Sequenz ist in SEQ ID Nr. 3 im Basenpaarabschnitt 27801-53789 gezeigt. Die DNA Sequenz des in Beispiel 5 beschriebenen 5,7 kb KpnI Fragments kommt in pNE112 vor und ist in SEQ ID Nr. 3 in der Basenpaarregion 43093 bis 48768 gezeigt.

Beispiel 10: Identifizierung und Charakterisierung von Cosmidklonen mit chromosomalen DNA Fragmenten von *A. mediterranei*, die mit einem Ende der 26 kb *A. mediterranei* Region von pNE112 überlappen

[0110] Um Cosmidklone zu identifizieren, die chromosomale DNA Fragmente von *A. mediterranei* umfassen, die direkt vor der 26 kb Region von pNE112 liegen, wird das Plasmid pNE112 mit dem Restriktionsenzym BamHI geschnitten und das entstehende BamHI Fragment mit einer Größe von 3,2 kb wird aus den anderen BamHI Fragmenten in einem Agarosegel abgetrennt und aus dem Gel isoliert. Dieses BamHI Fragment liegt an einem Ende der in pNE112 eingebauten *A. mediterranei* DNA (siehe Beispiel 8) und kann so als DNA Sonde für die Auffindung der erforderlichen Cosmidklone verwendet werden. Etwa 0,5 µg des isolierten 3,2 kb BamHI DNA Fragments werden mit ³²P-dCTP durch das Nicktranslationsystem aus Gibco/BRL (Basel) gemäß den Anleitungen des Herstellers radioaktiv markiert.

[0111] Die Cosmidgenbank von *A. mediterranei*, die in Beispiel 7 beschrieben ist, wird dann durch Koloniehybridisierung (Verfahren von Beispiel 3) mittels der 3,2 kb DNA Sonde für Klone mit Überlappungen verwendet. Zwei Cosmidklone mit einem starken Hybridisierungssignal können auf diese Weise identifiziert werden und ihnen werden die Nummern pNE95 und pRi44-2 gegeben. Es ist durch Restriktionsanalyse und Southern Blot möglich, zu bestätigen, dass die Plasmide pNE95 und pRi44-2 die chromosomalen DNA Fragmente von *A. mediterranei* enthalten, die mit dem 3,2 kb BamHI Fragment von pNE112 überlappen und zusammen eine 35 kb große chromosomale Region von *A. mediterranei* abdecken, die direkt an das 26 kb *A. mediterranei* Fragment von pNE112 benachbart liegt, die in pNE112 kloniert ist.

Beispiel 11: Restriktionsanalyse der chromosomalen *A. mediterranei* DNA Region, die mit den Cosmidklonen pNE112, pNE95 und pRi44-2 kloniert wurde

[0112] Die chromosomale *A. mediterranei* DNA Region, die mit den Cosmidklonen pNE112, pNE95 und pRi44-2 kloniert wurde, wird durch Ausführung einer Restriktionsanalyse charakterisiert. Der Verdau der Plasmid DNA der drei Cosmide mit den Restriktionsenzymen EcoRI, BglII und HindIII (einzeln und in Kombination) liefert eine grobe Restriktionskarte der klonierten Region von *A. mediterranei*. Es werden auf diese Weise überlappende Fragmente der 3 Plasmide etabliert und durch Southern Blot bestätigt. Diese chromosomale Region von *A. mediterranei* hat eine Größe von etwa 61 kb und wird durch die folgenden Restriktionsschnittstellen mit dem angegebenen Abstand in kb von einem Ende charakterisiert: EcoRI in Position 7,2 kb, HindIII in Position 21 kb, BglII in Position 31 kb, HindIII in Position 42 kb, BglII in Position 47 kb und BglII in Position 59 kb. In dieser Region im *A. mediterranei* Chromosom deckt das Plasmid pRi44-2 eine Region von Position 1 bis etwa 37 kb ab, das Plasmid pNE95 deckt eine Region etwa von Position 9 kb bis 51 kb ab und das Plasmid pNE112 deckt eine Region etwa von Position 35 kb bis 61 kb ab.

Beispiel 12: Bestimmung der Sequenz der chromosomalen *A. mediterranei* DNA Region, die in Beispiel 11 beschrieben wurde, von der EcoRI Schnittstelle an der Position 7,2 kb bis zum 61 kb Ende

[0113] Die Bestimmung der DNA Sequenz der in Beispiel 11 beschriebenen chromosomalen Region von *A. mediterranei* (EcoRI Schnittstelle an der Position 7,2 kb bis 51 kb) wird mit den Plasmiden pRi44-2 und pNE95

unter Verwendung desselben Verfahrens ausgeführt, wie dies in Beispiel 9 beschrieben ist. Die Analyse der auf diese Weise erhaltenen DNA Sequenz bestätigt die grobe Restriktionskarte, die in Beispiel 11 beschrieben wurde, und überlappt mit den klonierten *A. mediterranei* Fragmenten in den Plasmiden pNE112, pNE95 und pRi44-2.

[0114] Die DNA Sequenz der chromosomalen *A. mediterranei* DNA Region, die in Beispiel 11 beschrieben ist, von der EcoRI Schnittstelle an der Position 7,2 kb bis zum Ende bei 61 kb ist in SEQ ID Nr. 3 beschrieben (Länge 53789 Basenpaare).

Beispiel 13: Analyse einer ersten für ein Protein kodierenden Region (ORF A) der klonierten in SEQ ID Nr. 3 gezeigten chromosomalen Region von *A. mediterranei*

[0115] Die in SEQ ID Nr. 3 gezeigte Nukleotidsequenz wird mit dem Codenpräferenzcomputerprogramm (Genetics Computer Group, University of Wisconsin, 1994) analysiert. Diese Analyse zeigt, dass ein sehr großer offener Leserahmen (ORF A), der für ein Protein kodiert, im ersten Drittel der Sequenz vorkommt (Position 1825 bis 15543 einschließlich des Stopcodons in SEQ ID Nr. 3). Die in ORF A verwendeten Codons sind für Actinomycetengene mit einem hohen G + C Gehalt typisch.

[0116] Ein Vergleich der Aminosäuresequenz von ORF A (SEQ ID Nr. 4, Größe 4572 Aminosäuren) mit anderen Polyketidsynthasen und speziell mit der sehr gut charakterisierten Polyketidsynthase von *Saccharopolyspora erythraea* (Donadio, Science, (1991), 252, 675-679, DNA Sequenz Gen/EMBL Hinterlegungsnummer M63676) ergibt die folgenden Ergebnisse:

Die Region von ORF A, SEQ ID Nr. 4, Aminosäuren 370 bis 451 ist zu 50 % identisch zur Acylcarrierprotein-domäne von Modul 1 des *eryA* Lokus von *Saccharopolyspora erythraea*.

Die Region von ORF A, SEQ ID Nr. 4, Aminosäuren 469 bis 889 ist zu 65 % identisch zur Ketoacylsynthase-domäne von Modul 1 des *eryA* Lokus von *Saccharopolyspora erythraea*.

Die Region von ORF A, SEQ ID Nr. 4, Aminosäuren 982 bis 1292 ist zu 54 % identisch zur Acyltransferasedomäne von Modul 1 des *eryA* Lokus von *Saccharopolyspora erythraea*.

Die Region von ORF A, SEQ ID Nr. 4, Aminosäuren 1324 bis 1442 ist zu 42 % identisch zur Dehydratasedomäne von Modul 4 des *eryA* Lokus von *Saccharopolyspora erythraea*.

Die Region von ORF A, SEQ ID Nr. 4, Aminosäuren 1664 bis 1840 ist zu 56 % identisch zur Ketoreduktasedomäne von Modul 1 des *eryA* Lokus von *Saccharopolyspora erythraea*.

Die Region von ORF A, SEQ ID Nr. 4, Aminosäuren 1929 bis 2000 ist zu 53 % identisch zur Acylcarrierprotein-domäne von Modul 1 des *eryA* Lokus von *Saccharopolyspora erythraea*.

Die Region von ORF A, SEQ ID Nr. 4, Aminosäuren 2032 bis 2453 ist zu 64 % identisch zur Ketoacylsynthasedomäne von Modul 1 des *eryA* Lokus von *Saccharopolyspora erythraea*.

Die Region von ORF A, SEQ ID Nr. 4, Aminosäuren 2554 bis 2865 ist zu 37 % identisch zur Acyltransferasedomäne von Modul 1 des *eryA* Lokus von *Saccharopolyspora erythraea*.

Die Region von ORF A, SEQ ID Nr. 4, Aminosäuren 2918 bis 2991 ist zu 54 % identisch zur Acylcarrierprotein-domäne von Modul 1 des *eryA* Lokus von *Saccharopolyspora erythraea*.

Die Region von ORF A, SEQ ID Nr. 4, Aminosäuren 3009 bis 3431 ist zu 65 % identisch zur Ketoacylsynthasedomäne von Modul 1 des *eryA* Lokus von *Saccharopolyspora erythraea*.

Die Region von ORF A, SEQ ID Nr. 4, Aminosäuren 3532 bis 3847 ist zu 53 % identisch zur Acyltransferasedomäne von Modul 1 des *eryA* Lokus von *Saccharopolyspora erythraea*.

Die Region von ORF A, SEQ ID Nr. 4, Aminosäuren 4142 bis 4307 ist zu 43 % identisch zur Ketoreduktasedomäne von Modul 1 des *eryA* Lokus von *Saccharopolyspora erythraea*.

Die Region von ORF A, SEQ ID Nr. 4, Aminosäuren 4405 bis 4490 ist zu 50 % identisch zur Acylcarrierprotein-domäne von Modul 1 des *eryA* Lokus von *Saccharopolyspora erythraea*.

[0117] Zusätzlich zu diesen signifikanten Homologien mit der *eryA* Polyketidsynthase von *S. erythraea* ist die Region von ORF A, SEQ ID Nr. 4, Aminosäuren 1 bis 356 zu 53 % identisch zur postulierten Starteinheitsaktivierungsdomäne der Rapamycinpolyketidsynthase von *Streptomyces hygroscopicus* (Aparicio et al., GENE (1996), 169, 9-16).

[0118] Die großen Ähnlichkeiten, die in der Aminosäuresequenz der enzymatischen Domänen gefunden werden, legen unzweifelhaft nahe, dass die für das Protein kodierende Region (ORF A) der chromosomalen Region von *A. mediterranei*, die in SEQ ID Nr. 3 gezeigt ist, für eine typische modulate Polyketidsynthase (Typ 1) kodiert. Diese sehr große *A. mediterranei* Polyketidsynthase, die von ORF A kodiert wird, umfasst drei komplette bioaktive Module, die jeweils für die Kondensation einer C2 Einheit im Makrolidring des Moleküls und die korrekte Modifikation der anfänglich gebildeten β -Ketogruppen verantwortlich sind. Aufgrund der Homologie

mit den Aktivierungsdomänen der Rapamycinpolyketidsynthase umfasst das erste oben beschriebenen Modul sehr wahrscheinlich eine enzymatische Domäne zur Aktivierung der aromatischen Starteinheit der Rifamycinbiosynthese, nämlich 3-Amino-5-hydroxybenzoesäure (Ghisalpa et al., *Biotechnology of Industrial Antibiotics*, E.J. Vandamme, Herausgeber, Decker Inc. New York, (1984) 281-327).

Beispiel 14: Analyse einer zweiten für Protein kodierenden Region (ORF B) der klonierten chromosomalen Region von *A. mediterranei*, die in SEQ ID Nr. 3 gezeigt ist

[0119] Die Nukleotidsequenz von SEQ ID Nr. 3 wird mittels des Codonpräferenzcomputerprogramms (Genetics Computer Group, University of Wisconsin, 1994) analysiert. Diese Analyse zeigt, dass ein weiterer großer offener Leserahmen (ORF B), der für ein Protein kodiert, in der mittleren Region der Sequenz vorkommt (Position 15550 bis 30759 einschließlich des Stopcodons der SEQ ID Nr. 3). Die in ORF B verwendeten Codons sind für Actinomycetengene mit einem hohen G + C Gehalt typisch.

[0120] Ein Vergleich der Aminosäuresequenz von ORF B (SEQ ID Nr. 5, Länge 5069 Aminosäuren) mit anderen Polyketidsynthasen und speziell mit der sehr gut charakterisierten Polyketidsynthase von *Saccharopolyspora erythraea* (Donadio, *Science* (1991), 252, 675-679, DNA Sequenz Gen/EMBL Hinterlegungsnummer M63676) ergibt die folgenden Ergebnisse:

Die Region ORF B, SEQ ID Nr. 5, Aminosäuren 44 bis 468 ist zu 62 % identisch zur Ketoacylsynthasedomäne von Modul 1 des *eryA* Lokus von *Saccharopolyspora erythraea*.

Die Region ORF B, SEQ ID Nr. 5, Aminosäuren 571 bis 889 ist zu 56 % identisch zur Acyltransferasedomäne von Modul 1 des *eryA* Lokus von *Saccharopolyspora erythraea*.

Die Region ORF B, SEQ ID Nr. 5, Aminosäuren 921 bis 1055 ist zu 47 % identisch zur Dehydratasedomäne von Modul 4 des *eryA* Lokus von *Saccharopolyspora erythraea*.

Die Region ORF B, SEQ ID Nr. 5, Aminosäuren 1353 bis 1525 ist zu 49 % identisch zur Ketoreduktasedomäne von Modul 1 des *eryA* Lokus von *Saccharopolyspora erythraea*.

Die Region ORF B, SEQ ID Nr. 5, Aminosäuren 1621 bis 1706 ist zu 53 % identisch zur Acylcarrierproteindomäne von Modul 1 des *eryA* Lokus von *Saccharopolyspora erythraea*.

Die Region ORF B, SEQ ID Nr. 5, Aminosäuren 1726 bis 2148 ist zu 62 % identisch zur Ketoacylsynthasedomäne von Modul 1 des *eryA* Lokus von *Saccharopolyspora erythraea*.

Die Region ORF B, SEQ ID Nr. 5, Aminosäuren 2251 bis 2560 ist zu 55 % identisch zur Acyltransferasedomäne von Modul 1 des *eryA* Lokus von *Saccharopolyspora erythraea*.

Die Region ORF B, SEQ ID Nr. 5, Aminosäuren 2961 bis 3132 ist zu 49 % identisch zur Ketoreduktasedomäne von Modul 1 des *eryA* Lokus von *Saccharopolyspora erythraea*.

Die Region ORF B, SEQ ID Nr. 5, Aminosäuren 3228 bis 3313 ist zu 52 % identisch zur Acylcarrierproteindomäne von Modul 1 des *eryA* Lokus von *Saccharopolyspora erythraea*.

Die Region ORF B, SEQ ID Nr. 5, Aminosäuren 3332 bis 3755 ist zu 63 % identisch zur Ketoacylsynthasedomäne von Modul 1 des *eryA* Lokus von *Saccharopolyspora erythraea*.

Die Region ORF B, SEQ ID Nr. 5, Aminosäuren 3857 bis 4173 ist zu 52 % identisch zur Acyltransferasedomäne von Modul 1 des *eryA* Lokus von *Saccharopolyspora erythraea*.

Die Region ORF B, SEQ ID Nr. 5, Aminosäuren 4664 bis 4799 ist zu 47 % identisch zur Ketoreduktasedomäne von Modul 1 des *eryA* Lokus von *Saccharopolyspora erythraea*.

Die Region ORF B, SEQ ID Nr. 5, Aminosäuren 4929 bis 5014 ist zu 52 % identisch zur Acylcarrierproteindomäne von Modul 1 des *eryA* Lokus von *Saccharopolyspora erythraea*.

Beispiel 15: Analyse einer dritten für Protein kodierenden Region (ORF C) der klonierten chromosomalen Region von *A. mediterranei*, die in SEQ ID Nr. 3 gezeigt ist

[0121] Die Nukleotidsequenz in SEQ ID Nr. 3 wird mittels des Codonpräferenzcomputerprogramms (Genetics Computer Group, University of Wisconsin, 1994) analysiert. Diese Analyse zeigt, dass ein großer offener Leserahmen (ORF C), der für ein Protein kodiert, in der mittleren Region der Sequenz vorkommt (Position 30895 bis 36060 einschließlich des in SEQ ID Nr. 3 verwendeten Stopcodons). Die in ORF C verwendeten Codons sind für Actinomycetengene mit einem hohen G + C Gehalt typisch.

[0122] Ein Vergleich der Aminosäuresequenz von ORF C (SEQ ID Nr. 6, Länge 1721 Aminosäuren) mit anderen Polyketidsynthasen und insbesondere der sehr gut charakterisierten Polyketidsynthase von *Saccharopolyspora erythraea* (Donadio, *Science*, (1991) 252, 675-679, DNA Sequenz Gen/EMBL Hinterlegungsnummer M63676) ergibt die folgenden Ergebnisse:

Die Region von Ort C, SEQ ID Nr. 6, Aminosäuren 1 bis 414 ist zu 63 % identisch zur Ketoacylsynthasedomäne von Modul 1 des *eryA* Lokus von *Saccharopolyspora erythraea*.

Die Region von Orf C, SEQ ID Nr. 6, Aminosäuren 514 bis 828 ist zu 54 % identisch zur Acyltransferasedomäne von Modul 1 des *eryA* Lokus von *Saccharopolyspora erythraea*.

Die Region von Ort C, SEQ ID Nr. 6, Aminosäuren 1290 bis 1399 ist zu 49 % identisch zur Ketoreduktasedomäne von Modul 1 des *eryA* Lokus von *Saccharopolyspora erythraea*.

Die Region von Ort C, SEQ ID Nr. 6, Aminosäuren 1563 bis 1648 ist zu 55 % identisch zur Acylcarrierprotein-domäne von Modul 1 des *eryA* Lokus von *Saccharopolyspora erythraea*.

Beispiel 16: Analyse einer vierten für Protein kodierenden Region (ORF D) der klonierten chromosomalen Region von *A. mediterranei*, die in SEQ ID Nr. 3 gezeigt ist

[0123] Die Nukleotidsequenz in SEQ ID Nr. 3 wird mittels des Codonpräferenzcomputerprogramms (Genetics Computer Group, University of Wisconsin, 1994) analysiert. Diese Analyse zeigt, dass ein großer offener Leserahmen (ORF D), der für ein Protein kodiert, in der mittleren Region der Sequenz vorkommt (Position 36259 bis 41325 einschließlich des Stopcodons in SEQ ID Nr. 3). Die in ORF D verwendeten Codons sind für Actinomycetengene mit einem hohen G + C Gehalt typisch.

[0124] Ein Vergleich der Aminosäuresequenz von ORF D (SEQ ID Nr. 7, Länge 1688 Aminosäuren) mit anderen Polyketidsynthasen und speziell mit der sehr gut charakterisierten Polyketidsynthase von *Saccharopolyspora erythraea* (Donadio, Science, (1991) 252, 675-679, DNA Sequenz Gene/EMBL Hinterlegungsnummer M63676) ergibt die folgenden Ergebnisse:

Die Region von Orf D, SEQ ID Nr. 7, Aminosäuren 1 bis 418 ist zu 64 % identisch zur Ketoacylsynthasedomäne von Modul 1 des *eryA* Lokus von *Saccharopolyspora erythraea*.

Die Region von Orf D, SEQ ID Nr. 7, Aminosäuren 524 bis 841 ist zu 54 % identisch zur Acyltransferasedomäne von Modul 1 des *eryA* Lokus von *Saccharopolyspora erythraea*.

Die Region von Orf D, SEQ ID Nr. 7, Aminosäuren 1260 bis 1432 ist zu 51 % identisch zur Ketoreduktasedomäne von Modul 1 des *eryA* Lokus von *Saccharopolyspora erythraea*.

Die Region von Ort D, SEQ ID Nr. 7, Aminosäuren 1523 bis 1608 ist zu 53 % identisch zur Acylcarrierprotein-domäne von Modul 1 des *eryA* Lokus von *Saccharopolyspora erythraea*.

Beispiel 17: Analyse einer fünften für Protein kodierenden Region (ORF E) der klonierten chromosomalen Region von *A. mediterranei*, die in SEQ ID Nr. 3 gezeigt ist

[0125] Die Nukleotidsequenz in SEQ ID Nr. 3 wird mittels des Codonpräferenzcomputerprogramms analysiert (Genetics Computer Group, Wisconsin, 1994). Diese Analyse zeigt, dass ein großer offener Leserahmen (ORF E), der für ein Protein kodiert, in der hinteren Region der Sequenz vorkommt (Position 41373 bis 51614 einschließlich des Stopcodons in SEQ ID Nr. 3). Die in ORF E verwendeten Codons sind für Actinomycetengene mit einem hohen G + C Gehalt typisch.

[0126] Ein Vergleich der Aminosäuresequenz von ORF E (SEQ ID Nr. 8, Länge 3413 Aminosäuren) mit anderen Polyketidsynthasen und insbesondere der sehr gut charakterisierten Polyketidsynthase von *Saccharopolyspora erythraea* (Donadio, Science (1991), 252, 675-679, DNA Sequenz Gen/EMBL Hinterlegungsnummer M63676) ergibt die folgenden Ergebnisse:

Die Region von ORF E, SEQ ID Nr. 8, Aminosäuren 31 bis 451 ist zu 64 % identisch zur Ketoacylsynthasedomäne von Modul 1 des *eryA* Lokus von *Saccharopolyspora erythraea*.

Die Region von ORF E, SEQ ID Nr. 8, Aminosäuren 555 bis 874 ist zu 37 % identisch zur Acyltransferasedomäne von Modul 1 des *eryA* Lokus von *Saccharopolyspora erythraea*.

Die Region von ORF E, SEQ ID Nr. 8, Aminosäuren 907 bis 1036 ist zu 49 % identisch zur Dehydratasedomäne von Modul 4 des *eryA* Lokus von *Saccharopolyspora erythraea*.

Die Region von ORF E, SEQ ID Nr. 8, Aminosäuren 1336 bis 1500 ist zu 52 % identisch zur Ketoreduktasedomäne von Modul 1 des *eryA* Lokus von *Saccharopolyspora erythraea*.

Die Region von ORF E, SEQ ID Nr. 8, Aminosäuren 1598 bis 1683 ist zu 51 % identisch zur Acylcarrierprotein-domäne von Modul 1 des *eryA* Lokus von *Saccharopolyspora erythraea*.

Die Region von ORF E, SEQ ID Nr. 8, Aminosäuren 1702 bis 2124 ist zu 62 % identisch zur Ketoacylsynthasedomäne von Modul 1 des *eryA* Lokus von *Saccharopolyspora erythraea*.

Die Region von ORF E, SEQ ID Nr. 8, Aminosäuren 2229 bis 2543 ist zu 53 % identisch zur Acyltransferasedomäne von Modul 1 des *eryA* Lokus von *Saccharopolyspora erythraea*.

Die Region von ORF E, SEQ ID Nr. 8, Aminosäuren 2573 bis 2700 ist zu 47 % identisch zur Dehydratasedomäne von Modul 4 des *eryA* Lokus von *Saccharopolyspora erythraea*.

Die Region von ORF E, SEQ ID Nr. 8, Aminosäuren 3054 bis 3227 ist zu 52 % identisch zur Ketoreduktasedomäne von Modul 1 des *eryA* Lokus von *Saccharopolyspora erythraea*.

Die Region von ORF E, SEQ ID Nr. 8, Aminosäuren 3324 bis 3405 ist zu 51 % identisch zur Acylcarrierprotein-domäne von Modul 1 des *eryA* Lokus von *Saccharopolyspora erythraea*.

Beispiel 18: Analyse einer sechsten für Protein kodierenden Region (ORF F) der klonierten chromosomalen Region von *A. mediterranei*, die in SEQ ID Nr. 3 gezeigt ist

[0127] Die Nukleotidsequenz in SEQ ID Nr. 3 wird mittels des Codonpräferenzcomputerprogramms analysiert (Genetics Computer Group, University of Wisconsin, 1994). Die Analyse zeigt, dass ein offener Leserahmen (ORF F), der für ein Protein kodiert, in der hinteren Region der Sequenz vorkommt (Position 51713 bis 52393 einschließlich des Stopcodons in SEQ ID Nr. 3). Die verwendeten Codons in ORF F sind typisch für Actinomycetengene mit einem hohen G + C Gehalt.

[0128] Ein Vergleich der Aminosäuresequenz von ORF F (SEQ ID Nr. 9, Länge 226 Aminosäuren) mit Proteinen aus der EMBL Datenbank (Heidelberg) zeigt eine große Ähnlichkeit mit der N-Hydroxyarylamino-O-acyltransferase von *Salmonella typhimurium* (29 % Identität über eine Region von 134 Aminosäuren). Es besteht auch eine signifikante Homologie mit Arylaminoacyltransferasen aus anderen Organismen. Es kann aus diesen Übereinstimmungen geschlossen werden, dass der ORF F, der in *A. mediterranei* in SEQ ID Nr. 3 gefunden wird, für eine Arylaminoacyltransferase kodiert und es kann angenommen werden, dass dieses Enzym für die Anbindung der langkettigen Acylkette, die durch die Polyketidsynthase gebildet wird, an die Aminogruppe des Ausgangsmoleküls 3-Amino-5-hydroxybenzoesäure verantwortlich ist. Diese Reaktion würde das Rifamycin-ringsystem korrekt nach der Vervollständigung der Kondensationsschritte durch die Polyketidsynthase schließen.

Beispiel 19: Zusammenfassende Evaluierung der Funktion der Proteine, die durch ORF A bis F in SEQ ID Nr. 3 kodiert werden und ihre Rolle in der Biosynthese von Rifamycin

[0129] Die 5 für Protein kodierenden Regionen (ORF A bis E), die in den Beispielen 13 bis 17 der SEQ ID Nr. 3 beschrieben sind, umfassen Proteine mit einer sehr großen Ähnlichkeit (in der Aminosäuresequenz und der Anordnung der enzymatischen Domänen) zu Polyketidsynthasen für Polyketide vom Makrolidtyp. Zusammengefasst umfassen diese 5 multifunktionellen Enzyme 10 Polyketidsynthasemoleküle, die für einen Kondensationsschritt in der Polyketidsynthese verantwortlich sind. Es sind ebenfalls 10 solcher Kondensationsschritte für die Rifamycinbiosynthese erforderlich (Ghiselba et al., *Biotechnology of Industrial Antibiotics*, E.J. Vandamme, Decker Inc. New York, (1984) 281-327). Die Prozessierung der bestimmten Ketogruppen, die von den enzymatischen Domänen innerhalb der Module erforderlich sind, entsprechen im wesentlichen der Aktivität, die für das Rifamycinmolekül erforderlich ist, falls angenommen wird, dass die Polyketidsynthese „colinear“ mit der Anordnung der Module im Gencluster von *A. mediterranei* stattfindet (dies ist der Fall für andere Makrolidantibiotika, wie Erythromycin und Rapamycin). Es kann hier angefügt werden, dass es nicht sicher ist, ob die Transkription der 5 ORFs zu 5 Proteinen führt, wobei insbesondere ORF C und ORF D möglicherweise in ein größeres Protein translatiert werden.

[0130] Eine enzymatische Domäne, die sehr wahrscheinlich für die Aktivierung des Startmoleküls 3-Hydroxy-5-aminobenzoesäure der Rifamycinbiosynthese ist, kann am N-Terminus des ORF A, dem Beginn der Polyketidsynthase, gefunden werden. Direkt hinter dem beschriebenen Rifamycinpolyketidsynthasegencluster liegt ein Gen (ORF F), das sehr wahrscheinlich für ein Protein kodiert, das den Ringschluss des Rifamycinmoleküls nach der Vervollständigung der Kondensationsschritte durch die Polyketidsynthase herbeiführt.

[0131] Es kann auf der Grundlage dieser Feststellungen geschlossen werden, dass die chromosomale Region von *A. mediterranei*, die in SEQ ID Nr. 3 beschrieben ist, für die zehn Kondensationsschritte verantwortlich ist, die für die Rifamycinpolyketidsynthese erforderlich sind, einschließlich der Aktivierung des Startmoleküls 3-Hydroxy-5-aminobenzoesäure und dem abschließenden Ringschluss.

Hinterlegte Mikroorganismen

[0132] Die folgenden Mikroorganismen und Plasmide wurden bei der Deutschen Sammlung von Mikroorganismen und Zellkulturen GmbH (DSMZ), Mascheroder Weg 1b, D-38124 Braunschweig gemäß den Anforderungen des Budapest-Vertrags hinterlegt.

Mikroorganismus / Plasmid	Hinterlegungsdatum	Hinterlegungsnummer
E. coli mit Plasmid pRi7-3	10.8.96	DSM 11114
E. coli mit Plasmid pNE112	14.7.97	DSM 11657
E. coli mit Plasmid pNE95	14.7.97	DSM 11656
E. coli mit Plasmid pRi44-2	14.7.97	DSM 11655

Sequenzliste

(1) Allgemeine Information:

(i) Anmelder:

(A) Name: Novartis AG
 (B) Straße: Schwarzwaldallee 215
 (C) Stadt: Basel
 (E) Land: Schweiz
 (F) Postleitzahl (ZIP): 4058
 (G) Telefon: +41 61 324 1111
 (H) Telefax: +41 61 322 75 32

(ii) Titel der Erfindung: Rifamycinbiosynthesegencluster

(iii) Anzahl an Sequenzen: 9

(iv) Computerlesbare Form:

(A) Mediumtyp: Diskette
 (B) Computer: IBM PC kompatibel
 (C) Betriebssystem: PC-DOS/MS-DOS
 (D) Software: PatentIn Release Nr. 1.0, Version Nr. 1.30 (EPO)

(xi) Sequenzbeschreibung: SEQ ID Nr: 1

GGTACCCGGT GTTCGCGACG GCGTTCGACG AGGCTTGCGA GCAGCTGGAC GTCTGTCTGG	60
CCGGCCGTGC CGGGCACC GC GTGCGGGACG TCGTGCTCGG CGAAGTGCCC GCCGAAACCG	120
GGCTGCTGAA CCAGACGGTC TTCACCCAAG CCGGGCTGTT CGCGGTGGAG AGCGCGCTGT	180
TCCGGCTCGC CGAATCCTGG GGTGTCCGGC CGGACGTGGT GCTCGGCCAC TCCATCGGGG	240
AGATCACCGC CGCGTATGCC GCGGGCGTCT TCTCGCTGCC GGACGCCGCC CGGATCGTCG	300
CGGCGCGCGG CCGGCTGATG CAGGCGCTGG CGCCGGGCGG GCGGATGGTC GCCGTGCGCG	360
CCTCCGAAGC CGAGGTGGCC GAACTGCTCG GCGACGGCGT GGAAGTCGCC GCCGTCAACG	420
GCCCTTCGGC GGTAAGTCTT TCCGGGGACG CGGACGCGGT CGTCGCGGCC GCCGCCCGCA	480
TGCGCGAGCG CGGGCACAAG ACCAAGCAGC TCAAGGTTTC GCACGCGTTC CACTCCGCGC	540
GGATGGCGCC GATGCTGGCG GAGTTCGCCG CCGAGCTGGC CGGCGTGACG TGGCGCGAGC	600
CGGAGATCCC GGTGGTCTCC AACGTGACCG CCCGGTTCCG CGAGCCCGGC GAACTGACCG	660
AGCCGGGCTA CTGGGCCGAG CACGTGCGGC GGCCGGTGCG GTTCGCCGAG GCGGTGCGCG	720
CCGCGACGGA GTCCGGCGGC TCGCTGTTCTG TGGAGCTCGG GCCGGGGGCG GCGCTGACCG	780
CCCTCGTCGA GGAGACGGCC GAGGTCACCT GCGTCGCGGC CCTGCGGGAC GACCGCCCGG	840
AGGTCACCGC GCTGATCACC GCGGTGCGCG AGCTGTTCGT CCGCGGGGTT GCGGTGATT	900
GGCCGGCCCT GCTGCCGCG GTCACCGGGT TCGTCGACCT GCCGAAGTAC GCCTTCGACC	960
AGCAGCACTA TTGGCTGCAG CCCGCCGCGC AGGCCACGGA CGCGGCCTCG CTCGGGCAGG	1020

TCGCGGCCGA CCACCCGCTG CTGGGCGCGG TGGTCCGGCT GCCGCAGTCG GACGGCCTGG	1080
TCTTCACCTC GCGGCTGTCA TTGAAATCGC ACCCGTGGCT GGCCGACCAC GTCATCGGCG	1140
GGGTGGTGCT CGTCGCGGGC ACCGGGCTCG TCGAGCTGGC CGTCCGGGCC GGGGACGAGG	1200
CCGGCTGCCC GGTCTCTGAA GAACTCGTCA TCGAGGCTCC GCTGGTCGTC CCCGACCACG	1260
GCGGGGTCCG GATCCAGGTC GTCGTGGGG CACCGGGGA GACCGGTTTCG CGCGCGGTTCG	1320
AGGTGTACTC CCTGCGCGAG GACGCCGGTG CCGAAGTGTG GGCCCGGCAC GCCACCGGGT	1380
TCCTGGCTGC GACGCCGTCG CAGCACAAGC CGTTCGACTT CACCGCCTGG CCGCCGCCCCG	1440
GCGTCGAGCG CGTCGACGTC GAGGACTTCT ACGACGGCTT CGTCGACCGC GGGTACGCCT	1500
ACGGGCCGTC GTTCCGGGGC CTGCGGGCGG TGTGGCGGCG CGGCGACGAA GTGTTGCGCCG	1560
AGGTCGCCCT GGCCGAGGAC GACCGCGCGG ACGCGGCCCG GTTCGGCATC CACCCCGGCC	1620
TGCTGGACGC CGCCCTGCAC GCGGGCATGG CCGGTGCCAC CACCACGGAA GAGCCCGGCC	1680
GGCCGGTGCT GCCGTTGCCC TGGAACGGCC TGGTGCTGCA CGCGGCCGGG GCGTCCGCGC	1740
TGCGGGTCCG GCTCGCCCCG AGCGGTCCGG ACGCCCTGTC GGTCGAGGCC GCGGACGAGG	1800
CCGGCGGTCT CGTTGTGACG GCGGACTCGC TGGTCTCCCG GCCGGTGTCTG GCCGAACAGC	1860
TGGGCGCGGC GGCGAACCAC GACGCGTTGT TCCGCGTGA GTGGACCGAG ATTTCTCTCGG	1920
CTGGAGACGT TCCGGCGGAC CACGTCAAG TGCTCGAAGC CGTCGGCGAG GATCCCCTGG	1980
AACTGACCGG CCGGGTCCTG GAGGCCGTGC AGACCTGGCT CGCCGACGCA GCCGACGACG	2040
CTCGCCTGGT CGTGGTGACC CGCGGCGCCG TCCACGAGGT GACTGACCCG GCCGGTGCCG	2100
CGGTGTGGGG CCTGATCCGG GCCGCGCAGG CGGAAAACCC GGACCGGATC GTGCTGCTGG	2160

ACACCGACGG TGAAGTGCCG CTAGGCCGGG TGCTGGCCAC CGGCGAGCCC CAAACAGCCG	2220
TCCGAGGCGC CACGCTGTTC GCCCCGCGC TGGCCGCGC CGAGGCCGCG GAGGCACCGG	2280
CAGTGACCGG CGGGACGGTC CTGATCTCGG GCGCCGGCTC GCTGGGCGCG CTCACCGCCC	2340
GGCACCCTGGT CGCCCGGCAC GGAGTCCGGC GGCTGGTGCT CGTCAGCCGC CGTGGCCCCG	2400
ACGCCGACGG CATGGCCGAA CTGACCGCTG AACTCATCGC TCAGGGCGCC GAGGTGCGCG	2460
TAGTCGCTTG CGACCTGGCC GACCGGGACC AGGTCCGGGT ACTGCTGGCC GAGCACCGCC	2520
CGAACGCCGT CGTGACACG GCCGGTGTTC TCGACGACGG CGTCTTCGAG TCGCTGACGC	2580
GGGAGCGGCT GGCCAAGGTC TTCGCGCCCA AAGTTACTGC TGCCAATCAC CTCGACGAGC	2640
TGACCCGCGA ACTGGATCTT CGCGCGTTCG TCGTGTTCTC CTCGCGCTCC GGGGTCTTCG	2700
GCTCCGCCGG GCAGGGCAAC TACGCCGCTG CCAACGCCTA CCTGGACGCC GTGGTCGCCA	2760
ACGCCCGGGC CGCGGGCCTG CCCGGCACAT CGCTGGCCTG GGGCCTGTGG GAACAGACCG	2820
ACGGGATGAC CGCGCACCTC GCGGACGCCG ACCAGGCGCG GCGGAGTCGC GGCGGGGTCC	2880
TCGCCATCTC ACCCGCCGAA GGCATGGAGC TGTTGACGC AGCGCCGGAC GGGCTCGTCG	2940
TCCCGGTCAA GCTGGACCTG CGCAAGACCC GCGCCGGCGG GACGGTGCCG CACCTGCTGC	3000
GCGGCCTGGT CCGCCCGGGA CGGCAGCAGG CCCGTCCGGC GTCCACTGTG GACAACGGAC	3060
TGGCCGGGCG ACTCGCCGGG CTCGCGCCGG CGGAGCAGGA GGCGCTGCTG CTCGACGTG	3120
TCCGCACGCA GGTGCGCTG GTGCTCGGGC ACGCCGGGCC GGAGGCCGTC CGCGCGGACA	3180
CGGCGTTCAA GGACACCGGC TTCGACTCGC TGACGTCGGT GGAAGTGGC AACC GGCTGC	3240

GCGAGGCGAG CGGGCTGAAG CTGCCC CGGA CGCTCGTCTT CGACTACCCG ACGCCGGTCG	3300
CGCTGGCCCC CTACCTGCGT GACGAATTCG GCGACACGGT GGCAACAACCT CCGGTGGCCA	3360
CCGCGGCCCC AGCGGACGCC GCGGAGCCGA TCGCCATCGT CGGCATGGCG TGCCGGCTGC	3420
CGGCGGGGGT CACCGATCCC GAAGGCCTGT GCGCCTGGT GCGCGACGGC CTCGAAGGGC	3480
TGTCTCCCTT CCCCAGGAC CGGGGCTGGG ACCTGGAGAA CCTGTTCGAC GACGACCCCG	3540
ACCGCTCCGG CACGACGTAC ACCAGCCGGG GCGGGTTCCT CGACGGCGCC GGCCTGTTCG	3600
ACGCGGGCTT CTTCGGGATT TCGCCGCGG AGGCGCTGGC CATGGACCCG CAGCAGCGGC	3660
TGCTGCTCGA GCGGCCTGG GAAGCCCTCG AAGGCACCGG TGTCGACCCG GGCTCGTTGA	3720
AGGGCGCCGA CGTCGGGGTG TTCGCCGGGG TGTCCAACCA GGGCTATGGG ATGGGCGCGG	3780
ATCCGGCCGA ACTGGCGGGG TACGCGAGCA CGGCGGGCGC TTCGAGCGTC GTCTCGGGCC	3840
GAGTCTCGTA CGTCTTCGGG TTCGAAGGAC CGGCGGTAC GATCGACACG GCTTGCTCGT	3900
CGTCGCTGGT GCGGATGCAC CTGGCCGGGC AGGCGCTGCG GCAGGGCGAG TGCTCGATGG	3960
CCCTGGCCGG TGGCGTCACG GTGATGGGGA CGCCCGGCAC GTTCGTGGAG TTCGGAAGC	4020
AGCGCGCCTT GGCCGGCGAC GGCCGGTGCA AGCCCTACGC CGAAGGCGCG GACGGCACGG	4080
GCTGGGCCGA GGGCGTCGGG GTCGTCGTGC TGGAGCGGCT GTCGGTGGCG CCGAGCGCG	4140
GGCACCGGGT GCTGGCCGTG CTGCGCGGCA GCGCGGTCAA CTCCGACGGC GCGTCCAACG	4200
GCTGACCGC CCCCAACGGG CCGTCGCAGC AACGGGTGAT CCGCCGGGCC CTGGCCGGCG	4260
CCGGCCTCGA ACCGTCCGAT GTGGACATCG TGAAGGGCA CGGCACCGG ACGGCGCTGG	4320
GCGACCCGAT CGAGGCGCAG GCCCTGCTGG CCACCTACGG CAAGGACCGC GACCCGGAGA	4380

CGCCGTGTGTG GCTGGGGTTCG GTGAAGTCGA ACTTCGGCCA CACGCAGTCC GCGGCCGGCG	4440
TGGCCGGGGT GATCAAGATG GTGCAGGCGC TGCGCCACGG CGTCATGCCG CCCACCCTGC	4500
ACGTGGACCG GCCCACCAGC CAGGTCTGACT GGTCCGCGGG GGCCGTCGAA GTGCTGACCG	4560
AGGCACGGGA GTGGCCGCGG AACGGCCGTC CGCGCCGGGC CGGGGTGTCC TCGTTCGGGA	4620
TCAGCGGCAC GAACGCCAC CTGATCATCG AAGAAGCACC GGCCGAGCCA CAGCTTGCCG	4680
GACCACCGCC GGACGGCGGT GTGGTGCCGC TGGTCGTCTC GGCTCGCAGC CCCGGTGCCC	4740
TGGCCGGTCA GCGCGTTCGG CTGGCCACGT TCCTCGGCGA CGGGCCCCCTT TCCGACGTCC	4800
CCGGTGCGCT GACGAGCCGC GCCCTGTTTCG GCGAGCGCGC GGTCGTCTGT GCGGATTCGG	4860
CCGAGGAAGC CCGCGCCGGT CTGGGCGCAC TGGCCCGCGG CGAAGACGCG CCGGGCCTGG	4920
TCCGCGGCCG GGTGCCCCG TCCGGCCTGC CGGGCAAGCT CGTGTGGGTG TTCCCCGGGC	4980
AGGGGACGCA GTGGGTGGGC ATGGGCCGCG AACTCCTCGA AGAGTCTCCG GTGTTCGCCG	5040
AGCGGATCGC CGAGTGTGCG GCCGCGCTGG AGCCGTGGAT CGGCTGGTCG CTGTTCGACG	5100
TCCTCCGTGG CGACGGTGAC CTCGATCGGG TCGATGTGCT GCAGCCCGCG TGCTTTGCGG	5160
TGATGGTCCG CTTGGCCGCG GTGTGGTCCT CGGCCGGGGT GTCCCCGAT GCGGTGCTCG	5220
GCCACTCCCA GGGTGAATC GCCGCGGCGT GCGTGTCCGG TGCCTTCTCG CTGGAGGATG	5280
CGGCGAAGGT GGTTCGCCCTG CGCAGCCAGG CCATCGCCGC GAAGCTCTCC GGCCGCGGCG	5340
GGATGGCTTC GTTCGCCCTG GCGGAAGCCG ATGTGGTGTC GCGGCTGGCG GACGGGGTCG	5400
AGGTGGCTGC CGTCAACGGT CCGGCGTCCG TGGTGATCGC GGGGGATGCC CAGGCCCTCG	5460

```

ACGAAACGCT GGAAGCGCTG TCCGGTGCGG GAATCCGGGC TCGGCGGGTG GCGGTGGACT 5520

ACGCCTCGCA CACCCGGCAC GTCGAAGACA TCGAAGACAC CCTCGCCGAA GCGCTGGCCG 5580

GGATCGACGC CCGGGCGCCG CTGGTGCCGT TCCTCTCCAC CCTCACCGGC GAGTGGATCC 5640

GGGACGAGGG CGTCGTGGAC GCGGGCTACT GGTACC 5676

```

(2) Information für SEQ ID Nr: 2

(i) Sequenzeigenschaften:

(A) Länge: 1891 Aminosäuren
 (B) Typ: Aminosäure
 (C) Strangart: Einzelstrang
 (D) Topologie: linear

(ii) Molekültyp: Peptid

(xi) Sequenzbeschreibung: SEQ ID Nr: 2

```

Tyr Pro Val Phe Ala Thr Ala Phe Asp Glu Ala Cys Glu Gln Leu Asp
1           5           10           15

Val Cys Leu Ala Gly Arg Ala Gly His Arg Val Arg Asp Val Val Leu
20           25           30

Gly Glu Val Pro Ala Glu Thr Gly Leu Leu Asn Gln Thr Val Phe Thr
35           40           45

Gln Ala Gly Leu Phe Ala Val Glu Ser Ala Leu Phe Arg Leu Ala Glu
50           55           60

Ser Trp Gly Val Arg Pro Asp Val Val Leu Gly His Ser Ile Gly Glu
65           70           75           80

```

Ile Thr Ala Ala Tyr Ala Ala Gly Val Phe Ser Leu Pro Asp Ala Ala
 85 90 95

Arg Ile Val Ala Ala Arg Gly Arg Leu Met Gln Ala Leu Ala Pro Gly
 100 105 110

Gly Ala Met Val Ala Val Ala Ala Ser Glu Ala Glu Val Ala Glu Leu
 115 120 125

Leu Gly Asp Gly Val Glu Leu Ala Ala Val Asn Gly Pro Ser Ala Val
 130 135 140

Val Leu Ser Gly Asp Ala Asp Ala Val Val Ala Ala Ala Ala Arg Met
 145 150 155 160

Arg Glu Arg Gly His Lys Thr Lys Gln Leu Lys Val Ser His Ala Phe
 165 170 175

His Ser Ala Arg Met Ala Pro Met Leu Ala Glu Phe Ala Ala Glu Leu
 180 185 190

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

Thr Gly Arg Phe Ala Glu Pro Gly Glu Leu Thr Glu Pro Gly Tyr Trp
 210 215 220

Ala Glu His Val Arg Arg Pro Val Arg Phe Ala Glu Gly Val Ala Ala
 225 230 235 240

Ala Thr Glu Ser Gly Gly Ser Leu Phe Val Glu Leu Gly Pro Gly Ala
 245 250 255

Ala Leu Thr Ala Leu Val Glu Glu Thr Ala Glu Val Thr Cys Val Ala
 260 265 270

Ala Leu Arg Asp Asp Arg Pro Glu Val Thr Ala Leu Ile Thr Ala Val
 275 280 285

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

Pro Pro Val Thr Gly Phe Val Asp Leu Pro Lys Tyr Ala Phe Asp Gln
 305 310 315 320

Gln His Tyr Trp Leu Gln Pro Ala Ala Gln Ala Thr Asp Ala Ala Ser
 325 330 335

Leu Gly Gln Val Ala Ala Asp His Pro Leu Leu Gly Ala Val Val Arg
 340 345 350

Leu Pro Gln Ser Asp Gly Leu Val Phe Thr Ser Arg Leu Ser Leu Lys
 355 360 365

Ser His Pro Trp Leu Ala Asp His Val Ile Gly Gly Val Val Leu Val
 370 375 380

Ala Gly Thr Gly Leu Val Glu Leu Ala Val Arg Ala Gly Asp Glu Ala
 385 390 395 400

Gly Cys Pro Val Leu Glu Glu Leu Val Ile Glu Ala Pro Leu Val Val
 405 410 415

Pro Asp His Gly Gly Val Arg Ile Gln Val Val Val Gly Ala Pro Gly
 420 425 430

Glu Thr Gly Ser Arg Ala Val Glu Val Tyr Ser Leu Arg Glu Asp Ala
 435 440 445

Gly Ala Glu Val Trp Ala Arg His Ala Thr Gly Phe Leu Ala Ala Thr
 450 455 460

Pro Ser Gln His Lys Pro Phe Asp Phe Thr Ala Trp Pro Pro Pro Gly

465	470	475	480
Val Glu Arg Val Asp Val Glu Asp Phe Tyr Asp Gly Phe Val Asp Arg			
485	490	495	
Gly Tyr Ala Tyr Gly Pro Ser Phe Arg Gly Leu Arg Ala Val Trp Arg			
500	505	510	
Arg Gly Asp Glu Val Phe Ala Glu Val Ala Leu Ala Glu Asp Asp Arg			
515	520	525	
Ala Asp Ala Ala Arg Phe Gly Ile His Pro Gly Leu Leu Asp Ala Ala			
530	535	540	
Leu His Ala Gly Met Ala Gly Ala Thr Thr Thr Glu Glu Pro Gly Arg			
545	550	555	560
Pro Val Leu Pro Phe Ala Trp Asn Gly Leu Val Leu His Ala Ala Gly			
565	570	575	
Ala Ser Ala Leu Arg Val Arg Leu Ala Pro Ser Gly Pro Asp Ala Leu			
580	585	590	
Ser Val Glu Ala Ala Asp Glu Ala Gly Gly Leu Val Val Thr Ala Asp			
595	600	605	
Ser Leu Val Ser Arg Pro Val Ser Ala Glu Gln Leu Gly Ala Ala Ala			
610	615	620	
Asn His Asp Ala Leu Phe Arg Val Glu Trp Thr Glu Ile Ser Ser Ala			
625	630	635	640
Gly Asp Val Pro Ala Asp His Val Glu Val Leu Glu Ala Val Gly Glu			
645	650	655	
Asp Pro Leu Glu Leu Thr Gly Arg Val Leu Glu Ala Val Gln Thr Trp			
660	665	670	

Leu Ala Asp Ala Ala Asp Asp Ala Arg Leu Val Val Val Thr Arg Gly
 675 680 685

Ala Val His Glu Val Thr Asp Pro Ala Gly Ala Ala Val Trp Gly Leu
 690 695 700

Ile Arg Ala Ala Gln Ala Glu Asn Pro Asp Arg Ile Val Leu Leu Asp
 705 710 715 720

Thr Asp Gly Glu Val Pro Leu Gly Arg Val Leu Ala Thr Gly Glu Pro
 725 730 735

Gln Thr Ala Val Arg Gly Ala Thr Leu Phe Ala Pro Arg Leu Ala Arg
 740 745 750

Ala Glu Ala Ala Glu Ala Pro Ala Val Thr Gly Gly Thr Val Leu Ile
 755 760 765

Ser Gly Ala Gly Ser Leu Gly Ala Leu Thr Ala Arg His Leu Val Ala
 770 775 780

Arg His Gly Val Arg Arg Leu Val Leu Val Ser Arg Arg Gly Pro Asp
 785 790 795 800

Ala Asp Gly Met Ala Glu Leu Thr Ala Glu Leu Ile Ala Gln Gly Ala
 805 810 815

Glu Val Ala Val Val Ala Cys Asp Leu Ala Asp Arg Asp Gln Val Arg
 820 825 830

Val Leu Leu Ala Glu His Arg Pro Asn Ala Val Val His Thr Ala Gly
 835 840 845

Val Leu Asp Asp Gly Val Phe Glu Ser Leu Thr Arg Glu Arg Leu Ala
 850 855 860

Lys Val Phe Ala Pro Lys Val Thr Ala Ala Asn His Leu Asp Glu Leu
 865 870 875 880

Thr Arg Glu Leu Asp Leu Arg Ala Phe Val Val Phe Ser Ser Ala Ser
 885 890 895

Gly Val Phe Gly Ser Ala Gly Gln Gly Asn Tyr Ala Ala Ala Asn Ala
 900 905 910

Tyr Leu Asp Ala Val Val Ala Asn Arg Arg Ala Ala Gly Leu Pro Gly
 915 920 925

Thr Ser Leu Ala Trp Gly Leu Trp Glu Gln Thr Asp Gly Met Thr Ala
 930 935 940

His Leu Gly Asp Ala Asp Gln Ala Arg Ala Ser Arg Gly Gly Val Leu
 945 950 955 960

Ala Ile Ser Pro Ala Glu Gly Met Glu Leu Phe Asp Ala Ala Pro Asp
 965 970 975

Gly Leu Val Val Pro Val Lys Leu Asp Leu Arg Lys Thr Arg Ala Gly
 980 985 990

Gly Thr Val Pro His Leu Leu Arg Gly Leu Val Arg Pro Gly Arg Gln
 995 1000 1005

Gln Ala Arg Pro Ala Ser Thr Val Asp Asn Gly Leu Ala Gly Arg Leu
 1010 1015 1020

Ala Gly Leu Ala Pro Ala Glu Gln Glu Ala Leu Leu Leu Asp Val Val
 1025 1030 1035 1040

Arg Thr Gln Val Ala Leu Val Leu Gly His Ala Gly Pro Glu Ala Val
 1045 1050 1055

Arg Ala Asp Thr Ala Phe Lys Asp Thr Gly Phe Asp Ser Leu Thr Ser

1060	1065	1070
Val Glu Leu Arg Asn Arg Leu Arg Glu Ala Ser Gly Leu Lys Leu Pro		
1075	1080	1085
Ala Thr Leu Val Phe Asp Tyr Pro Thr Pro Val Ala Leu Ala Arg Tyr		
1090	1095	1100
Leu Arg Asp Glu Phe Gly Asp Thr Val Ala Thr Thr Pro Val Ala Thr		
1105	1110	1115 1120
Ala Ala Ala Ala Asp Ala Gly Glu Pro Ile Ala Ile Val Gly Met Ala		
1125	1130	1135
Cys Arg Leu Pro Gly Gly Val Thr Asp Pro Glu Gly Leu Trp Arg Leu		
1140	1145	1150
Val Arg Asp Gly Leu Glu Gly Leu Ser Pro Phe Pro Glu Asp Arg Gly		
1155	1160	1165
Trp Asp Leu Glu Asn Leu Phe Asp Asp Asp Pro Asp Arg Ser Gly Thr		
1170	1175	1180
Thr Tyr Thr Ser Arg Gly Gly Phe Leu Asp Gly Ala Gly Leu Phe Asp		
1185	1190	1195 1200
Ala Gly Phe Phe Gly Ile Ser Pro Arg Glu Ala Leu Ala Met Asp Pro		
1205	1210	1215
Gln Gln Arg Leu Leu Leu Glu Ala Ala Trp Glu Ala Leu Glu Gly Thr		
1220	1225	1230
Gly Val Asp Pro Gly Ser Leu Lys Gly Ala Asp Val Gly Val Phe Ala		
1235	1240	1245
Gly Val Ser Asn Gln Gly Tyr Gly Met Gly Ala Asp Pro Ala Glu Leu		
1250	1255	1260

Ala Gly Tyr Ala Ser Thr Ala Gly Ala Ser Ser Val Val Ser Gly Arg
 1265 1270 1275 1280

Val Ser Tyr Val Phe Gly Phe Glu Gly Pro Ala Val Thr Ile Asp Thr
 1285 1290 1295

Ala Cys Ser Ser Ser Leu Val Ala Met His Leu Ala Gly Gln Ala Leu
 1300 1305 1310

Arg Gln Gly Glu Cys Ser Met Ala Leu Ala Gly Gly Val Thr Val Met
 1315 1320 1325

Gly Thr Pro Gly Thr Phe Val Glu Phe Ala Lys Gln Arg Gly Leu Ala
 1330 1335 1340

Gly Asp Gly Arg Cys Lys Ala Tyr Ala Glu Gly Ala Asp Gly Thr Gly
 1345 1350 1355 1360

Trp Ala Glu Gly Val Gly Val Val Val Leu Glu Arg Leu Ser Val Ala
 1365 1370 1375

Arg Glu Arg Gly His Arg Val Leu Ala Val Leu Arg Gly Ser Ala Val
 1380 1385 1390

Asn Ser Asp Gly Ala Ser Asn Gly Leu Thr Ala Pro Asn Gly Pro Ser
 1395 1400 1405

Gln Gln Arg Val Ile Arg Arg Ala Leu Ala Gly Ala Gly Leu Glu Pro
 1410 1415 1420

Ser Asp Val Asp Ile Val Glu Gly His Gly Thr Gly Thr Ala Leu Gly
 1425 1430 1435 1440

Asp Pro Ile Glu Ala Gln Ala Leu Leu Ala Thr Tyr Gly Lys Asp Arg
 1445 1450 1455

Asp Pro Glu Thr Pro Leu Trp	Leu Gly Ser Val Lys Ser Asn Phe Gly	
1460	1465	1470
His Thr Gln Ser Ala Ala Gly Val Ala Gly Val Ile Lys Met Val Gln		
1475	1480	1485
Ala Leu Arg His Gly Val Met Pro Pro Thr Leu His Val Asp Arg Pro		
1490	1495	1500
Thr Ser Gln Val Asp Trp Ser Ala Gly Ala Val Glu Val Leu Thr Glu		
1505	1510	1515 1520
Ala Arg Glu Trp Pro Arg Asn Gly Arg Pro Arg Arg Ala Gly Val Ser		
1525	1530	1535
Ser Phe Gly Ile Ser Gly Thr Asn Ala His Leu Ile Ile Glu Glu Ala		
1540	1545	1550
Pro Ala Glu Pro Gln Leu Ala Gly Pro Pro Pro Asp Gly Gly Val Val		
1555	1560	1565
Pro Leu Val Val Ser Ala Arg Ser Pro Gly Ala Leu Ala Gly Gln Ala		
1570	1575	1580
Arg Arg Leu Ala Thr Phe Leu Gly Asp Gly Pro Leu Ser Asp Val Ala		
1585	1590	1595 1600
Gly Ala Leu Thr Ser Arg Ala Leu Phe Gly Glu Arg Ala Val Val Val		
1605	1610	1615
Ala Asp Ser Ala Glu Glu Ala Arg Ala Gly Leu Gly Ala Leu Ala Arg		
1620	1625	1630
Gly Glu Asp Ala Pro Gly Leu Val Arg Gly Arg Val Pro Ala Ser Gly		
1635	1640	1645
Leu Pro Gly Lys Leu Val Trp Val Phe Pro Gly Gln Gly Thr Gln Trp		

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

Ala Leu Ala Gly Ile Asp Ala Arg Ala Pro Leu Val Pro Phe Leu Ser
 1860 1865 1870

Thr Leu Thr Gly Glu Trp Ile Arg Asp Glu Gly Val Val Asp Gly Gly
 1875 1880 1885

Tyr Trp Tyr
 1890

(2) Information für SEQ ID Nr: 3

(i) Sequenzeigenschaften:

(A) Länge: 53789 Basenpaare
 (B) Typ: Nukleinsäure
 (C) Strangart: Doppelstrang
 (D) Topologie: linear

(ii) Molekültyp: DNA (genomisch)

(xi) Sequenzbeschreibung: SEQ ID Nr: 3

```

GAATTCCAGG CCGTCGACGG CTGCGACATC GCGGTCTTCC GGTGGTCGCA CCGCACGAAG      60
ATCGCCGAAT AAGAATTTCG GGATCTCCCA CGGGAAAGGT TTCCATGACC GACGCAATAT      120
CCTTCGAGGT GCCGTGGGAC CGGACCGACA AGTTCGACCC GCCCGCGGTG TTCGACTCTC      180
TGCGCGAAGA ACGTCCGCTC GCGAAGATGG TTTACCCGGA TGGGCACGTC GGCTGGATCG      240
TTTCCAGCTA CGAGCTGGTC CGCGAGGTCC TCAGCGACCT GCGGTTTCAGC CACAGCTGCG      300
AAGTCGGCCA CTTCCCGGTG ACCCACCAGG GCCAGGTCAT CCCGACCCAC CCGCTGATCC      360

```

CCGGCATGTT CATCCACATG GACCCGCCCC AGCACACGCG CTACCGCAAG CTGCTGACCG	420
GCGAGTTCAC CGTCCGCCGC GCCAGCAGGC TGATCCCGCG GGCCGAGGCC GTGGCCGCCG	480
AGCAGATCGA GGTTCATGCGG GCCAAGGGCG CCCCCGCGGA CGTGGTCATG GACTTCGCCA	540
AGCCGCTGGT GCTGCGGATG CTGGGCGAGC TCGTCGGCCT GCCCTACGAG GAACGCGACC	600
GGTACGTGCC CGCGGTGACC CTCCTGCACG ACGCCGAAGC GGACCCGGCC GAGGCCGCGG	660
CCGCCTACGA GGTGGCCGGG AAGTTCTTCG ACGAGGTCAT CGAGCGCCGC CGGCAGCGGC	720
CCCAGGACGA CCTCATCAGC TCGCTCGTCA CCGAGGACCT GACCCAGGAG GAGCTGCGCA	780
ACATCGTCAC CCTGCTGCTG TTCGCCGGGT ACGAGACCAC CGAGGGCGCG CTCGCCACCG	840
GCGTCTTCGC GCTGCTGCAC CACACCGATC AGCTGGCGGC ACTGCGCGCG GAGCCGGAAA	900
AGCTCGACGC CGCGATCGAA GAGCTGCTGC GCTACCTGAC CGTCAACCAG TACCACACCT	960
ACCGCACCGC GCTGGAGGAC GTGAAGCTGG AGGGCGAGCT GATCAAGAAG GGCGACACGG	1020
TGACGGTGTC GCTGCCCGCG GCCAACC GCG ACGCGGCCAA GTTCGGCTGT CCCGCGGAGC	1080
TCGACATCGA GCGGGACACC TCCGGCCACG TCGCGTTCGG CTCGGCATC CACCAGTGCC	1140
TGGGCCAGAA CCTGGCGCGC ATCGAGCTGC GGGCCGGCTT CACGGCGCTC CTGCGGGCGT	1200
TCCCCGAGCT CCGGCTGGCC GTCCCGGCCG ACGAGGTTC GCTGCGGCTG AAGGGTTCG	1260
TCTTCTCGGT GAAGAAGCTG CCCGTCTCCT GGTGAGCGTT CTCCTCTCG AACACCCGAA	1320
AGGATCTGCG GCACAGTGCG CACCGATCTC ATCAAGCCAC TTCACGTGCG ACTCCTGGAG	1380
AACGCGACCC GCTTCGCCCG CAAGCCGGCC TTCGCCGACG ACCACCGGAC GGTACCTAC	1440
GGCGACCTCG AGGCGCGGAC GCGCCGGCTG GCCGGGCACC TGGCCGGCCT CCGTGTCCGG	1500

CACGGCGACC GGGTGGCGAT CTGCCTCGGC AACCGGGTGT CCACTGTGGA GAGTTACTTC	1560
GCGATCCTGC GCGCGGGTGC CGTCGGCGTG CCGCTCAACC CCGGTTCCGC GACGGCCGAG	1620
CTCGAGCACC CGCTGACCGA CAGCGGCGCC ACGGTGGTCG TCACCGACGC CGCCCAGGCG	1680
GCCCCGGCTCC GGCTCGCGCC GCACGTCGAG CTGCTGGTGA CCGGCGACGA CGTCCCAGAG	1740
GGCGCCCACT CCTACGACGA ACTCGCCCTC AGCGAACCGG CCGAGCCCGC CGCGGACGAC	1800
CTCGAGCTCG ACGAGCCGGC GTGGATGTTT TACACGTCGG GCACGACCGG GCGGCCCAAG	1860
GGCGTCGTGT CCACGCAGCG CAACTGCCTC TGGTCCGTCG CTTCTGCTA CGTGCCGTTT	1920
CCCGGGTTGT CGGACCAGGA CCGGGTGCTC TGGCCGCTCC CGCTGTTCCA CAGCCTTTTCG	1980
CACATCGCCT GCGTCCTGTC CGCCACCGTG GTCGGGGCCA GCGTCCGGAT CGCCGACGGC	2040
AGTCCGCCG ACGACGTGAT GCGGCTGATC GAGGCGGAGA GCTCGACCTT CCTGGCCGGC	2100
GTGCCGACCA CCTACCACCA CCTGGTGCGG GCCGCCCGGC AGCGCGGTTT CTCCGCGCCG	2160
AGCCTGCGGA TCGGCCTGGC CGGGGGCGCG GTCCTCGGCG CCGGGCTGCG AAGCGAGTTC	2220
GAAGAGACCT TCGGGGTCCC GCTGATCGAC GCCTACGGCA GCACCGAGAC CTGCGGGGCG	2280
ATCACCATGA ACCCGCCGGA CGGCGCCCGC GTCGAGGGCT CGTGCGGCTT GGCCGTGCCG	2340
GGCGTCGACG TCGGGTCTGT CGACCCCGAC ACCGGGCTCG ACGTCCCCGC CGGCGAGGAG	2400
GGCGAGGTCT GGGTCAGCGG GCCGAACGTC ATGCTCGGCT ACCACAACAG CCCGGAGGCG	2460
ACCGCCGCGG CGATGCGGGA CGGCTGGTTC CGGACCGGGG ACCTGGCCCG CCGCGACGAC	2520
GCCGGTTACT TCACCATCTG CGGCCGGATC AAGGAACTCA TCATCCGCGG CGGCGCGAAC	2580

ATCCACCCCG GCGAGGTCCA GGCGGTCCTG CGCACGGTCG ACGGCGTCGC GGACGCGGCG	2640
GTCGGCGGTG TGCCGCACGA CACGCTCGGC GAGGTGCCGG TGCCTACGT CATCCCCGA	2700
CCGACCGGTT TCGATCCTGC GCGTTGATC GAGAAGTGCC GCGAACAGCT GTCCGCCTAC	2760
AAGGTGCCGG ACCGGATCCT CGAGGTGCGC CACATTCCCC GGACCGCGTC GGGCAAGATC	2820
CGGCGCGGGC TGCTGACCGA CGAGCCCGCG CAGCTGCGGT ACGCCGCGAC CGAACACGAG	2880
GAACAGTCCC GGCACGCCGA CGAGTCCGTC GCGGCGGCGC TCGCGCGCG ACTGTCCGGT	2940
TTGGACGAAC GCGCCAGTG CGAGCTCCTG GAAGACCTCG TCCGCACCCA GGCGGCCGAC	3000
GTGCTGGGGC AGCCGGTCCC GGACGGGCGT GCGTTCCGCG ACCTCGGCTT CACGTGCTG	3060
GCCATCGTGG AGCTGCGCAA CCGGCTGACC GAGCACACCG GGCTCTGGCT GCGCGCCAGC	3120
GCCGTCTTCG ACCACCCAC GCCGGCGGCG CTGGCCGCC GCGTCCGGG TGAGCTCCTC	3180
GGGATCACGC AGGCCGTCGC GGAGCCGGTC GTCGCGGCCG ACCCGGGCGA GCCGATCGCG	3240
ATCGTGGGGA TGGCTCCCG CCTGCCGGGT GCGTGCGGT CCCCAGAAGA CCTGTGGCGG	3300
CTGGTGCCCG AGCGCGTCGA CGCCGTTTCG GAGTTCCCCG GCGACCGCGG CTGGGACCTG	3360
GACAGCCTGA TCGACCCGA CCGGGAGCGC GCCGGGACGT CGTACGTCGG CCAGGGCGGA	3420
TTCTGCACG ACGCCGGCGA GTTCGACGCC GGGTCTTCG GGATCTCGCC GCGTGAGGCC	3480
GTCGCGATGG ACCCGCAGCA GCGGTTGCTG CTGGAGACGT CGTGGGAGGC CCTCGAAAAC	3540
GCCGGAGTCG ACCCGATCGC GTTGAAGGGC ACCGACACCG GCGTGTTCTC CGGCCTCATG	3600
GGCCAGGGGT ACGGTCCGG CGCGGTGGCG CCGGAGCTCG AAGGTTTCGT CACCACCGGG	3660
GTGCGTCGA GCGTGGCCTC GGGCCGGGTG TCGTACGTGC TGGGACTGGA AGGCCCGGCG	3720

GTCACCGTGG ACACCGCGTG TTCGTCGTGG CTGGTCGCGA TGCACCTGGC CGCGCAGGCC	3780
CTGCGGCAGG GCGAATGCTC GATGGCGCTC GCCGGCGGGG TCACGGTGAT GGCCACGCCG	3840
GGCTCGTTCC TCGAGTTCTC CCGCCAGCGG GCCCTGGCGC CCGACGGGCG CTGCAAGGCC	3900
TTGCGGGCGG CGGCCGACGG GACCGGCTGG TCCGAGGGTG TCGGCGTGGT CGTCCTCGAG	3960
CGGCTGTCCG TGGCGCGCGA GCGGGGCCAC CGGATCCTGG CCGTTTTCGG TGGCAGCGCG	4020
GTCAACCAGG ACGGCGCGTC CAACGGGCTC ACCGCGCCGA ACGGCCTCTC GCAGCAGCGG	4080
GTCAATCCGC GCGCGCTGGC CGCGGCCGGG CTGGCACCCT CCGATGTGGA CGTCGTGAG	4140
GCGCAGGCA CCGGGACCAC GCTGGGTGAC CCGATCGAGG CGCAGGCCCT GCTGGCGACC	4200
TACGGCCAGG AGCGGAAGCA GCCGTGTGG CTCGGTTCGC TCAAGTCGAA CATCGGCCAC	4260
GCGCAGGCGG CCGCGGGCGT TCGGGCGTC ATCAAGATGG TGCAGGCGCT GCGGCACGAG	4320
ACCTTGCCGC CGACGCTGCA TGTCGACAAG CCGACTCTG AGGTGGACTG GTCCGCCGGT	4380
GCCATTGAAC TCTGACGGA GGCCCGTGG TGGCCGCGA ACGGCCGTCC GCGCCGGGCC	4440
GGGGTGTGCT CGTTCGGCGT CAGCGGGACC AACGCGCACC TGATCCTGGA GGAGGCGCCG	4500
GCGAGGAGC CGGTGCTGC CCCGGAAGTG CCGGTGCTGC CCCTGGTGGT GTCGGCGCGG	4560
AGCACGGAGT CGCTGTCCGG GCAGGCCGAG CGGCTGGCGT CCCTCCTCGA AGGGGACGTC	4620
TGCTGACCG AGGTGGCCGG GCGCTGGTG TCCCGCCGGG CCGTGCTGGA CGAGCGGGCC	4680
GTGTCGTGG CCGGTTCGG CGAGGAAGCC GTGACCGGC TCGGGGCGCT GAACACGGCC	4740
GTTTCGGGA CGCCGGGCAA GGTGCTGTGG GTGTTCGGG GGCAGGGGAC GCAGTGGGCC	4800

GGGATGGGCC GTGAGCTGCT GGCCGAGTCC CCGGTGTTTCG CCGAGCGGAT CGCCGAGTGC	4860
CGGCCCGCGT TGGCGCCGTG GATCGACTGG TCGCTCGTCG ACGTCCTGCG CGGCGAGGGC	4920
GACCTGGGTC GGGTCGATGT GCTGCAGCCG GCCTGTTTCG CCGTGATGGT CGGGCTGGCT	4980
GCCGTCTGGG AGTCCGTGGG GGTCCGGCCG GACGCCGTTCG TCGGGCACTC GCAGGGTGAG	5040
ATCGCGGCTG CCTGCGTTTC GGGGGCGTTG TCCCTCGAGG ACGCGGCGAA GGTGGTGGCC	5100
CTGCGCAGCC AGGCCATCGC GCGGAACTG TCCGGCCGCG GCGGGATGGC GTCGGTCGCC	5160
CTGGGCGAGG ACGACGTCGT TTCGCGGCTG GTGGACGGGG TCGAGGTCGC CGCCGTCAAC	5220
GGCCCGTCGT CCGTGGTGAT CGCCGGGGAT GCCCATGCCC TCGACGCGAC CCTGGAAATC	5280
TTGTCCGGGG AAGGCATCCG GGTTCGGCCG GTGGCGGTGG ACTACGCCTC GCACACCCCG	5340
CATGTCGAGG ACATCCGCGA CACTCTTGCC GAAACCTTGG CCGGGATCAG TGCGCAGGCG	5400
CCGGCTGTGC CGTTCTACTC CACCGTCACG AGCGAGTGGG TCGCGACGC GGGGGTGCTG	5460
GACGGCGGCT ACTGGTACCG GAACCTGCGC AACCAGGTCC GGTTCGGAGC GGCCGCGACG	5520
GCCCTGCTCG AGCAGGGCCA CACGGTGTTC GTCGAGGTCA GTCCGCACCC GGTGACGGTC	5580
CAGCCCTTGA GCGAGCTCAC CGGGGACGCG ATCGGGACAT TCGGGCGTGA AGACGGTGGC	5640
CTGCGGCGGT TGCTGGCTTC GATGGGTGAG CTGTTCTGTC CCGGCATCGA CGTGGACTGG	5700
ACGGCGATGG TGCCCGCGGC CGGCTGGGTC GACTTGCCGA CCTACGCGTT CGAACACCGG	5760
CACTACTGGC TCGAGCCCGC CGAGCCCGCT TCGGCCGAG ACCCGCTGCT GGGCACAGTC	5820
GTCAGCACTC CCGGTTCCGA CCGACTCACC GCCGTGGCGC AGTGGTCGCG CCGGGCGCAG	5880
CCCTGGGCGG TGGACGGCCT GGTGCCGAAC GCGGCCCTGG TCGAGGCGGC CATCCGGCTC	5940

GGCGACCTGG CCGGCACCCC CGTCGTCGGC GAACTGGTCG TCGACGCGCC GGTGGTGCTG	6000
CCGCGGCGCG GCAGCCGCGA GGTCCAGCTG ATCGTCGGCG AGCCCGGCGA GCAGCGGCGG	6060
CGTCCGATCG AGGTCTTTTC CCGGAAGCC GACGAGCCGT GGACGCGGCA CGCGCACGGC	6120
AACTCGCTC CCGCCGCCG TCGGTGCCA GAACCGGCGG CGGCGGGAGA CGCCACCGAC	6180
GTCACCGTGG CCGGCCTGCG CGACGCGGAC CGGTACGGGA TCCACCCCGC GCTGCTGGAC	6240
GCCGCCGTCC GCACGGTCGT CGGCGACGAC CTGCTCCCGT CGGTGTGGAC CGGCGTGTCC	6300
CTGCTGGCCT CCGGGGCCAC GGCCGTGACC GTGACGCCGA CGGCGACCGG CCTGCGGCTG	6360
ACCGACCCGG CCGGGCAGCC CGTCCTGACC GTCGAATCCG TCGCGGGCAC GCCGTTCGTC	6420
GCCGAGCAGG GGACCACCGA CGCGCTCTTC CGCGTCGACT GGCCGGAAAT CCCGCTGCCC	6480
ACCGCCGAAA CCGCGGACTT CCTGCCGTAC GAAGCCACGT CGGCCGAGGC GACCCTCTCC	6540
GCGCTCCAGG CCTGGCTGGC AGACCCCGCG GAAACCCGGC TGGCCGTGGT CACCGGGGAC	6600
TGCACCGAAC CCGGCGCGGC CGCGATCTGG GGCCTGGTGC GCTCGGCGCA GTCCGAACAC	6660
CCCGCCCGGA TCGTGCTGGC CGACCTCGAC GACCCCGCCG TGCTGCCCGC CGTGGTGGCG	6720
AGCGGCGAAC CGCAGGTGCG GGTGCGCAAC GGCCTCGCCT CGGTGCCCGC CTTGACCCGG	6780
GTTACTCCCC GGCAGGACGC GCGGCCGCTC GACCCCGAGG GCACCGTCCT GATCACCGGC	6840
GCACCGGCA CGCTCGGTGC GCTGACCGCC CGGCACCTCG TCACCGCGCA CGGCGTCCGG	6900
CACCTGGTGC TGGTCAGCCG CCGCGGTGAG GCTCCCGAGC TGCAGGAAGA ACTGACCGCA	6960
CTGGGGGCAT CCGTCGCCAT CGCCGCCTGC GACGTGGCAG ACCGGGCGCA GCTCGAAGCC	7020

GTCTTGCGCG CGATCCCGGC CGAGCACCCG CTCACCGCCG TGATCCACAC CGCGGGGGTC	7080
CTCGACGACG GCGTCGTAC CGAGCTGACC CCGGACCGGC TCGCCACCGT GCGGCGGCCG	7140
AAGGTCGACG CCGCCCGGCT CCTGGACGAG CTCACCCGGG AGGCCGATCT CGCCGCGTTC	7200
GTGCTGTTCT CCTCGGCGGC GGGTGTGCTG GGCAACCCCG GCCAGGCCGG GTACGCGGCC	7260
GCCAACGCCG AGCTGGATGC GTTGGCGCGC CAGCGGAACA GCCTCGACCT GCGCGCGGTG	7320
TCCATCGCAT GGGGCTACTG GCGGACGGTC AGCGGGATGA CCGAGCACCT GGGCGACGCC	7380
GACCTGCGGC GCAACCAGCG GATCGGCATG TCCGGGCTTC CCGCCGACGA GGGCATGGCG	7440
CTGCTGGACG CCGCCATCGC CACCGGTGGC ACGCTGGTCG CGGCCAAGTT CGACGTCGCC	7500
GCGCTGCGGG CGACGGCGAA GGCCGGCGGC CCGGTGCCGC CGCTGCTGCG TGGCCTGGCC	7560
CCGCTGCCGC GCCGGGCGGC GGCCAAGACC GCGTCGCTGA CCGAACGCCCT CGCCGGGCTG	7620
GCCGAGACCG AGCAGGCCGC GGCCCTGCTC GACCTGGTCC GCGGGCACGC CGCCGAGGTG	7680
CTCGGGCACA GCGGCGCCGA ATCCGTCCAT TCAGGACGGA CGTTCAAGGA CGCCGGCTTC	7740
GACTCGCTGA CCGCGTGGA ACTGCGGAAC CGCCTCGCGG CCGCGACCGG GCTCACCTG	7800
TCCCCGGCGA TGATCTTCGA CTACCCGAG CCCCCGGCGC TCGCGGACCA CCTGCGCGCC	7860
AAGCTCTTCG GATCGGCGGC GAACCGGCCG GCCGAGATCG GCACCGCCGC GGCCGAGGAG	7920
CCGATCGCGA TCGTCGCGAT GCGGTGCCGC TTCCCCGGTG GCGTGCACAG CCCCAGGAC	7980
CTGTGGCGGC TGGTCGCCGA CGGCGCCGAC GCCGTCACCG AGTTCCCCGC CGACCGCGGC	8040
TGGGACACCG ACCGGCTCTA CCACGAAGAC CCCGACCACG AAGGCACGAC GTACGTCCGG	8100
CACGGCGCCT TCCTCGACGA CGCCGCCGGG TTCGACGCCG CCTTCTTCGG CATCTCGCCG	8160

AACGAGGCGC TCGCCATGGA CCCGCAGCAG CGGCTGCTGC TGGAGACGTC CTGGGAGCTG	8220
TTGAGACGGG CCGCGATCGA CCCGACCACG CTGGCCGGCC AGGACATCGG CGTCTTCGCC	8280
GGCGTCAACA GCCACGACTA CAGCATGCGG ATGCACCGGG CCGCCGGTGT CGAGGGCTTC	8340
CGGCTCACCG GCGGTTCGGC CAGCGTGCTC TCCGGCCGCG TCGCCTACCA CTTCCGGCGTC	8400
GAAGGCCCGG CCGTCACGGT CGACACGGCC TGCTCGTCTT CGCTGGTCGC GCTGCACATG	8460
GCGGTGCAGG CCCTGCAGCG CGGCGAGTGC TCCATGGCGC TCGCGGGCGG CGTGATGGTG	8520
ATGGGCACGG TCGAGACGTT CGTCGAGTTC TCGCGGCAGC GCGGGCTGGC CCCCAGCGGC	8580
CGCTGCAAGG CGTTCGCCGA CGGCGCGGAC GGCACCGGCT GGTCCGAGGG CGTCGGGCTG	8640
CTCCTGGTGG AGCGGCTGTC CGAGGCTCAG CGTCGCGGGC ACCAGGTCCT CGCCGTGGTC	8700
CGCGGGTCGG CGGTCAACTC CGACGGCGCG TCGAACGECT TGACGGCCCC GAACGGCCCCG	8760
TCCAGCAGC GCGTGATCCG CAAGGCACTG GCCGCCGCC GACTGTCCAC ATCGGACGTC	8820
GACCGGTGG AGGCGCACGG CACCGGGACG ACCCTGGGCG ACCCGATCGA GGCCGAGGCG	8880
CTGCTGGCCA CCTACGGCCA GAACCGGGAA ACGCCGCTGT EGCTCGGGTC GGTGAAGTCG	8940
AACCTCGGGC ACACGCAGGC GCTGCGGGT GTCGCAGGCG TGATCAAGAT GGTGATGGCC	9000
ATGCGCCACG GCGTCCTGCC CCGGACGCTG CACCTCGACC GGCCGTCGTC CTATGTGGAC	9060
TGCTCGGCCG GTGCGGTCCA GCTGCTGACC GAGGCACGGG ACTGGGTGAG CAACGGCCAC	9120
CCGCGCCGCG CGGGCGTGTC GTCGTTCCGC ATCGGCGGCA CCAACGCGCA CGTCGTCTTC	9180
GAAGAGGTTG CCGCACCGAT CACCACGCCG CAGCCTGAGC CGGCCGAGTT CCTGGTGCCG	9240

GTGCTCGTCT CCGCGCGGAC GCGGGCGGGT CTGCGCGGCC AGGCCGGACG GCTCGCCGCG	9300
TTCTCGGCG ACCGGACCGA CGTCCGCGTC CCCGATGCCG CCTACGCACT GGCCACCACG	9360
CGCGCCAGC TCGACCACCG GGCCGTCGTC CTGGCCTCCG ACCGGGCACA GCTCTGCGCG	9420
GACCTTGCCG CGTTCGGCTC CGGCGTCGTG ACCGGAACGC CGGTTGACGG CAAGCTGGCC	9480
GTGCTCTTCA CCGGCCAGGG CAGCCAGTGG GCCGGGATGG GCCGTGAACT CGCCGAGACG	9540
TTCCCGGTCT TCCGCGACGC CTTCGAGGCC GCGTGCGAGG CCGTGGACAC GCACCTGCGT	9600
GAGCGTCCGC TCGCGGAGGT CGTGTTGAC GACAGCGCGC TGCTCGACCA GACGATGTAC	9660
ACCCAGGGCG CCCTGTTGCG CGTGGAGACC GCGTTGTTCC GGCTCTTCGA GTCCTGGGGT	9720
GTGCGGCCGG GTCTCCTCGC CGGTCACCTG ATCGGCGAGC TCGCCGCCGC GCACGTGTCC	9780
GGCGTGCTGG ACCTGGCCGA CGCGGGCGAG CTGGTCGCCG CGCGCGGCCG GCTGATGCAG	9840
GCCCTGCCCC CGGGCGGCGC GATGGTCGCC GTCCAGGCGA CCGAGGACGA AGTCGCGCCC	9900
CTGCTCGACG GCACGGTCTG CGTCGCCGCG GTCAACGGTC CGGACTCGGT GGTGCTCTCC	9960
GGCACC GAAG CCGCCGTGCT CGCCGTCGCG GATGAACTGG CTGGTCGCGG CCGTAAGACC	10020
CCACGGCTGG CCGTGAGCCA CGCCTTCCAC TCGCCGCTCA TGGAACCGAT GCTCGACGAC	10080
TTCCGCGCGG TCGCCGAACG CCTGACGTAC CGGGCCGGTT CGCTGCCCGT CGTCTCGACG	10140
CTGACCGGGG AACTCGCGGC GCTCGACAGC CCGGACTACT GGGTCGGCCA GGTGCGCAAC	10200
GCCGTGCGGT TCAGCGACGC CGTCACCGCG CTGGGCGCCC AAGGCGCGTC GACGTTCTC	10260
GAGCTCGGCC CGGGCGGTGC GCTCGCCGCG ATGGCGCTCG GCACGCTCGG CGGACCCGAG	10320
CAGAGCTGCG TCGCGACCTT GCGCAAGAAC GCGCCGAGG TGCCCGACGT CCTCACCGCG	10380

CTCGCCGAAC TGCACGTCCG GGGCGTGGGC GTCGACTGGA CGACCGTGCT CGACGAACCG	10440
GCCACGGCGG TCGGGACCGT CCTGCCCACC TACGCGTTCC AGCACCAGCG CTTCTGGGTC	10500
GACGTCGACG AAACAGCGGC CGTCAGCGTC ACCCCGCCGC CGGCGGAGCC GATCGTGGAC	10560
CGGCCGGTGC AGGACGTGCT GGAGCTGGTC CGGGAGAGCG CCGCGGTGGT GCTCGGGCAC	10620
CGGGACGCCG GCAGTTTCGA CCTCGACCGG TCCTTCAAGG ACCACGGCTT CGACTCGCTC	10680
AGCGCGGTCA AGCTGCGCAA CCGTCTGCGC GACTTCACCG GCGTGGAGCT GCCCAGCACC	10740
CTGATCTTCG ACTACCCGAA CCCGGCCGTC CTCGCGGACC ACCTGCGGGC CGAACTGCTC	10800
GGCGAGCGCC CGGCCGCGCC GGCCCCGGTG ACGAGGAGC TCTCCGACGA GCCGATCGCG	10860
ATCGTCGGCA TGAGCACCCG GCTGCCGGGT GGCGCCGACA GCCCCGAAGA GCTGTGGAAG	10920
CTCGTCGCGG AGGGACGGGA CGCCGTGTCC GGCTTCCCCG TCGACCGCGG CTGGGACCTC	10980
GACGGCCTCT ACCACCCGGA CCCC GCCCAC GCGGGACGA GCTACACGCG TTCGGGCGGC	11040
TTCTGACAG ACGCGGCCCA GTTCGACGCC GGGCTCTTCG GGATCTCACC GCGTGAGGCC	11100
CTGGCCATGG ACCCGCAGCA GCGGCTGCTG CTGGAGACGT CGTGGGAAGC CTTGGAGCGC	11160
GCGGGGGTCG ACCCGCTGTC CGCCCGCGGC AGCGACGTG GCGTCTTCAC CGGGATCGTC	11220
CACCACGACT ACGTGACCG GCTGCGCGAA GTGCCGAAG ACGTCCAGGG CTACACGATG	11280
ACCGGCACGG CTTGAGCGT GCGGTCGGGC CGGGTGGCGT ACGTCTTCGG CTTGAGGGC	11340
CCGGCGGTCA CCGTGGACAC CGCGTGTTTCG TCGTCGCTGG TCGCGATGCA CCTGGCGGCG	11400
CAGGCGCTGC GGCAGGGGGA GTGCTCGATG GCCCTGGCCG GCGGCGCGAC CGTGATGGCC	11460

AGCCCGGACG CCTTCCTCGA GTTCTCCCGC CAGCGCGGCC TGTCCGCGGA CGGCCGGTGC	11520
AAGGCGTACG CGGAAGGCGC GGACGGCAGC GGCTGGGCCG AGGCGCTCGG TGTCGTCGTC	11580
CTCGAACGGC TTTCGGTGGC ACGCGAACGT GGCCACCGGG TGCTGGCGGT CCTGCGCGGC	11640
AGCGCGGTGA ACCAGGACGG TGCTTCCAAC GGCCTGACCG CCCCGAACGG GCCGTCGCAG	11700
CAGCGGGTGA TCCGCGGCGC GCTGGCGAGC GCCGGGCTGG CACCGTCCGA TGTGGACGTC	11760
GTGGAGGGCC ACGGGACCGG GACCGCGCTG GGTGACCCGA TCGAGGTCCA GGCCTGCTG	11820
GCCACCTACG GGCAGGAGCG GGAACAGCCG TTGTGGCTCG GCTCGCTGAA GTCGAACCTC	11880
GGGCACACGC AGGCCGCGGC CGGGGTCTGT GCGGTGATCA AGATGATCAT GGCCATGCGC	11940
CACGGCGTCA TGCCGGCCAC GCTGCACGTC GACGAGCGCA CGAGCCAGGT CGACTGGTCG	12000
GCGGGCGCGA TCGAGGTGTT GACCGAGGCC CGGGAGTGGC CGCGCACCGG ACGTCCGCGC	12060
CGGGCCGGGG TGTCTCCTT CGGCGCCAGC GGCACCAACG CGCACCTGAT CATCGAGGAA	12120
GGTCCCGCCG AAGAGGCCGT GGACGAAGAG GTGGCCTCCG TGGTGCCGCT GGTCTCTTCC	12180
GCCCGCAGCG CCGGTTCGCT GGCCGGGCAG GCCGGGCGCC TGGCCGCGGT CCTCGAGAAC	12240
GAATCGTTGG CCGGGGTGGC CCGTGCCCTG GTTTCGGCC GCGCGACGCT GAACGAGCGC	12300
GCGGTCTGTA TCGCGGGCTC CCGCGACGAG GCCCAGGACG GCCTGCAGGC ACTGGCCCGC	12360
GGCGAGAACG CGCCCGGCGT CGTGACCGGG ACGGCGGGCA AGCCGGGCAA GGTCTCTTGG	12420
GTCTTCCCGG GCCAGGGCTC GCAGTGATG GGCATGGGCC GGGACCTCCT GGACTCCTCG	12480
CCGGTGTTCTG CCGCGCGGAT CAAGGAATGC GCTGCGGCAC TGGAACAGTG GACCGACTGG	12540
TCGCTGCTGG ACGTGCTGCG CGGCGACGCC GACCTGCTGG ACCGGGTCTGA CGTGGTGCAG	12600

CCGGCCAGCT TCGCGATGAT GGTCGGGCTC GCCGCGGTGT GGACCTCGCT GGGGGTGACC	12660
CCGGATGCGG TGCTCGGCCA CTCCCAGGGC GAGATCGCCG CGGCGTGCGT GTCCGGCGCG	12720
CTGTGCTGG ACACGCGGC GAAGGTGGTC GCGTTGCGCA GCCAGGCGAT CGCGGGGAG	12780
CTGGCGGGCC GCGGCGGGAT GGCGTCGGTC GCACTGAGCG AAGAGGACGC AGTCGCGCGG	12840
CTGACGCCGT GGGCGAACCG GGTCGAGGTG GCCGCGGTCA ACAGCCCGTC CTCGGTCGTC	12900
ATCGCGGGAG ACGCGCAGGC CCTCGACGAA GCCCTCGAAG CCCTGGCCGG CGACGGTGTC	12960
CEGGTCCGGC GGGTCGCGGT GGAATACGCC TCCCACACCC GGCACGTCGA GGCATCGCC	13020
GAAACCCTGG CCAAGACCTT GCCCGGGATC GACGCGCGGG TTCCGGCGAT TCCGTTCCTAT	13080
TCCACCGTCC TGGGCACGTG GATCGAGCAG GCCGTGCTCG ACGCGGGCTA CTGGTACCGG	13140
AACCTGCGGC AGCAGGTGCG GTTCGGCCCC TCGGTGGCGG ACCTGGCCGG GCTGGGGCAC	13200
ACGGTGTTTCG TGGAGATCAG CGCCACCCG GTGCTGGTCC AGCCGCTGAG CGAGATCAGC	13260
GACGACCGG TGGTGACCGG GTCGCTGCGG CGGACGACG GGGGACTGCG GCGCCTGCTG	13320
GCGTCGGCGG CCGAACTGTA CGTCCGGGGC GTGGCCGTGG ACTGGACGGC GGCCGTGCCC	13380
GCGGCCGGCT GGGTGGACCT GCCGACGTAC GCCTTCGACC GCCGCCACTT CTGGCTGCAC	13440
GAAGCCGAGA CCGCCGAAGC CGCCGAGGGC ATGGACGGCG AGTTCTGGAC GGCGATCGAA	13500
CAGTCCGATG TGGACAGCTT GGCCGAGCTG CTCGAGCTGG TGCCGGAGCA GCGCGGGGCG	13560
CTCAGCACCG TCGTGCCCGT GCTGGCGCAG TGGCGGGACC GCGCCGCGA GCGCTCGACC	13620
GCGGAGAAGC TCGCTACCA GGTACCTGG CAGCCCCTGG AGCGCGAAGC CGCCGGCGTG	13680

CCGGGCGGGC GCTGGCTGGC CGTCGTCCCG GCCGGCACCA CCGACGCGCT CCTGAAGGAG	13740
CTGACCGGCC AGGGACTCGA CATCGTCCGG CTGGAGATCG AGGAAGCTTC GCGGGCACAG	13800
CTCGCCGAGC AGCTGCGGAA CGTCCTGGCG GAGCACGACC TCACCGGCGT GCTGTCGCTG	13860
CTCGCTCTCG ACGGCGGGCC CGCGGACGCG GCCGAGATCA CCGCGTCGAC GCTCGCGCTG	13920
GTCCAGGCCC TGGGCGACAC CACCACGTCC GCGCCGCTGT GGTGCCTCAC TTCCGGCGCG	13980
GTGAACATCG GCATCCAGGA CGCCGTGACC GCACCGGCCC AGGCGGCCGT GTGGGGGCTC	14040
GGCCGGGCGG TCGCGCTGGA GCGCCTCGAC CGGTGGGGCG GCCTGGTCGA CTTGCCCGCC	14100
GCGATCGACG CCCGCACGGC TCAGGCCCTG CTCGGCGTCC TGAACGGCGC CGCCGGGGAA	14160
GACCAGCTCG CCGTCCGGCG CTCGGGCGTC TACCGCAGGC GGCTGGTCCG CAAGCCCGTG	14220
CCGGAGTCCG CGACGAGCCG GTGGGAACCC CGCGGCACGG TCCTGGTGAC CGGTGGGGCC	14280
GAAGGACTCG GCCGGCACGC CTCGGTCTGG CTCGCGCAGT CCGGCGCCGA ACGGCTCATC	14340
GTCACCGGCA CCGACGCGT CGACGAACTG ACGGCCGAGC TGGCCGAGTT CGGCACCACG	14400
GTCGAGTTCT GCGCCGACAC CGACCGGGAC GCGATCGCGC AGCTGGTGGC GGACTCGGAG	14460
GTCACCGCCG TGGTGACGC CGCGGACATC GCGCAGACCA GCTCCGTCGA CGACACCGGC	14520
GTGGCCGACC TCGACGAGGT GTTCGCCGCG AAGGTGACCA CCGCGGTGTG GCTGGACCAG	14580
CTGTTGAGG ACACCCGCT CGACGCGTTC GTCGTGTTCT CCTCGATCGC CGGCATCTGG	14640
GGCGGTGGCG GGCAGGGCCC GGCGGGTGCG GCGAACGCCG TCCTCGACGC CCTGGTCGAA	14700
TGGCGCCGGG CCCGCGCCT CAAGGCGACG TCGATCGCCT GGGGCGCGCT CGACCAGATC	14760
GGCATCGGCA TGGACGAGGC CGCCCTCGCC CAGCTGCGCC GCCGCGGTGT CATCCCGATC	14820

GCGCCGCCGC	TGGCGGTAC	CGCGATGGTG	CAGGCGGTGC	CCGGCAACGA	GAAGGCCGTG	14880
GCGGTGGCCG	ACATGGACTG	GGCCGCCTTC	ATCCCGGCGT	TCACCTCGGT	CCGGCCCAGC	14940
CCGCTGTTTC	CCGATCTGCC	CGAGGCGAAG	GCCATCCTCC	GGGCGGCGCA	GGACGACGGC	15000
GAAGACGGCG	ACACCGCGTC	GTCGCTCGCG	GACTCCCTGC	GCGCGGTCCC	CGACGCCGAG	15060
CAGAACCGCA	TCCTGCTGAA	GCTGGTCCGC	GGCCACGCTT	CGACGGTGCT	CGGCCACAGC	15120
GGCGCCGAAG	GCATCGGCCC	GCGCCAGGCG	TTCCAGGAGG	TCGGCTTCGA	CTCGCTGGCC	15180
GCGGTCAACC	TCCGCAACAG	CCTGCACGCG	GCCACCGGGC	TGCGGCTGCC	CGCGACGCTG	15240
ATCTTCGACT	ACCCACCCC	GGAGGCGCTG	GTCGGCTACC	TGCGCGTCGA	ACTCCTGCGG	15300
GAGGCCGACG	ACGGCCTGGA	CGGGCGGGAA	GACGACCTCC	GGCGAGTCCT	CGCGGCCGTG	15360
CCGTTGCCC	GGTTCAAGGA	GGCGGGCGTG	CTGGACACGC	TGCTCGGCCT	CGCCGACACC	15420
GGCACCGAAC	CGGGCACGGA	CGCCGAGACC	ACCGAAGCGG	CCCCGGCCGC	CGACGACGCA	15480
GAAGTGATCG	ACGCACTGGA	CATCTCCGGT	CTCGTGCAAC	GAGCCCTCGG	GCAGACGAGC	15540
TGACCGCCGA	TGGCGAACCA	ATCGTGAGG	AAGAACATGT	CCGCGCCGAA	CGAGCAGATC	15600
GTTGACGCAC	TGCGCGCGTC	GCTGAAGGAG	AACGTCCGGC	TTCAGCAGGA	GAACAGCGCG	15660
CTCGCCGCGG	CGCCCGCGGA	GCCCCTCGCG	ATCGTCTCCA	TGGCCTGCCG	CTACGCGGGC	15720
GGGATCCGCG	GCCCGGAGGA	CTTCTGGCGG	GTGGTGTCGG	AAGGCGCCGA	CGTCTACACC	15780
GGCTTCCCCG	AGGACCGCGG	CTGGGACGTC	GAAGGCCTCT	ACCACCCGGA	CCCCGACAAC	15840
CCCCGCACGA	CGTACGTGCG	GGAGGGCGCC	TTCTTCAGG	ACGCGGCCCA	GTTGACGCC	15900

GGGTTCTTCG GCATCTCGCC GCGCGAGGCG CTGGCCATGG ACCCCCAGCA GCGGCAGCTC	15960
CTGGAGGTGT CCTGGGAGAC CTTGGAACGG GCCGGCATCG ACCCGCATTC GGTGCGGGGC	16020
AGCGACATCG GCGTCTACGC CGGGGTCTGT CACCAGGACT ACGCCCCCGA CCTCAGCGGG	16080
TTCGAAGGCT TCATGAGCCT GGAGCGCGCC CTGGGCACCG CGGGCGGTGT CGCCTCCGGC	16140
CGGGTCGCCT ACACGCTCGG GCTCGAAGGC CCCGCCGTCA CCGTCGACAC GATGTGCTCG	16200
TCGTGCTGG TGGCGATTCA CCTTGCCGCG CAAGCTCTTC GCCGTGGTGA GTGCTCGATG	16260
GCCCTCGCGG GCGGCTCGAC CGTGATGGCG ACCCCGGGCG GGTTCGTGCG CTTCGCGCGT	16320
CAGCGGGCGT TGGCCTTCGA CGGGCGCTGC AAGTCCTACG CCGCGGCCGC CGACGGTTCC	16380
GGCTGGGCCG AGGGCGTCGG CGTGCTGCTG CTGGAGCGGC TGTGCGGTGGC GCGCGAGCGC	16440
GGGCACCAGG TGCTGGCCGT CATCCGCGGC AGCGCGGTCA ACCAGGACGG CGCTTCCAAC	16500
GGCCTGACCG CGCCCAACGG CCCGGCGCAG CAGCGGGTCA TCCGCAAGGC ACTGGCGAGC	16560
GCCGGGCTGA CACCGTCCGA TGTGGACACC GTGGAGGGCC ACGGCACCGG CACCGTCCTC	16620
GCGGACCCGA TCGAGGTCCA GCGCTGCTG GCCACCTACG GCCAGGGCCG CGACCCGCAG	16680
CAACCGCTGT GGCTGGGCTC GGTCAAGTCC GTCGTCGGGC ACACGCAGGC GGCATCCGGT	16740
GTGGCCGGCG TGATCAAGAT GGTCCAGTCG CTGCGGCACG GGCAGCTCCC GCGACCCAG	16800
CACGTCGACG CGCCACGCC GCAAGTGGAC TGGTCGGCCG GAGCGATCGA GCTGCTGGCC	16860
GAGGGCCGGG AGTGGCCGCG CAACGGCCAC CCGCGCCGGG GCGGCATCTC GTCGTCGGG	16920
GCCAGCGGCA CGAACGCGCA CATGATCCTC GAAGAAGCGC CCGAGGACGA GCCGGTGACC	16980
GAAGCGCCGG CGCCACGGG TGTCGTACCG CTGGTGGTGT CGGCGGCGAC CGCTGCTTCC	17040

CTGGCCGCC	AGGCCGGTCG	GCTGGCGGAG	GTCGGCGACG	TCTCCCTGGC	GGATGTCGCC	17100
GGGACGCTGG	TGTCCGGCCG	CGCGATGCTC	AGCGAGCGCG	CGGTCGTCGT	GGCCGGCTCC	17160
CACGAAGAAG	CCGTGACCGG	GCTGCGGGCG	CTGGCCCGCG	GCGAGAGCGC	GCCCCGCCCTG	17220
CTTTCCGGCC	GCGGCTCGGG	GGTCCCGGGC	AAGGTCGTCT	GGGTGTTCCC	CGGCCAGGGC	17280
ACGCAGTGGG	CCGGCATGGG	CCGCGAGCTG	CTGGACTCCT	CGGAGGTGTT	CGCCGCGCGG	17340
ATCGCCGACT	GCGAGACCGC	GCTCGGGCGG	TGGGTCGACT	GGTCGCTGAC	CGACGTGCTG	17400
CGCGGCGAGG	CCGACCTCCT	GGACCGGGTC	GACGTGGTGC	AACCGGCGAG	CTTCGCCGTG	17460
ATGGTCGGGC	TTGCCGCCGT	CTGGGCCTCC	CTCGGCGTCG	AGCCCGAGGC	CGTGGTGGGC	17520
CACTCGCAGG	GCGAGATCGC	GGCCGCATGC	GTGTCCGGGG	CACTGTCCCT	GGAGGACGCG	17580
GCGAAGGTGG	TGGCGTTGCG	CAGCCAGGCG	ATCGCCGCCT	CGCTGGCCGG	CCGGGGCGGC	17640
ATGCCGTCGG	TCGCGTTGAG	CGAAGAAGAC	GCGACCGCGC	GGCTCGAGCC	GTGGGCGGGC	17700
CGCGTGAGG	TCGCCGCCGT	CAACGGGCCG	ACGTCCGTGG	TGATCGCCGG	GGACGCCGAG	17760
GCGCTGGACG	AAGCCCTCGA	CGCGCTCGAC	GACCAAGGCG	TCCGGATCCG	GCGGGTGGCG	17820
GTGGACTACG	CCTCCACAC	CCGGCACGTC	GAAGCCGCGC	GCGACGCACT	GGCCGAGATG	17880
CTGGGCGGGA	TCCGCGCGCA	GGCGCCGGAA	GTECCGTTCT	ACTCGACCGT	GACCGGCGGC	17940
TGGGTCGAAG	ACGCCGGCGT	GCTCGACGGC	GGCTACTGCT	ACCGGAACCT	CCGCCGTCAG	18000
GTGCGGTTTCG	GCCCGGCGGT	GGCCGAGCTG	ATCGAGCAGG	GCCACCGGGT	GTTCGTCGAG	18060
GTCAGCGCGC	ATCCCGTGCT	GGTTCAGCCG	ATCAACGAAC	TCGTCGACGA	CACCGAAGCC	18120

GTGGTCACCG GGACGCTGCG GCGCGAGGAC GCGGGCCTCC GCGCCTGCT GGCTTCGGCG	18180
GCCGAGCTCT TCGTCCGCGG CGTGACCGTG GACTGGTCCG GTGTGCTGCC ACCGTCCCCG	18240
CGGGTCGAGC TGCCGACGTA CGCCTTCGAC CACCAGCACT ACTGGCTGCA GATGGGCGGG	18300
TCGGCCACCG ACGCCGTGTC GCTGGGCCTG GCCGGCGCCG ACCACCCGCT GCTGGGCGCG	18360
GTCGTCCCGC TGCCGAGTC CGACGGGCTC GTCTTCACCT CGCGGCTGTC GCTGAAGTCG	18420
CACCCGTGGC TGGCCGGGCA CGCGATCGGC GGGGTCGTGC TCATTCCGGG CACGGTGTAC	18480
GTCGACCTCG CGCTGCGCGC CGGCGACGAG CTCGGCTTCG GCGTCCTGGA AGAGCTCGTG	18540
ATCGAGGCAC CGCTGGTGCT GGGCGAGCGC GCGGGCGTTC GCGTGCAGGT CGCCGTGAGC	18600
GGGCCGAACG AGACCGGCTC GCGTGCGGTG GACGTCTTCT CCATGCGGGA AGACGGCGAC	18660
GAATGGACCC GGCACGCGAC CGGTCTCCTC GGGGCGTCGA CGTCCCGGA ACCGAGCCGC	18720
TTGCACTTCG CCGCCTGGCC GCCGGCCGGG GCGGAGCCGA TCGACGTCGA AAACCTTCTAC	18780
ACCGACCTCA CCGAGCGCGG GTACGCCTAC AGCGGCGCCT TCCAGGGCAT GCGGGCGGTC	18840
TGSCGGCGCG GTGACGAGGT CTTGCGCGAG GTCGCGCTGC CTGACGACCA CCGCGAGGAC	18900
GCCGGCAAGT TCGGCCTCCA CCCC GCCCTC CTCGACGCCG CTCTGCACAC GAACGCCTTC	18960
GCGAACCCGG ACGACGACCG CAGTGTGCTG CCGTTCGCGT GGAACGGCCT GGTCTGTCAC	19020
GCCGTGGGCG CGTCGGCGCT GCGGGTGCGG GTGGCGCCGG GCGGTCCGGA CGCGCTGACG	19080
TTCCAGGCCG CCGACGAGAC CGGTGGCCTG GTCCTACCA TGGATTGCT GGTGTCCCGC	19140
GAGGTGTGCG CCGCGCAGCT GGAGACGGCG GCGGGCGAAG AGCGCGACTC GCTGTTCCAG	19200
GTGGACTGGA TCGAGGTCCC CGCGACCGAG ACCGCGGCCA CCGAGCACGC CGAGGTGCTC	19260

GAAGCCTTCG GCGAGGCAGC GCCCCTCGAG CTGACCAGCC GGGTGCTGGA GGCCGTGCAG	19320
TCCTGGCTCG CCGACGCGGC CGACGAAGCA CGGTGGGTCTG TGGTGACCCG TGGCGCCGTG	19380
CGCGAGGTGA CGGACCCGGC CGGTGCCGCC GTGTGGGGTT TGGTGCGAGC CGCCCAGGCG	19440
GAGAACCCGG GCCGGATCAT CCTCGTCGAC ACCGACGGCG ACGTCCCGCT GGGTGCGGTG	19500
CTGGCCAGTG GTGAGCCGCA GCTCGCCGTG CGCGGCAACG CTTTCTCCGT CCCGCGCCTC	19560
GCCCCGGCCA CCGGCGAGGT GCCGGAGGCC CCCGCGGTGT TCAGTCCGGA AGGGACGGTC	19620
CTGCTCACC GCGGCACCG CTGCTGGGC GGTCTGGTGG CCAAGCACCT GGTGCCCCG	19680
CACGGCGTCC GCGGCTGGT GCTCGCCAGC CGCCGAGCGG TGGCCGCGGA AGACCTCGTC	19740
ACCGAGCTGA CCGAGCAGG CGCGACGGTG TCCGTGGTGG CTTGCGACGT CTCCGACCGC	19800
GACCAGGTGG CCGCGTTGCT GGCCGAACAC CGCCCGACCG GCATCGTGCA CCTGGCCGGC	19860
CTGCTGGACG ACGGCGTCAT CGGAGCCCTG AACC GGGAGC GGCTGGCCGG GGTGTTCCG	19920
CCCAAGGTG ATGCCGTCCA GCACCTCGAC GAACTGACCC GCGACCTCGG CCTCGACGCG	19980
TTCGTGCTGT TCTCGTCCGC AGCCGCGCTC ATGGGCTCCG CCGGCCAGGG CAACTACGCG	20040
GCCGCCAACG CTTTCCTCGA CGGCTTGATG GCCGGCGGCC GCGCGGCGGG CCTGCCAGGC	20100
GTGTCCCTGG CGTGGGGCCT GTGGGAGCAG GCGGACGGCC TGACCGCGAA CCTCAGCGCC	20160
ACCGACCAGG CCCGGATGAG CCGCGGCGGC GTGCTGCCGA TGACACCGGC CGAGGCCCTG	20220
GACATCTTCG ACATCGGCCT GGCCGCCGAG CAGGCCCTGC TGGTCCCGAT CAAGCTCGAC	20280
CTGCGGACGC TGCGCGGCA GGCCACCGCC GGCGGCGAGG TGCCGCACCT GCTGCGCGGC	20340

CTGGTCCGCG CGAGCCGCCG CGTGACCCGC ACGGCTGCCG CGAGTGGCGG CGGTGGCCTG	20400
GTCCACAAGC TCGCCGGGCG GCCAGCCGAA GAGCAGGAAG CCGTGCTGCT GGGCATCGTC	20460
CAGGCGGAGG CGGCCGCGGT GCTCGGCTTC AACGCCCCCG AGCTGGCCCA GGGCACCCGC	20520
GGGTTTACGG ACCTCGGCTT CGACTCGCTG ACCGCGGTG AGCTGCGGAA CCGGCTGAGC	20580
GCCGCGACCG GCGTCAAATT GCCCGCCACG CTCGTCTTCG ACTACCCGAC GCCGGTCGCG	20640
CTCGCCCGCC ACCTGCGCGA AGAGCTGGGC GAGACGGTGG CGGGTGCGCC GGCCACGCCG	20700
GTGACGACCG TCGCCGACGC GGGCGAGCCG ATCGCCATCG TCGGCATGGC GTGCCGCCTG	20760
CCGGGCGGCG TGATGAGCCC CGACGACCTC TGGCGGATGG TCGCCGAGGG CCGCGATGGG	20820
ATGTGCGCGT TCCCCGAGA CCGCGGCTGG GACCTGGACG GCCTGTTCGA CTCGGACCCC	20880
GAGCGCCCGG GCACCGCCTA CATCCGCCAA GCGGGCTTCC TGCACGAGGC CGCGCTGTTC	20940
GACCCGGGCT TCTTCGGGAT CTCGCCGCGC GAAGCCCTGG CCATGACCC GCAGCAGCGG	21000
CTGCTECTCG AAGCCTCCTG GGAAGCCCTG GAGCGCGCGG GCATCGACCC GACCAAGGCC	21060
CGCGGTGACG CCGTCGGCGT CTTCTCCGSC GTCTCCATCC ACGACTACCT CGAGTCCCTG	21120
AGCAACATGC CCGCCGAGCT CGAAGGCTTC GTCACCACGG CCACGGCGGG CAGCGTCGCC	21180
TCGGGCCGGG TGTCTACAC CTTTCGGGTTT GAGGGCCCCG CGGTCACGGT GGACACGGCG	21240
TGCTCGTCGT CGCTGGTCGC GATCCACCTG GCCGCACAGG CACTGCGGCA GGGCGAGTGC	21300
ACGATGGCCC TGGCCGGCGG TGTCGCCGTG ATGGGCTCGC CGATCGGTGT CATCGGCATG	21360
TCGCGGCAGC GCGGCATGGC CGAGGACGSC CGGGTCAAGG CGTTCGCCGA CGGCGCGGAC	21420
GGCACCGTCC TGTCCGAAGG CGTCGGCATC GTCGTCCTCG AACGGCTTTC GGTGGCCCCG	21480

GAACGCGGGC ACCGGGTGCT CGCCGTGCTC CGCGGCAGCG CGGTCAACCA GGACGGCGCT	21540
TCGAACGGCC TGACCGCGCC CAACGGGCCG TCGCAGCAGC GGGTGATCCG CAGCGCGCTG	21600
GCCGGGGCCG GACTGCAACC GTCCGAAGTG GACGTCGTCTG AAGCGCACGG CACCGGGACC	21660
GCGCTGGGCG AACCAGATCGA AGCCCAGGCC CTGCTGGCCA CCTACGGCAA GAGCCGCGAG	21720
ACGCCGTTGT GGCTCGGGTC GCTGAAGTCG AACATCGGCC ACACCCAGGC GGCCGCGGGC	21780
GTGGCGGCCG TGATCAAGAT GGTCCAGGCG CTGCGGCAGG ACACCCTGCC GCCGACCCTC	21840
CACGTGCAGG AACCACCAA GCAGGTGGAC TGGTCCGCGG GTGCGGTCCA GCTGCTGACC	21900
GAAGCCCGGG AGTGGGCCC GCAACGGCCAC CCGCGCCGGG CCGGTGTCTC GTCGTTCCGC	21960
ATCAGCGGCA CCAACGCGCA CCTCATCTTG GAAGAGGCGC CCGCCGACGA CACCGCCGAG	22020
GCGGACGTGC CCGACGCCGT GGTGCCCCTG GTGATCTCCG CGCGCAGCAC CGGATCCCTG	22080
GCGGGCCAGG CCGGACGCCT GCGGCGGTTT CTCGACGGAG ACGTCCCGCT GACCCGCGTG	22140
GCGGGTGCCC TGCTGTCGAC CCGGGCGACG CTGACCGACC GGGCCGTCGT CGTGGCGGGC	22200
TCGGCCGAGG ACGCCCGGGC GGGGCTGACC GCGCTGGCCC GCGGCGAGAG CGCGAGCGGG	22260
CTTGTGACCG GTACCGCAGG GATGCCGGGC AAGACGGTCT GGGTGTTCCT CGGCCAGGGG	22320
ACGCAGTGGG CGGGCATGGG CCGGGAGCTC CTCGAAGCGT CCCCAGTGTT CGCCGAGCGC	22380
ATTGAGGAAT GCGCGGCCG GCTGCAGCCG TGGATCGACT GGTGCTGCT GGACGTCTCT	22440
CGTGGCGAAG GTGAGCTGGA TCGGGTCGAC GTGCTGCAGC CGGCGTGTTT CGCGGTGATG	22500
GTGGGGCTGG CCGCCGTCTG GGCCTCGGTC GCGCTCGTGC CGGACCGGCT CCTGGGCCAC	22560

TCCCAGGGCG AGATTGCCGC CGCCTGCGTG TCGGGTGCAC TGTCCCTCGA GGACGCAGCC	22620
AAGGTCGTCTG CGCTGCGCAG CCAGGCGATC GCGGCGGAGC TGTCTGGGCCG CGGGGGCATG	22680
GCGTCGATCC AGCTGAGCCA CGACGAGGTG GCTGCCCGGC TCGCGCCGTG GCGGGGCCGC	22740
GTCGAGATCG CCGCCGTCAA CGGTCCGGCC TCGGTCGTGA TCGCCGGTGA CGCCGAAGCG	22800
CTCACCAGAG CCGTCGAAGT CCTCGGCGGT CGGCGGGTGG CCGTGGACTA CGCGTCCCAC	22860
ACGCGGCACG TCGAGGACAT CCAGGACACC CTCGCCGAGA CTCTGGCCGG GATCGACGCG	22920
CAGGCCCCCG TGGTGCCCTT CTACTCCACG GTCGCCGGCG AGTGGATCAC CGATGCCGGG	22980
GTCGTCGACG GCGGGTACTG GTACCGGAAC CTGCGCAACC AGGTCGGCTT CGGCCCCGCC	23040
GTGGCCGAGC TGATCGAGCA GGGGCACGGG GTGTTCGTCTG AGGTCAGTGC GCATCCGGTG	23100
CTGGTGACG CGATCAGCGA GCTCACCAGT GCGGTCTCTA CCGGGACGTT GCGGCGCGAC	23160
GACGGTGSGG TCGGCGGGCT GCTGACCTCG ATGECGGAAC TGTTCGTCCG CCGTGTCCCG	23220
GTCGACTGGG CCACGATGGC GCCGCCCGCG CGCGTCGAGC TGCCGACCTA CGCCTTCGAC	23280
CACCAGCACT TCTGGCTCAG CCCGCCCGCC GTGGCGGACG CGCCCGCGCT CGGCCTGGCC	23340
GGCGCCGACC ACCCGCTGCT GGGGGCGGTT CTCCCGCTGC CGCAGTCCGA CGGCCTGGTG	23400
TTACCTTCG GCCTGTCCGT GCGGACGCAT CCGTGGCTGG CCGACGGCGT CCCC GCCGCC	23460
GCCTTGGTGG AGCTGGCCGT GCGGGCCGGT GACGAAGCCG GTTGCCCGGT CCTCGCCGAC	23520
CTGACCGTCG AAAAGCTGCT GGTGCTGCCG GAGAGCGGTG GCCTGCGCGT CCAGGTGATC	23580
GTGAGCGGCG AGCGCACGGT CGAGGTGTAT TCGCAGCTCG AAGGCGCCGA AGACTGGATC	23640
CGGAACGCCA CCGGGCACCT GTCCGCCACG GCTCCGGCGC ACAGGCCTT CGACTTCACC	23700

GCCTGGCCGC CCGCCGGAGC CCAGCAGGTC GACGGCCTCT GGCGGCGCGG CGACGAGATC	23760
TTCCCCGAGG TCGCCCTGCC GGAGGAGCTG GACGCCGGCG CGTTCGGCAT CCACCCCTTC	23820
CTGCTGGACG CGGCCGTGCA GCCGGTCCTC GCGGACGACG AGCAGCCGGC GGAGTGGCGC	23880
AGCCTGGTCC TGCACGCCGC GGGTGCCTCG GCGCTGCGCG TCGGCTGGT GCCCGGCGGT	23940
GCCCTCCAAG CGGCGGACGA AACCGGCGGG CTGGTCCTCA CGGCGGATTC GGTGGCAGGC	24000
CGGGAACCTCT CGGCCGGGAA GACCCGCGCC GGATCGCTGT ACCGGGTGCA CTGGACCGAA	24060
GTGTCCATG CAGACAGTGC GGTGCCGGCC AACATCGAGG TCGTCGAAGC CTTCGGTGAA	24120
GAGCCCTGG AACTGACCG CCGGGTCCTG GAGGCTGTGC AGACCTGGCT CGTCACCGCG	24180
GCCGACGATG CGCGGCTGGT CGTGGTGACC CGCGGCGCCG TCGCGAGGT GACCGACCCC	24240
GCCGGTGCGG CCGTGTGGGG CCTGGTCCGA GCCGCGCAGG CGGAGAACCC CGGTGCGATC	24300
TTCTGATCG ACACCGACGG CGAGATCCCG GCCCTGACCG GTGACGAGCC CGAGATCGCG	24360
GTGCGCGGCG GGAAGTTCTT CGTGCCCCGC ATCACTCGCG CGGAGCCGAG CGGGGCCGCC	24420
GTGTTCCGCC CGGACGGGAC AGTGCTGATC TCGGGCGCGG GTGCGCTCGG TGGCCTGGTG	24480
GCCCGGCGTC TCGTCGAACG CCACGGCGTG CGGAAGCTCG TGCTGGCGTC CCGGCGCGGC	24540
CGAGACGCCG ACGCGGTGGC GGACCTGGTC GCCGACCTGG CCGCGGACGT GTCCGTGGTG	24600
GCTTGCGACG TCTCCGATCG CGCCAGGTG GCGGCCCTGC TCGACGAGCA CCGGCCGACC	24660
GCCGTCGTGC ACACCGCCGG CGTCATCGAC GCGGGCGTGA TCGAGACGCT GGACCGGGAC	24720
CGGCTGGCCA CGGTGTTGCG GCCGAAGGTC GACGCCGTGC GGCACCTCGA CGAGCTGACC	24780

CGCGACCGCG	ACCTCGACGC	CTTCGTCGTC	TACTCCTCGG	TCTCGGCCGT	GTTTCATGGG	24840
GCGGGCAGCG	GCAGTTACGC	CGCGGCGAAC	GCCTTCCTGG	ACGGCCTGAT	GGCGAACCGC	24900
CGGGCGGCGG	GCCTGCCGGG	CCTGTCGCTG	GCGTGGGGCC	TGTGGGACCA	GAGCACCGGT	24960
ATGGCCGCCG	GCACCGACGA	GGCCACCCGG	GCGCGGATGA	GCCGCCGCGG	TGGCCTGCAG	25020
ATCATGACGC	AGGCCGAGGG	CATGGACCTG	TTCGACGCCG	CGCTGTCGTC	GGCCGAGTCG	25080
CTGCTGGTGC	CCGCCAAGCT	CGACCTGCGT	GGGGTGC GCG	CCGACGCCGC	CGCGGGCGGG	25140
GTCGTGCCGC	ACATGCTGCG	TGGCCTGGTC	CGCGCGGGCC	GGGCGCAGGC	CCGCGCGGGC	25200
TCCACTGTGG	ACAACGGGCT	GGCCGGACGG	CTGGCCGGGC	TCGCCCCGGC	GGACCAGCTC	25260
ACGCTGCTCC	TGGACCTGGT	CCGGGCGCAG	GTCGCGGCCG	TGCTCGGGCA	CGCCGACGCG	25320
AGCGCCGTCC	GCGTCGACAC	GGCCTTCAAG	GACGCCGGCT	TCGACTCGCT	GACCGCGGTC	25380
GAGCTGCGCA	ACCGCATGCG	GACCGCCACC	GGCCTGAAGC	TGCCCCGCGAC	GCTCGTCTTC	25440
GACTACCCGA	ACCCCCAGGC	GCTCGCCCCG	CACCTGCGCG	ACGAACTCGG	TGGTGC GGCC	25500
CAGACGCCGG	TGACCACAGC	GGCCGCGAAG	GCCGACCTCG	ACGAGCCGAT	CGCCATCGTC	25560
GGGATGGCGT	GCCGCTTGCC	GGGCGGGGTC	GCCGGGCCCG	AGGACCTCTG	GCGGCTGGTC	25620
GCCGAGGGCC	GGGACGCGGT	GTCGAGCTTC	CCGACCGACC	GCGGCTGGGA	CACCGACAGC	25680
CTGTACGACC	CCGATCCGGC	CCGCCCGGGC	AAGACCTACA	CCCGGCACGG	CGGCTTCCTG	25740
CACGAAGCCG	GGCTCTTCGA	CGCGGGCTTC	TTCGGGATCT	CGCCACGCGA	GGCCGTCGCC	25800
ATGGACCCGC	AGCAGCGGCT	GCTGCTGGAG	GCCTCCTGGG	AGGCCATGGA	AGACGCCGGG	25860
GTCGACCCAC	TTTCGCTGAA	GGGCAACGAC	GTCGGCGTGT	TCACCGGCAT	GTTCGGCCAG	25920

GGTTACGTCTG CTCCCGGGGA CAGCGTCGTC ACGCCGGAGC TGGAGGGTTT CGCGGGCACG	25980
GGCGGGTCGT CGAGTGTGCG GTCCGGCCGC GTGTCTGTACG TGTTCGGGTT CGAAGGCCCG	26040
GCCGTGACGA TCGACTCGGC GTGCTCGTCC TCGCTGGTCG CGATGCACCT CGCCGCGCAG	26100
TCGCTGCGGC AGGGCGAGTG CTCGATGGCC TTGGCCGGCG GCGCGACGGT GATGGCGAAC	26160
CCCGGGCGCAT TCGTGGAGTT CTCGCGGCAG CGGGGCCTCG CCGTCGACGG TCGCTGCAAG	26220
GCGTTTCGCCG CCGCGGCCGA CGGCACCGGC TGGGCCGAGG GCGTCGGTGT GGTTCATCCTC	26280
GAGCGGCTGT CCGTGGCGCG GGAACCGCCG CACCGGATCC TGGCCGTGCT GCGCGGCAGC	26340
GCGGTCAACC AGGACGGCGC CTCGAACGGC CTGACCGCGC CGAACGGGCC GTCGCAGCAG	26400
CGGGTGATCC GCGGGGCGCT GGTGAGCGCC GGGCTGGCAC CGTCCGATGT GGACGTCGTC	26460
GAGGCGCACG GCACCGGGAC CACGCTGGGT GACCCGATCG AGGCGCAAGC TCTGCTGGCT	26520
ACCTACGGCA AGGACCGCGA GTCGCCGCTG TGGCTCGGCT CGCTGAAGTC GAACATCGGC	26580
CACGCGCAGG CCGCCCGGGG GGTCCCGGC GTCATCAAGA TGGTCCAGGC GCTCCGGCAC	26640
GAAGTCCTGC CGCCGACGCT GCACGTCGAC CGGCCTACCC CCGAGGTCGA CTGGTCGGCC	26700
GGTGCCGTCG AACTGCTGAC GGAAGCCCGC GAGTGGCCGC GCAACGGGCG CCCGCGCCGG	26760
GCCGGGGTCT CCGCGTTCGG CGTCAGCGGC ACGAACCGGC ACCTGATCCT GGAGGAGGCG	26820
CCCGCCGAAG AGCCGGTGCC CACACCCGAG GTTCCCCTGG TGCCGGTCGT GGTCTCCCGC	26880
CGGAGCAGGG CGTCCCTGGC CGGTCAGGCC GGTGCGCTCG CCGGATTCGT GGCGGGTGAC	26940
GCGTCCTTGG CCGGTGTGEC CCGGGCGCTG GTGACGAACC GGGCCGCGCT GACCGAGCGC	27000

GCGGTCATGG TCGTGGGCTC TCGCGAAGAA GCCGTGACGA ACCTGGAAGC GCTGGCCCCG	27060
GGCGAAGACC CCGCCGCGGT GGTACCCGGC CGGCGGGTT CGCCGGGCAA GCTCGTCTGG	27120
GTCTTCCCCG GCCAGGGCTC GCAGTGGATC GGGATGGGCC GGGAACTCCT GGAATCTTCG	27180
CCGGTCTTCG CCGAGCGGGT CGCCGAATGC GCGCCCGCCC TGGAACCGTG GATCGATTGG	27240
TCACTGCTCG ACGTGCTGCG CGGGGAGTCC GACCTGCTGG ACCGGGTCGA CGTCGTGCAG	27300
CCCGCCAGCT TCGCGATGAT GGTCCGGCTG GCCGCGGTGT GGCAGTCGGT GGGTGTCCGC	27360
CCGGATGCCG TCGTCGGCCA CTCGCAGGGC GAGATCGCCG CCGCCTGCGT CTCGGGCGCG	27420
CTGTGCTGC AGGACCGCCG GAAGGTGGTT GCCTGCGCA GCCAGGCGAT CGCCACCCGG	27480
CTGGCCGGG CCGGCGGCAT GGCTTCCGTG GCGTTGAGCG AAGAAGACGC GACCGCGTGG	27540
CTGGCGCCGT GGGCCGACCG GGTCCAGGTG GCCGCGGTCA ACAGCCCTGC CTCCGTGGTG	27600
ATCGCCGGGG AAGCCAGGC CCTCGACGAG GTCGTGACG CGTTGTCCGG TCAGGAAGTC	27660
CGCGTCCGGC GGGTGGCCGT GGAATACGGG TCCACACCA ACCAGGTCGA AGCCATCGAG	27720
GATCTGCTGG CCGAGACCTT GGCCGGCATC GAGGCGCAGG CCCCAGAGGT GCCCTTCTAC	27780
TCGACCCTGA TCGGTGACTG GATCCGTGAC GCCGGGATCG TCGACGGCGG CTAATGGTAC	27840
CGGAACCTGC GCAACCAGGT CCGGTTCGGT CCGGCCGTG CGGAGCTCGT TCGCCAGGGC	27900
CACGGGGTGT TCGTCGAGGT CAGCGCGCAC CCGGTGCTGG TCCAGCCGCT CAGTGAATC	27960
AGCGACGACG CCGTGGTGAC CCGGTGCTG CCGCGCAAG ACGGTGGCCT GCGCCGCTG	28020
CTGACGTCGA TGGCCGAGCT GTACGTGCAG GTGTCCCGC TCGACTGGAC CGCGGTCTTG	28080
CCGCGGACCG GCCGGTCTGA CCTGCCGAAG TACGCTTCG ACCACCGCA CTAATGGCTG	28140

CGGCCCGCCG AGTCCGCGAC CGACGCGGCT TCGCTGGGCC AGGCGGCGGC CGACCACCCG	28200
CTGCTGGGCG CGGTCTCGA GCTGCCGAG TCCGACGGCC TGGTGTTCAC CTCGCGGCTG	28260
TCCGTGCGGA CGCACCCTG GCTGGCCGAC CACGCGGTG GTGGCGTGGT CATCCTCCCC	28320
GGCTCCGGGC TGGCCGAAC TGGCGTCCG GCCGGCGAC AAGCCGGGTG CACCGCCCTC	28380
GACGAGCTGA TCATCGAAGC TCCGCTGGTC GTGCCCGCCC AAGGCGCGGT CCGCGTCCAG	28440
GTCGCGTTGA GCGGCCCGGA CGAGACCGGC TCGCGCACGG TGGACCTCTA CTCCCAGCGC	28500
GACGGCGGCG CGGGGACGTG GACGCGGCAC GCCACCGGCG TGCTGTGAC GGGCCCCGCT	28560
CAGGAACCCG AGTTCGACTT CCACGCCTGG CCGCCCGCG ATGCCGAGCG GATCGACGTC	28620
GAGACCTTCT ACACCGACCT GGCCGAGCG GTTACGGCT ACGGGCCGCG GTTCCAGGGG	28680
CTGCAAGCGG TGTGGCGGCG TGACGGCGAC GTCTTCGCG AGGTGCGCCT GCCCGAGGAC	28740
CTGCGCAAGG ACGCGGGCCG GTTCGGCGTC CACCCGGCG TGCTCGACGC GCGCTGCAG	28800
GCCGCCACCG CCGTGGGCG CGACGAGCCC GGTGAGCCG TGCTGGCGTT CCGGTGGAAC	28860
GGCCTGGTCC TGACGCGCG GGGCGGCTG GCCCTGCGGG TCCGGCTCGC GCGAGCGGC	28920
CCGACACGC TGTCCGTGGC AGCCGCCGAC GAAACCGGCG GCTTGGTCCT GACCATGGAA	28980
TCGCTGGTCT CCCGGCCGGT TTCGGCCGAG CAGCTCGGCG CCGCGGCCGA CGCGGGCCAC	29040
GACGCGATGT TCCGCTCGA CTGGACCGAG CTGCCTGCC TGCCCCGCG GGAAGTGGCG	29100
CCGTGGGTGC GGATCGACAC CGCCGACGAC GTCCGGCCT TGGCGGAGAA GGCGGACGCA	29160
CCACCGGTGG TGGTCTGGGA AGCCGCCGGG GGAGACCGG CCCTGGCCGT GAGTTCCCGG	29220

GTGCTCGAGA TCATGCAGGC CTGGCTGGCC GCGCCCGCGT TCGAGGAGGC CCGGCTGGTC	29280
GTGACGACCC GCGGCGCGGT ACCCGCCGGC GGTGACCACA CACTGACCGA CCCGGCCGCG	29340
GCCGCGGTGT GGGGCTGGT CCGGTCCGCG CAGGCGGAAC ACCCGGACCG GGTCGTCTTG	29400
CTGGACACCG ACGGCGAAGT TCCGCTGGGC GCGGTGCTGG CCTCCGGTGA GCCGCAGCTC	29460
GCGGTGCGCG GAACGACGTT CTTCGTGCCC CGGCTGGCCC GCGCCACCCG GCTCTCGGAC	29520
GCGCCTCTTG CGTTCGACCC GGACGGGACC GTGCTGGTCT CGGGCGCCGG ATCGCTGGGC	29580
ACCTTGGTGG CCCGGCACCT GGTACCCCG CACGGCGTGC GCCGGGTGGT GCTGGCCAGC	29640
CGGCAGGGCC GGGACGCCGA GGGCGCCAG GACCTGATCA CCGAGCTCAC CGGCGAAGGC	29700
GCGGACGTGT CCTTCGTGGC CTGTGACGTC TCCGATCGCG ACCAGGTGGC CGCGCTGCTC	29760
GCGGGCCTCC CGGACCTGAC CGGGGTGGTG CACACCGCCG GCGTCTTCGA GGACGGCGTG	29820
ATCGAGGCGC TGACGCCCGA CCAGCTCGCG AACGTGTACG CGGCCAAGGT CACGGCCGCG	29880
ATGCACCTCG ACCAGCTCAC CCGCGACCGG GATCTCGGCG CGTTCGTCTG GTTCTCCTCC	29940
GTCGCGGGGG TGATGGGTGG TGGCGGTCAA GGCCCGTACG CGGCGGCGAA CGCCTTCCTG	30000
GACGCGGCGA TGGCGAGTCG TCAGGCCGCG GGCCTGCCGG GCCTGTCCCT GCGGTGGGGC	30060
CTCTGGGAAC GCAGCAGCGG CATGGCCGCC CACCTCAGCG AGGTCGACCA CGCGCGGGCG	30120
AGCCGCAACG GTGTCTTGA ACTGACCCGG GCCGAGGGCC TGGCGCTGTT CGACCTCGGG	30180
CTGCGGATGG CCGAGTCGCT GCTCGTGCCG ATCAAGCTCG ACCTCGCCGC GATGCGGGCG	30240
AGCACGGTCC CGGTCTGTT CCGCGGCCCTG GTCCGGCCGA GCCGGACCCA GCGCGCACG	30300
GCGTCCACTG TGGACCGGG GCTGGCCGGG CGGCTCGCCG GGCTGCCGGT GGCCGAGCGG	30360

GCGGCGGTGC TGGTCGACCT GGTGCGCGGG CAGGTCGCGG TCGTGCTCGG CTACGACGGG	30420
CCGGAGGCCG TCCGCCCGGA CACGGCGTTC AAGGACACCG GGTTCGACTC GCTGACGTCG	30480
GTGGAAGTGC GCAACCGGCT GCGCGAGGCG ACCGGGCTCA AGCTCCCCGC CACGCTCGTC	30540
TTCGACTACC CGAACCCCTT GCGGTGGCG CGCTACCTGG GCGCGCGGCT GGTCCCGGAC	30600
GGGACCGCGA ACGGCAACGG GAACGGGAAT GGGCACAGCG AAGACGACCG GCTGCGGCAC	30660
GCGCTGGCGG CCATCCGCGC CGAGGACGCG GCGGAGGAGC GGTGATCGC CGACCTGGGC	30720
GTCGACGACC TCGTGCAACT GGCTTTCGGC GACGAGTGAT TGGGGCAAGT GGTGAGTGGC	30780
TCGTATGAAA AGGTCGTCGA GCGCTGCGG AAGTCGCTCG AAGAGGTCGG CACGCTGAAG	30840
AAGCGGAACC GGCAGCTCGC CGACGCGGCC GCGGAGCCGA TCGCCATCGT CGGCATGGCC	30900
TGCCGGCTGC CCGGTGGCGT CACCGGGCCC GGTGACCTCT GCGGGCTGGT GGCCGAGGGC	30960
GGCGACGCCG TCTCGGGGTT CCCCACCGAC CGCTGCTGGG ACCTGGACAC CCTGTTCGAC	31020
CCGGATCCCG ACCACGCGGG GACGTCGTAC ACCGACCAGG GCGGCTTCCT CCACGACGCG	31080
GCCCTGTTCG ACCCGGGCTT CTTCGGGATT TCGCCGCGCG AGGCGCTGGC CATGGACCCG	31140
CAGCAGCCGT TGCTGCTGGA GSCGTCTGG GAGGCGCTGG AAGGTGTCGG CCTCGACCCG	31200
GCTTCGTTGC AGGGCACCGA CGTCGGCGTG TTCACCGGCG CGGGCGGGTC GGGCTACGGC	31260
GGCGGCCTCA CCGGGCCGGA GATGCAGAGT TTCGCGGGCA CCGGGCTGGC CTCGAGCGTG	31320
GCTTCGGGCC GGGTGTCTTA CGTCTTCGGG TTCGAGGGAC CGGCGGTCAC GATCGACACG	31380
GCGTGCTCGT CGTCGCTGGT GGCATGCAC CTCGCCGCGC AGGCCCTGCG CCAAGGCGAC	31440

TGCTCGATGG	CACTGGCCGG	CGGCGCGATG	GTGATGTCCG	GCCCCGACTC	CTTCGTCTGC	31500
TTCTCCCGGC	AGCGGGGGCT	GGCCACCGAC	GGGCGGTGCA	AGGCGTTCGC	GTCGGGCGCC	31560
GACGGCATGG	TGCTCGCCGA	GGGCATCAGC	GTGGTCGTGC	TGGAGCGGCT	TTCCGTCTCGC	31620
CGGGAACGCG	GGCACC GG GT	GCTGGCCGTG	CTGCGCGGCA	GCGCGGTGAA	CCAGGATGGC	31680
GCGTCCAACG	GCCTGACCGC	CCCGAACGGC	CCTTCCCAGC	AGCGCGTGAT	CCGCGCCGCG	31740
CTEGCCAACG	CCGGAATCGG	ACCGTCCGAT	GTGGACCTCG	TCGAGGCGCA	CGGGACCGGG	31800
ACGAGCCTGG	GTGATCCCAT	CGAGGCGCAG	GCCTTGCTGG	CGACCTACGG	CCAGGACCGG	31860
GAGACGCCGT	TGTGGCTCCG	CTCGCTGAAG	TCGAACATCG	GGCACACGCA	GGCGGCCGCG	31920
GGCGTGCCGA	GCGTGATCAA	GGTCGTGCAG	GCGCTGCGGC	ACGGCGTCAT	GCCGCCGACC	31980
CTGCACGTGC	ACGAGCCCAG	CTCGCAGGTC	GA CTGGTCCG	AAGGCGCGGT	GGA ACTGCTG	32040
ACCGGGAGCC	GGGACTGGCC	GCGCGGGGAC	CGGCCGCGCC	GGGCCGGGGT	GTCGTCTGTT	32100
GGCGTCAGCG	GGACGAACGT	GCACCTGATC	ATCGAGGAAG	CCCCCGAGGA	GCCCGCTGCG	32160
GCCGTGCCGA	CGTCCGCGGA	CGTCGTGCCG	CTGGTG GTTT	CCGCACGCAG	CACGGGTTC	32220
CTGGCCGGTC	AGGCCGACCG	GCTGACCGAG	GTGGACGTCC	CCCTCGGACA	CCTCGCCGGG	32280
GCGCTEGTGG	CCGGGCGCGC	GGTGCTCGAG	GAACGCGCGG	TCGTGGTCCG	CGGTTCGGCC	32340
GAAGAAGCCC	GCGCGGGGCT	GGGTGCGCTG	GCTCGCGGTG	AAGCCGCGCC	CGGCGTCTGT	32400
ACCGGGACCG	CGGGCAAGCC	GGGCAAGGTC	GTCTGGGTGT	TCCCGGGACA	GGGGACGCAG	32460
TGGGTGGGCA	TGGGCCGGGA	GCTCCTCGAC	GCGTCCCCGG	TGTTCCCGCA	GCGGATCAAG	32520
GAGTGCGCGG	CGGCACTGGA	CCAGTGGAAC	GACTGGTCCG	TGCTGGACGT	CCTGCGTG GT	32580

GACGGTGACC TGGATTCTGT CGAGGTGCTG CAGCCCGCGT GCTTCGCGGT GATGGTGGGG	32640
CTGGCCGCGG TCTGGGAGTC GGCGGGGGTC CGGCCGGACG CCGTCGTCGG CCACTCGCAG	32700
GGCGAGATCG CCGCGGCCTG CGTGTCGGC GCGCTCACC TCGACGACGC CGCGAAGGTG	32760
GTGGCCCTGC GCAGCCAGGC GATCGCGGCG CGGCTGTCCG GCCGCGGCGG GATGGCGTCG	32820
GTGCGGTTGA GCGAGGACGA GGCGAACGCA CGGCTGGGTT TGTGGGACGG CCGGATCGAG	32880
GTGGCCGCGG TCAACGGCCC CGCCTCCGTG GTGATCGCGG GGGACGCCCA AGCCCTCGAC	32940
GAGGCTTTGG AGGTGCTGGC CGGGGACGGC GTCCGCGTCC GGCAGGTGCG GGTGCGACTAC	33000
GCCTCCACACA CCCGGCACGT CGAGGACATC CGCGACACCC TCGCCGAGAC GCTGGCCGGG	33060
ATCACCGCGC AGGCCCCGGA CGTGCCGTTC CGCTCCACCG TCACCGGCGG CTGGGTGCGG	33120
GACGCCGACG TCCTGGACGG CGGGTACTGG TACCGCAACC TCGCGAACCA GGTCCGGTTC	33180
GGCCCGGCCG TGGCCGAGCT GCTCGAGCAG GGCCACGGGG TGTTCGTGCA GGTGAGCGCC	33240
CACCCCGTGC TGGTGAGCC GATCAGCGAG CTCACCGACG CGGTCGTAC CCGGACGCTG	33300
CGGCGGACG ACGGCGGCCT GCGCCGCTG CTGACGTGCA TGGCCGAGCT GTTCGTCCGC	33360
GGTGTCGCG TCGACTGGGC CACGCTGGTG CCGCCCGCGC GCGTGGACCT CCCGACGTAC	33420
GCCTTCGACC ACCAGCACTT CTGGCTCCGG CCGGCCCGCG AGGCGGACGC CGTCTCGCTC	33480
GGCCAGGCCG CGGCGGAGCA CCCGCTGCTC GGCGCGGTG TCCGGCTGCC GCAGTCGGAC	33540
GGCCTGGTCT TCACCTCGCG GCTGTGCTG CGGACGCACC CGTGGCTGGC CGACCACACC	33600
ATCGGCGGCG TGGTGCTGTT CCGCGGCACC GGGCTGGTCG AACTGGCCGT GCGGGCCGGC	33660

GACGAGGCCG GGTGCCCGGT CCTGGACGAA CTCGTGACCG AGGCGCCGCT GGTGCTGCCC	33720
GGGCAGGGCG GAGTGAACGT CCAGGTCACG GTGAGCGGCC CGGACCAGAA CGGCTTGCGC	33780
ACGGTGGACA TCCACTCCCA GCGCGACGAC GTGTGGACCC GGCACGCGAC CGGAACGGTC	33840
TCGGCGACCC CGGCGAGCAG CCCC GGCTTC GACTTCACCG CGTGGCCGCC GCCGGACGGG	33900
CAGCGCGTCG AGATCGGCGA CTTCTACGCC GACCTCGCCG AGCGCGGGTA CGCGTACGGG	33960
CCCTTGTTCC AGGGCGTGCG GGCGGTGTGG CAGCGCGGCG AAGACGTGTT CGCCGAGGTC	34020
GCGCTGCCCC AAGACCGGCG GGAGGACGCC GCCCGGTTCTG GCCTGCACCC GCGGTTGCTG	34080
GACGCGGCCC TGCAGACCGG GACGATCGCC GCGGCCGCGT CCGGTCAGCC GGGCAAGTCC	34140
GTGATGCCGT TCTCGTGGA CCGGCTGGCG CTGCACGCCG TCGGGGCCGC GGGCCTCCGG	34200
GTCCCGTGCG CCCCCGGCGG ACCGGACGCG CTGACCGTCG AGGCCGCCGA CGAGACCGGC	34260
GCCCCGGTCC TCACCATGGA CTCGCTGATC CTGCGTGAAG TCGCCCTCGA CCAGCTGGAC	34320
ACTGCGCGCG CCGGCTCGCT CTACCGGGTG GACTGGACGC CACTGCCCAC TGTGGACAGT	34380
GCGGTGCCCC CTGGTCGGGC CGAGGTGCTG GAAGCTTTCTG GCGAGGAGCC CCTGGACCTG	34440
ACCGGCCGGG TGCTGGCCGC CCTGCAGGCG TGGCTTTCCG ACGCGGCGGA GGAAGCCCGC	34500
CTGGTCGTGG TGACCCGGGG TCGGCTGCCC GCCGGAGACG GTGTGGTGAG CGATCCGGCG	34560
GGTGCCGCGG TGTGGGGCCT GTTCCGGGCC GCGCAGGCGG AGAACCCGGA CCGGTTCTGTC	34620
CTGCTCGACA CCGACGCGCA GGTGCCGCTG GAAGCGGTGC TGGCGACCGG TGAGCCGCAG	34680
CTCGCGCTGC GCGGCACGAC GTTCTCGGTG CCCC GGCTCG CCCGCGTCAC CGAACCGGCG	34740
GAAGCCCCGC TGACGTTCCG TCCGGACGGG ACGGTCCTGG TCTCCGGCGC CGGGACGCTG	34800

GGTGCGCTCG CCGCCCCGGA CCTCGTCACC CGGCACGGCG TCCGGCGGCT CGTCTGGCC	34860
AGCCGGCGCG GCCGGGCCG CGAGGGCATC GACGACCTCG TCGCCGAGCT GACCGGGCAC	34920
GGCGCCGAAG TGACGGTCGC CGCCTGCGAC GTCTCCGACC GCGACCAGGT GGCGGCGCTG	34980
CTCAAGGAAC ACGCGCTGAC CGCGGTGGTG CACACGGCGG GCGTGTTTGA CGCCGGTGTC	35040
ACCGGCGCGC TGACCCGGGA GCGGCTGGCC AAGGTGTTTG CGCCCAAGGT CGACGCGGCC	35100
AACCACCTCG ACGAGCTGAC CCGCGACCTG GACCTCGACG CGTTCATCGT CTACTCGTCC	35160
GCCTCCTCGA TCTTCATGGG CGCGGGCAGC GGCGGGTACG CGGCGGCGAA CGCCTACCTC	35220
GACGECCTGA TGGCCGCCCC GCGCGCGGCG GGCTGCCC GGCTGTGCTT GGCCTGGGGC	35280
CCGTGGGAGC AGCTCACC GG CATGGCCGAC ACCATCGACG ACCTCACCCT GGCCCGGATG	35340
AGCCGGCGCG AAGGCCGCGG CGGCGTCCGC GCGCTCGGCT CCGCCGACGG CATGGAGCTG	35400
TTGACGCGG CGCTCGCGG CCGGCAGGCG CTGCTGGTGC CGATCGAGCT CGACCTGCGC	35460
GAGGTGCGGG CCGACGCGG CGGCGGCGG ACGGTGCGG ACCTGCTGCG CGGCTGGTTC	35520
CGCGCGGGCC GGCAGGCGG GCGGACGGG GCCACCGAGG ACGGCGGCCT GGAACGCCGG	35580
CTGGCCGGGC TCACCGTGGC CGAACAGGAA GCGCTGCTGC TCGACCTCGT CCGCGGTCAG	35640
GTCGCCGTCG TGCTCGGGCA CGCCGACAGC TCCGGCGTCC GCGCCGACGC GGCGTTCAAG	35700
GACGCCGGGT TCGACTCGCT GACGTGCGTG GAGCTGCGCA ACCGGCTGCG CGAGACGACC	35760
GGCTGAAAC TGCCCGCGAC GCTGGTCTTC GACCATCCGA ACCCGCTGGC ACTGGCCCGG	35820
CACCTGCGGG CGGAACTCGC CGTCGACGAG GCATCCCCGG CCGATGCGGT GCTGGCCGGG	35880

CTCGCCGGGC TGGAGGCGGC CATCGCGGCC GCCGCGCCC CGGACGGCGA CCGGATCACC	35940
GCGCGGCTGC GGGAACTGCT CAAGGCCGCC GAGGCGGCC AGGCCCGGCC GGGCACCTCC	36000
GGCGATCTCG ACACGGCCAG CGACGAGGAA CTGTTGCCCC TCGTCGACGG GCTCGACTGA	36060
AACCGCTGTG ACATCCGGGG CTTCGCCACC CGGCCCCGA AAAGCAAGCA CACGTGAGAG	36120
TTCTGGGAGT TGAGTTCAGT GGCTGACGAG GGACAACTCC GCGACTACCT CAAGCGGGCC	36180
ATCGCCGACG CCCGCGACGC CCGCACGCGG CTGCGCGAGG TCGAGGAGCA GCGCGGGAG	36240
CCGATCGCCA TCGTCGCCAT GGCGTGCCGG TACCGGGCG GGGTGTCCTC GCCCGAGGAC	36300
CTGTGGCGGC TGGTGGCCGA GGGGACCGAC GCCGTCTCCG CGTTCCCCGG CGACCGCGGC	36360
TGGGACGTGC ACGGGCTCGT CGACCCGGAC CCCGACCGCC CGGGCACGAC GTACACGGAC	36420
CAGGTGGCT TCCTCCACGA GGCCGGCCTC TTCGACGCGG GGTTCCTCGG GATCTCGCCG	36480
CGGGAGGCCG TCGCGATGGA CCCGCAGCAG CGGCTGCTGC TGGAGACGTC CTGGGAGGCC	36540
ATCGAACGCA CCGGCACCGA CCCGCTTTTC CTGAAGGGCA GCGACATCGG CGTCTTCACC	36600
GGCGTCGCGA GCATGGGTTA CGGCGCCGGT GCGGCGGTGG TCGCGCCGGA GCTGGAGGGT	36660
TTCTGCGCA CCGGTCCGGC GCCGTGCATC GCGTCCGCC GGGTGTGTA CGTCTCGGC	36720
TTGAAGGCC CGGCGTCCAC CGTCGACACC GGGTGTTCGT CGTCGCTGGT GCGATGCAC	36780
CTCGCCGCGC AGGCGCTGCG GCGGGGTGAG TGCTCGATGG CTCTGGCCGG CGGCGCGATG	36840
CTGATGGCCC AGCCGGGTTT GTTCGTGTCC TTCTCGCGC AACCGGGCT CGCCCTGGAC	36900
GGGCGCTGCA AGGCGTTTTC GGACAGCGCC GACGGGATGG GACTGGCCGA GGGCGTCGGC	36960
GTCATCGCGC TGGAACGGCT GTCGGTCGCC CGTGAGCGTG GGCACCGGT GCTGGCCGTG	37020

CTGCGCGGTA TCGCGGTGAA CCAGGATGGC GCGTCGAACG GCTTGACCGC CCCGAACGGC	37080
CCGTCCCAGC AGCGGGTGAT CCGCGCCGCG CTGGCCGAAG CCGGGCTGTC GCCGTCCGAT	37140
GTGGACGCCG TCGAAGGGCA CGGGACGGGC ACGACGCTGG GCGATCCGAT CGAAGCGCAG	37200
GCGTTGCTGG CCACCTACGG CAAGGGCCGG GACCCGGAGA AGCCGCTCTG GCTGGGCTCG	37260
GTGAAGTCGA ACCTCGGGCA CACGCAAGCG GCCGCGGGCG TGGCCAGCGT GATCAAGATG	37320
GTGCAGGCGC TGCGCCACGG CGTGCTGCCC CCGACGCTGC ACGTCGACCG GCCGTCCACC	37380
GAAGTCGACT GGTGCGCCGG TCGGCTCTCG CTGTTGACGG AGGCTCGGGA GTGGCCGCGC	37440
GAAGGGCGGC CGCGCCGGGC CGGGGTGTCC TCGTTCGGGA TCAGCGGGAC CAACGCGCAC	37500
CTCATCCTGG AGGAAGCGCC CGAGGAGGAG CCGCCCGTCG CCGAAGCGCC TTCCGCCGGA	37560
GTGGTGCCCG TGGTGGTGTC GGCTCGTGGG GCCCTGGCGG GTCAGGCCGG CCGGCTGGCC	37620
GCGTTCCTCG AGGCGTCCGA CGAGCCGTTG GTGACCGTCG CCGGGGCGCT GATCTGCCGC	37680
CGGTCCCGGT TCGGCGACCG GGCCGTCTGT GTGGCGGGCA CGCGCGCAGA GCGACGGCC	37740
GGGCTGGCCG CGCTGGCCCG CGGCGAAAGC GCCGCCGACG TCGTGACCGG CACGGTCGCG	37800
GCCTCGGGCG TGCCGGGCAA GCTCGTGTGG GTGTTCCCGG GCCAGGGTTC GCAGTGGGTG	37860
GGCATGGGCC GGGAGCTCCT CGAAGCCTCG CCGGTGTTCT CCGCGCGGAT CGCGGAGTGC	37920
GCGGCTGCCC TCGAACCGTG GATCGACTGG TCGCTGCTGG ACGTCTCTCG TGGCGAGGGC	37980
GACCTCGACC GCGTCGACGT GGTGCAGCCC GCGAGTTTCG CCGTGATGGT CGGCCTGGCC	38040
GCGGTGTGGT CGTCCGTCTG GGTGGTGCCC GACGCGGTGC TCGGGCACTC GCAGGGGGAG	38100

ATCGCGGCGG CGTGCGTGTC GGGGGCGTTG TCGCTGCAGG ACGCGGCGAA GGTGGTCGCG	38160
TTGCGCAGCC AGGCGATCGC GGCGAAGCTG GCCGGCCGCG GCGGCATGGC CTCGGTCGCG	38220
CTGAGCGAGG AAGACGCGGT CGCGCGGTTG CGGCACTGGG CGGACCGGGT CGAGGTGGCC	38280
GCGGTCAACA GCGCGTCGTC GGTGGTGATC GCCGGCGACG CCGAAGCCCT CGACCAGGCC	38340
CTCGAAGCAC TGACCGGCCA GGACATCCGG GTCCGGCGGG TGGCGGTGGA CTACGCCTCG	38400
CACACCCGGC ACGTCGAAGA CATCCAGGAG CCCCTCGCCG AGGCACTGGC CGGGATCGAG	38460
GCGCACGCGC CGACCCCTGCC GTTCTTCTCG ACCCTCACCG GTGACTGGAT TCGCGAAGCG	38520
GCGGTCGTGG ACGGCGGCTA CTGGTACCGG AACCTGCGCA ACCAGGTCGG TTTCGGCCCCG	38580
GCGGTGGCCG AGCTGCTCGG CCTCGGCCAC CGGGTGTTCTG TCGAGGTCAG CGCGCACCCC	38640
GTGCTCGTCC AGGCGATCAG CGCGATTGCC GACGACACCG ACGCGGTCTGT CACCGGCTCG	38700
CTGCGGCGCG AGGAGGGCGG CCTGCGGCGG CTGCTGACGT CGATGGCCGA GCTGTTCGTC	38760
CGCGGAGTCG ACGTGGAATG GGCCACGATG GTGCCGCCAG CGCGGGTCGA TTGCGCGACC	38820
TACGCCTTCG ACCACCAGCA CTACTGGCTG CGGTACGTCG AGACCGCGAC CGACGCGGCC	38880
GGTCCGGTGG TCCGGCTGCC GCAGACGGGC GGCCTGGTCT TCACCACCGA GTGGTCGCTG	38940
AAGTCACAGC CGTGGCTGGC CGAGCACACC CTGGAAGACC TGGTCGTCTGT CCCC GGCGCG	39000
GCACTGGTCG AGCTGGCCGT CCGGGCCGGT GACGAGGCCG GGACCCCGGT GCTGGACGAA	39060
CTCGTCATCG AGACGCCCCCT GGTCTGTGCC GAACGCGGCG CGATCCGGGT GCAGGTCACG	39120
GTGAGCGGAC CGGACGACGG CACACGGACC CTGGAAGTGC ATTCCCAGCC CGAAGACGCC	39180
ACCGACGAAT GGACCCGGCA CGCCACCGGC ACGCTGTCTG CGACCCCGGA CGAAAGCAGC	39240

GGGTTCGACT TCACGGCCTG GCCGCCCCCG GCGCCCCGGC AGCTCGACGG CGTTCCGGCG	39300
ATCTGGCGGG CCGGCGACGA GATCTTCGCC GAAGTCTCCC TGCCCGACGA TGCGGACGCC	39360
GAGGCATTTC GCATCCACCC CGCGCTCCTG GACGCGGGCC TGCACCCCGC CCTGCCCCGGC	39420
GATGACGGTC TGACGCAGCC CATGGAATGG CGTGGCCTGA CGCTGCACGC CGCGGGGGCG	39480
TGCACGCTGC GGGTCCGGTT GGTGCCCCGGC GGGTTCCTGG AAGCGGCCGA CGGCGCCGGC	39540
AGCCTGGTCG TCACGGCGAA GGAGGTGGCC CTCCGCCCCG TGACGATCGC GCGGTTCGCGC	39600
ACCACCACCC GAGACTCGCT GTTCAGCTG AACTGGATCG AGCTGCCCGA GAGTGGCGTG	39660
GTGGCCGCGG CAGACGACAC CGAGGTGCTG GAGGTGCCCC CGGGCGATTTC CCCGCTGGCG	39720
GCGACCTCCC GAGTCTTGGA GCGGCTCCAG ACCTGGCTGA CCGAGCCCGA GCGGGAACAG	39780
CTGGTCGTCG TGACGCGCGG CGCGGTGCCC GCCGGGGACA CCCCAGTGAC CGACCCGGCC	39840
GCGGCGGCGG TCTGGGGCCT GGTCCGGTCC GCGCAGGCGG AGAACCCGGA CCGGATCGTC	39900
CTGCTCGACA CCGACGGCGA AGTCCCGCTG GGTGCGGTGC TGGCCGGCGG CGAGCCGCAG	39960
GTGCGGGTGC GCGGCACGEC GCTGTACGTC CCGCGCCTGG CCGCGCCGA CGCGGCCCCG	40020
GTATCCGGTC TACATGGGAC GGTCCCTCGTC TCCGGTGCCG GTGTGCTCGG CGAGATCGTG	40080
GCGCGGCACC TGGTCACCCG CCACGGCGTG CGCAAGCTGG TGCTCGCCAG CCGCCGCGGC	40140
CTGGACGCCG ACGGCGCGAA GGACCTCGTC ACCGACCTCA CCGGCGAGGG CGCGGACGTG	40200
TCCGTCGTCG CCTGCGACCT GCGCGATCGG AACCAGGTGG CCGCGCTGCT GCGCGACCAC	40260
CGCCCGGCGA GCGTCATCCA CACGGCGGGC GTCCTCGACG ACGGCGTCAT CGGGACGCTG	40320

ACCCCGGAGC GGCTGGCCAA GGTGTTTCGCG CCCAAGGTCG ACGCGGTCCG CCATCTCGAC	40380
GAGCTGACTC GCGACCTCGA CCTCGACGCG TTCGTCTGT TCTCCTCCGG CTCCGGCGTG	40440
TTCGGTTCGC CGGGGCAGGG CAACTACGCG GCGGCGAAG CGTTCTCTGA CGCGGCGATG	40500
GCGAGCCGCC GCGCGGCGGG TCTTCCTGGT CTCTCGCTGG CGTGGGGCCT GTGGGAACAG	40560
GCCACCGGCA TGACCGCGCA CCTCGGCGGC ACCGACCAGG CCCGGATGAG CCGGGGCGGG	40620
GTGCGGCCGA TCACGGCCGA GGAAGGCATG GCCCTGTTCG ACACGGCACT GGGTGCGCAG	40680
CCCGCGCTGC TCGTGCCGGT CAAGCTCGAC CTGCGGGAGG TCGGGGCCGG CGGGGCCGTG	40740
CCGCACCTGC TCGCGGGGCT GTCCCGGGCC GGGCGGCGGC AGGCCCAAGC CGCGTCCACA	40800
GTGGACAACC AGCTGCTGGG CCGGCTGGCC GGGCTGGGCG CGCCCGAGCA GGAGGCGCTG	40860
CTCGTCGACC TCGTGCGCGG CCAGGTCGCG GCGGTGCTCG GGCACGCCGG GCCGGACCGG	40920
GTCCGCGCCG ACACGGCGTT CAAGGACGCC GGGTTCGACT CGCTCACCTC GTTCGACCTG	40980
CGCAACCGGC TCGGGGAGAG CACCGGGCTG AAGCTGCCCG CCACGCTCGC CTTGACTAC	41040
CCGACCCCGC TGGTCTTCGC CCGGCACCTG CGTGACGAGC TCGGGGCCGG CGACGACCGG	41100
CTTTCGGTGG TGCACGCGCG GCTCGAAGAC GTCGAGGCGC TGCTCGGCGG GCTGCGCCTC	41160
GACGAATCCA CGAAGACCGG TCTCACCTC CGGCTGCAGG GCCTGGTCGC CCGGTGCAAC	41220
GGCGTGAACG ACCAGACCGG CGGCGAAACG CTGGCGGACC GGCTCGAGGC CGCGTCCGCC	41280
GACGAAGTCC TCGACTTCAT CGACGAGGAG CTGGGTCTCA CCTGACCCCG GTTCGAGACC	41340
GACGTTCCAG CAACCCTTGT GAGGACCCGA GAATGGCCAC GGACGAGAAA CTCCTCAAAT	41400
ACCTCAAGCG CGTCACGCGG GAGCTGCACA GCCTGCGCAA GCAGGGTGCC CGGCACGCCG	41460

ACGAGCCGCT CGCCGTCGTC GGGATGGCCT GCCGGTTCCC GGGTGGGGTG TCCTCGCCCG	41520
AAGACCTGTG GCAGCTCGTG GCCGGCGGGG TCGACGCCCT TTCGGACTTC CCCGACGACC	41580
GGGGCTGGGA GCTGGACGGC CTGTTGACCG CGGACCCCGA CCACCCCGGG ACGTCGTACA	41640
CCAGCCAGGG CGGCTTCCTG CGTGGCGCCG GGCTGTTGGA CGCGGGCCTG TTCGGCATCT	41700
CGCCGCGCGA GGCCCTCGTC ATGGACCCGC AGCAGCGGGT GCTGCTGGAG ACGTCGTGGG	41760
AGGCCCTCGA AGACGCCGGG GTCGACCCGC TTTCGCTGAA GGGCAGCGAC GTCGGCGTGT	41820
TCTCCGCGCT CTTACCCAG GGCTACGGCG CCGGGCGGAT CACGCCGGAC CTCGAGGCGT	41880
TCGCGGGCAT CGGGCGGGC TCGAGCGTGG CGTCGGGCCG GGTGTCCTAC GTCTTCGGGC	41940
TCGAAGGACC GGCGGTCACC ATCGACACCG CGTGTTCGTC GTCGCTGGTG GCCATCCACC	42000
TCGCCGCGCA GGCCCTGCGC GCGGGCGAGT GCTCGATGGC GCTCGCCGGC GGGGCGACGG	42060
TGATGCCGAC GCCCGGCACC TTCGTGCGGT TCTCGCGGCA GCGGGTGCTG GCTGCCGACG	42120
GCCGGTCCAA GGCCTTCTCC TCGACCGCGG ACGGCACCGG CTGGGCCGAG GGCGCCGGGG	42180
TGCTCGTCCT CGAACGGCTT TCGGTGCGCG AGGAGCGCGG CCACCGGATT CTCGCCGTGC	42240
TGCGCGGCAG CGCGGTCAAC CAGGATGGCG CCTCCAACGG CCTGACCGCG CCGAACGGGC	42300
CTTCGCAGCA GCGGGTGATC CGCAAGGCGC TCGCGGGCGC CGGGCTGGTC GCGTCCGATG	42360
TGGACGTCGT GGAGGCGCAC GGCACGGGCA CCGCGCTGGG CGACCCGATC GAAGCGCAGG	42420
CGCTGCTGGC GACCTACGGC CAGGGCCGTG AGCGGCCGCT GTGGCTGGGG TCGGTCAAGT	42480
CGAACTTCGG GCACACGCAG GCGGCCGCCG GGGTCGCGGG CGTGATCAAG ATGGTCCAGG	42540

CCCTGCGGCA CGGCGCCATG CCGCCGACCC TGCACGTGGC CGAGCCGACG CCGGAGGTGC	42600
ACTGGTCCGC CGGTGCGGTG GAACTGCTGA CCGAGCCGCG CGAGTGGCCC GCCGGTGATC	42660
GGCCGCGCCG GGCCGGGGTG TCCGCGTTTC GGATCAGCGG GACGAACGCC CACCTGATCC	42720
TGGAGGAGGC GCGCCCGGCC GACGCGGTTC CGGAAGAACC GGAGTTCAAG GGGCCGGTGC	42780
CGCTGGTCGT CTCGGCGGGC AGCCCCACAT CTTTGCGCGC TCAGGCCGGC CGGCTCGCGG	42840
AGGTCCGTGC GTCCGGTGGT GTGTCCCGGG CCCGGCTGGC GAGCGGGCTG CTGTCCGGCC	42900
GGGCGCTGCT CGGTGACCGC GCGGTGCTGG TCGCGGGAAC GGACGAGGAC GCGGTGGCCG	42960
GCTTGCCTGC GCTGGCCCGC GGGGACCGCG CGCCCGGCGT GCTGACCGGT TCGGCCAAGC	43020
ACGGCAAGGT CGTCTACGTC TTCCCCGGCC AGGGTTCCGA GCGGCTCGGG ATGGGCCGCG	43080
AGCTCTACGA CCGGTACCCG GTGTTCGCGA CGGCGTTCGA CGAGGCTTGC GAGCAGCTGG	43140
ACGTCTGTCT GGCCGGCCGT GCCGGGCACC GCGTGCGGGA CGTCGTGCTC GCGGAAGTGC	43200
CCGCCGAAAC CCGGCTGCTG AACCAGACGG TCTTCACCCA AGCCGGGCTG TTCGCGGTGG	43260
AGAGCGCGCT GTTCCGGCTC GCCGAATCCT GGGGTGTCCG GCCGGACGTG GTGCTCGGCC	43320
ACTCCATCGG GGAGATCACC GCCGCGTATG CCGCGGGCGT CTTCTCGCTG CCGGACGCCG	43380
CCCGGATCGT CGCGGCGCGC GGCCGGCTCA TGCAGGCGCT GGCGCCGGGC GGGGCGATGG	43440
TCGCCGTGCG CGCCTCCGAA GCCGAGGTGG CCGAACTGCT CGGCGACGGC GTGGAACTCG	43500
CCGCCGTCAA CGGCCCTTCG GCGGTAGTCC TTTCCGGGGA CGCGGACGCG GTCGTCCGGG	43560
CCGCCGCCCG CATGCGCGAG CGCGGGCACA AGACCAAGCA GCTCAAGGTT TCGCACGCGT	43620
TCCACTCCGC GCGEATGGCG CCGATGCTGG CGGAGTTCGC CGCCGAGCTG GCCGGCGTGA	43680

CGTGGCGCGA GCCGGAGATC CCGGTGGTCT CCAACGTGAC CGGCCGGTTC GCCGAGCCCC	43740
GCGAACTGAC CGAGCCGGGC TACTGGGCCG AGCACGTGCG GCGGCCGGTG CGGTTCGCCG	43800
AGGGCGTCGC GGCCGCGACG GAGTCCGGCG GCTCGCTGTT CGTGGAGCTC GGGCCGGGGG	43860
CGGCGCTGAC CCCCCGTC GAGGAGACGG CCGAGGTCAC CTGCGTCGCG GCCCTGCGGG	43920
ACGACCGCCC GGAGGTCACC GCGCTGATCA CCGCGGTGCG CGAGCTGTTC GTCCGCGGGG	43980
TTGCGGTGCA TTGGCCGGCC CTGCTGCCCG CGGTCACCGG GTTCGTCGAC CTGCCGAAGT	44040
ACGCCTTCGA CCAGCAGCAC TATTGGCTGC AGCCCGCCGC GCAGGCCACG GACGCGGCCT	44100
CGCTCGGGCA GGTGCGGGCC GACCACCCGC TGCTGGGCGC GGTGGTCCGG CTGCCGAGT	44160
CGGACGGCCT GGTCTTCACC TCGCGGCTGT CATTGAAATC GCACCCGTGG CTGGCCGACC	44220
ACGTCATCGG CGGGGTGGTG CTCGTCGCGG GCACCGGGCT CGTCGAGCTG GCCGTCCGGG	44280
CCGGGACGA GGCCGGCTGC CCGGTCCTCG AAGAACTCGT CATCGAGGCT CCGCTGGTCG	44340
TCCCCGACCA CGGCGGGGTC CGGATCCAGG TCGTCGTGGG GGCACCGGGG GAGACCGGTT	44400
CGCGCGCGGT CGAGGTGTAC TCCCTGCGCG AGGACGCCGG TGCCGAAGTG TGGGCCCGGC	44460
ACGCCACCGG GTTCCTGGCT GCGACGCCGT CGCAGCACA GCGGTTCGAC TTCACCGCCT	44520
GGCCGCCGCC CGGCGTCGAG CGCGTCGACG TCGAGGACTT CTACGACGGC CTCGTCGACC	44580
GCGGGTACGC CTACGGGCCG TCGTTCGGG GCCTGCGGGC GGTGTGGCGG CGCGGCGACG	44640
AAGTGTTCGC CGAGGTGCGC CTGGCCGAGG ACGACCGCGC GGACGCGGCC CGGTTCGGCA	44700
TCCACCCCGG CCTGCTGGAC GCCGCCCTGC ACGCGGGCAT GGCCGGTGCC ACCACCACGG	44760

AAGAGCCCCG CCGGCCGGTG CTGCCGTTG CCTGGAACGG CCTGGTGCTG CACGCCGCCG	44820
GGGCGTCCGC GCTGCGGGTC CGGCTCGCCC CGAGCGGTCC GGACGCCCTG TCGGTCGAGG	44880
CCGCCGACGA GGCCGGCGGT CTCGTTGTGA CGGCGGACTC GCTGGTCTCC CGGCCGGTGT	44940
CGGCCGAACA GCTGGGCGCG GCGGCGAACC ACGACGCGTT GTTCCGCGTG GAGTGGACCG	45000
AGATTTCCCTC GGCTGGAGAC GTTCCGGCGG ACCACGTCGA AGTGCTCGAA GCCGTCGGCG	45060
AGGATCCCCCT GGAAGTGACC GGCCGGGTCC TGGAGGCCGT GCAGACCTGG CTCGCCGACG	45120
CAGCCGACGA CGCTCGCCTG CTCGTGGTGA CCCGCGGCGC CGTCCACGAG GTGACTGACC	45180
CGGCCGGTGC CGCGGTGTGG GGCCTGATCC GGGCCGCGCA GCGGAAAAC CCGGACCGGA	45240
TCGTGCTGCT GGACACCGAC GGTGAAGTGC CGCTAGCCCG GGTGCTGGCC ACCGGCGAGC	45300
CCCAAACAGC CGTCCGAGGC GCCACGCTGT TCGCCCCGCG GCTGGCCCCG GCCGAGGCCG	45360
CGGAGGCACC GGCAGTGACC GGCGGGACGG TCCTGATCTC GGGCGCCGCG TCGCTGGGCG	45420
CGCTCACC GC CGGCACCTG GTCGCCCCGC ACGGAGTCCG GCGGCTGGTG CTCGTCAGCC	45480
GCCGTGCCCC CGACGCCGAC GGCATGGCCG AACTGACCGC TGAATCATC GCTCAGGGCG	45540
CCGAGGTCCG CGTAGTCGCT TCGACCTGG CCGACCGGA CCAGGTCCG GTACTGCTGG	45600
CCGAGCACCG CCCGAACGCC GTCGTGCACA CGGCCGGTGT TCTCGACGAC GCGGTCTTCG	45660
AGTCGCTGAC GCGGGAGCGG CTGGCCAAGG TCTTCGCGCC CAAAGTTACT GCTGCCAATC	45720
ACCTCGACGA GCTGACCCGC GAACTGGATC TTCGCGCGTT CGTCGTGTTT TCCTCCGCCT	45780
CCGGGGTCTT CGGCTCCGCC GGGCAGGGCA ACTACGCCG TGCCAACGCC TACCTGGACG	45840
CCGTGGTCGC CAACCGCCG GCCGCGGGCC TGCCCGGCAC ATCGCTGGCC TGGGGCCTGT	45900

GGGAACAGAC CGACGGGATG ACCGCGCACC TCGGCGACGC CGACCAGGCG CGGGCGAGTC	45960
GCGGCGGGGT CCTCGCCATC TCACCCGCCG AAGGCATGGA GCTGTTCGAC GCAGCGCCGG	46020
ACGGGCTCGT CGTCCCGGTC AAGCTGGACC TCGCAAGAC CCGCGCCGGC GGGACGGTGC	46080
CGCACCTGCT GCGCGGCCTG GTCCGCCCGG GACGGCAGCA GGCCCGTCCG GCGTCCACTG	46140
TGGACAACGG ACTGGCCGGG CCACTCGCCG GGCTCGCGCC GGCGGAGCAG GAGGCGCTGC	46200
TGCTCGACGT CGTCCGCACG CAGGTCGCGC TGGTGCTCGG GCACGCCGGG CCGGAGGCCG	46260
TCCGCGCGGA CACGGCGTTC AAGGACACCG GCTTCGACTC GCTGACGTGC GTGGAACATC	46320
GCAACCGGCT GCGCGAGGCG AGCGGGCTGA AGCTGCCCCG GACGCTCGTC TTCGACTACC	46380
CGACGCCGGT CGCGCTGGCC CGCTACCTGC GTGACGAACT CCGCGACACG GTGGCAACAA	46440
CTCCGGTGCC CACCGCGGCC GCAGCGGACG CCGGCGAGCC GATCGCCATC GTCGGCATGG	46500
CGTGCCGGCT GCCGGCGGG GTCAACGATC CCGAAGGCCT GTGGCGCCTG GTGCGCGACG	46560
GCCTCGAAGG GCTGTCTCCC TTCCCCGAGG ACCGGGGCTG GGACCTGGAG AACCTGTTTCG	46620
ACGACGACCC CGACCGCTCC GGCACGACGT ACACCAGCCG GGGCGGGTTC CTCGACGGCG	46680
CCGGCCTGTT CGACGCGGGC TTCTTCGGGA TTTCGCCGCG CGAGGCGCTG GCCATGGACC	46740
CGCAGCAGCG GCTGCTGCTC GAGGCGGCCT GGAAGCCCT CGAAGGCACC GGTGTGACCC	46800
CGGGCTCGTT GAAGGGCGCC GACGTCGGGG TGTTCCGCCG GGTGTCCAAC CAGGGCTATG	46860
GGATGGGCGC GGATCCGGCC GAACTGGCGG GGTACGCGAG CACGGCGGGC GCTTCGAGCG	46920
TCGTCTCGGG CCGAGTCTCG TACGTCTTCG GGTTCGAAGG ACCGGCGGTC ACGATCGACA	46980

CGGCTTGCTC GTCGTCGCTG GTGGCGATGC ACCTGGCCCG GCAGGCGCTG CGGCAGGGCG	47040
AGTGCTCGAT GGCCCTGGCC GGTGGCGTCA CGGTGATGGG GACGCCCGGC ACGTTCGTGG	47100
AGTTCGCGAA GCAGCGCGGC CTGGCCGGCG ACGGCCGTG CAAGGCCTAC GCCGAAGGCG	47160
CGGACGGCAC GGGCTGGGCC GAGGGCGTCG GGTGCTCGT GCTGGAGCGG CTGTCCGTGG	47220
CGCGCGAGCG CGGGCACCGG GTGCTGGCCG TGCTGCGCGG CAGCGCGGTC AACTCCGACG	47280
GCGCGTCCAA CGGCCTGACC GCCCCAACG GGCCGTCGCA GCAACGGGTG ATCCGCCGGG	47340
CCCTGGCCGG CGCCGGCCTC GAACCGTCCG ATGTGGACAT CGTGAAGGG CACGGCACCG	47400
GGACGGCGCT GGGCGACCG ATCGAGGCG AGGCCCTGCT GGCCACCTAC GGCAAGGACC	47460
GCGACCCGGA GACGCCGTTG TGGCTGGGGT CGGTGAAGTC GAACTTCGGC CACACGCAGT	47520
CCGCGGCCGG CGTGGCCGGG GTGATCAAGA TGGTGCAGGC GCTGCGCCAC GCGGTCATGC	47580
CGCCACCCCT GCACGTGGAC CGGCCACCA GCCAGGTCGA CTGGTCCGCG GGGGCCGTG	47640
AAGTGCTGAC CGAGGCACGG GAGTGGCCG GGAACGGCCG TCCGCGCCG GCCGGGGTGT	47700
CCTCGTTCGG GATCAGCGGC ACGAACGCC ACCTGATCAT CGAAGAAGCA CCGGCCGAGC	47760
CACAGCTTGC CGGACCACCG CCGGACGGCG GTGTGGTGCC GCTGGTCTGTC TCGGCTCGCA	47820
GCCCCGGTGC CCTGGCCGGT CAGGCGCGTC GGCTGGCCAC GTTCCTCGGC GACGGGCCCC	47880
TPTCCGACGT CGCCGGTGCG CTGACGAGCC GCGCCCTGTT CGGCGAGCG GCGGTCGTG	47940
TGGCGGATTC GGCCGAGGAA GCCCGCGCCG GTCTGGGCGC ACTGGCCCGC GGCGAAGACG	48000
CGCCGGGCCT GGTCCGCGGC CGGGTGCCG CGTCCGGCCT GCCGGGCAAG CTCGTGTGGG	48060
TGTTCCCCCG GCAGGGGACG CAGTGGGTGG GCATGGGCCG CGAACTCCTC GAAGAGTCTC	48120

CGGTGTTCCG	CGAGCGGATC	GCCGAGTGTG	CGGCCGCGCT	GGAGCCGTGG	ATCGGCTGGT	48180
CGCTGTTTGA	CGTCCTCCGT	GGCGACGGTG	ACCTCGATCG	GGTCGATGTG	CTGCAGCCCCG	48240
CGTGCTTTGC	GGTGATGGTC	GGCTTGGCCG	CGGTGTGGTC	CTCGGCCGGG	GTGGTCCCCG	48300
ATGCGGTGCT	CGGCCACTCC	CAGGGTGAGA	TGCCCGCGGC	GTGCGTGTCG	GGTGCGTTGT	48360
CGCTGSAGGA	TGCGGCGAAG	GTGGTTGCCC	TGCCCAGCCA	GGCCATCGCC	GCGAAGCTCT	48420
CCGGCCGCGG	CGGGATGGCT	TCGGTCGCCT	TGGGCGAAGC	CGATGTGGTG	TCGCGGCTGG	48480
CGGACGGGGT	CGAGGTGGCT	GCCGTCAACG	GTCCGGCGTC	CGTGGTGATC	GCGGGGGATG	48540
CCCAGGCCCT	CGACGAAACG	CTGGAAGCGC	TGTCCGGTGC	GGGAATCCGG	GCTCGGCGGG	48600
TGGCGGTGGA	CTACGCCTCG	CACACCCGGC	ACGTGGAAGA	CATCGAAGAC	ACCCTCGCCG	48660
AAGCGCTGGC	CGGGATCGAC	GCCCCGGGCG	CGCTGGTGCC	GTTCTCTTCC	ACCCTCACCG	48720
GCGAGTGGAT	CCGGGACGAG	GGCGTCGTGG	ACGGCGGCTA	CTGGTACCGG	AACCTGCGCG	48780
GCCGGGTGCG	GTTCCGGCCG	GCCGTGAGG	CGCTGCTGGC	CCAGGGGCAC	GGTGTGTTCC	48840
TCGAGCTCAG	CGCCCACCCG	GTGCTGGTCC	AGCCGATCAC	CGAGCTCACC	GACGAAACCG	48900
CCGCCGTGCT	CACCGTTTCG	CTGCGCCGGG	ACGACGGTGG	CCTGCGCCGG	CTGCTGACCT	48960
CGATG3CCGA	GCTCTTCGTC	CGTGGGGTCC	AAGTGGAATG	GACGTGCTTG	GTGCCGCCCG	49020
CCCGGGCCGA	CCTCCCGACG	TACGCCTTCG	ACCACGAGCA	CTACTGGCTC	CGCGCCGCGG	49080
ACACCGCTTC	CGACGCCGTC	TCGCTGGGGC	TGGCCGGGGC	GGACCACCCG	CTGCTCGGCG	49140
CGGTGCTGCA	GCTTCCGCAG	TCCGACGGCC	TGGTCTTCAC	TTCCCGGCTC	TCCCTGCGCT	49200

CGCACCCCTG GCTGGCCGAC CACGCGGTCC GGGACGTCGT GATCGTCCCC GGCACCGGGC	49260
TGCTCGAGCT GGCCGTGCGG GCCGGTGACG AAGCCGGCTG CCCGGTGCTC GACGAGCTGG	49320
TGATCGAGGC GCCGCTCGTG GTGCCCCGCC GCGGCGGGGT CCGCGTGCAG GTCGCCCTCG	49380
GCGGCCCCGC CGACGACGGT TCGCGCACGG TGGACGTCTT CTCCCTGCGC GAAGACGCGG	49440
ACAGCTGGCT CCGGCACGCC ACGGGCGTGC TGGTCCCGGA GAACCGGCCG CGGGGGACCG	49500
CCGCGTTCGA CTTCGCCGCC TGGCCGCCAC CGGAGGCGAA GCCCGTGGAC CTCACCGGTG	49560
CCTACGACGT GCTCGCGGAC GTCGGGTACG GCTACGGGCC CACGTTCCGG GCCGTGCGGG	49620
CCGTGTGGCG GCGCGGCAGC GGGAAACACCA CCGAGACCTT CGCCGAGATC GCCCTGCCCG	49680
AAGACGCCCC CGCGGAAGCC GGCCGGTTCG GCATCCACCC CGCGCTGCTG GACGCGGCC	49740
TGCACTCGAC GATGGTCAGC GCCCGGCGCG ACACCGAGTC CTACGGCGAC GAAGTGCGGC	49800
TGCCGTTTCG GTGGAACGGG CTGCGGCTGC ACGCGGCCCG CGCCTCGGTG CTGCGGGTGC	49860
GCGTCGCCAA GCCCGAGCGG GACAGTCTGT CGCTGGAGGC CGTCGACGAG TCCGGCGGCC	49920
TGCTCGTGAC GCTGGATTCC CTGCTCGGGC GCCCGGTGTC GAACGACCAG CTGACGACGG	49980
CGCGGGGGCC GCGGGGCGCC GGCTCGCTGT ACCGCGTGGA CTGGACGCCA TTGTCTCAG	50040
TGGACACTTC GGGACGGGTG CCGTCTGGC TTCCGGTCGC CACCGCGGAA GAGGTGGCGA	50100
CGCTGGCCGA CGACGTCCTG ACCGGCGCGA CCGAGGCGCC GCGGTGGCC GTCATGGAGG	50160
CCGTCGCCGA CGAGGGTTCC GTGCTGGCGC TCACCGTCCG GGTGCTGGAC GTGGTCCAGT	50220
GCTGGCTGGC CGGCGGCGGG CTGGAGGGGA CGAAGCTCGC GATCGTGACC CGCGGCGCGG	50280
TGCCCCCGG CGACGGCGTG GTGCACGACC CGGCCGCGSC CGCGGTGTGG GGGCTGGTCC	50340

GGGCCGCGCA GCGGAGAAC CCGGACCGGA TCGTCCTCCT CGACGTCGAG CCGGAAGCCG	50400
ACGTACCGCC GCTGCTGGGT TCGGTGCTCG CCGACGGCGA GCCGCAGGTC GCGGTGCGCG	50460
GAACCACGCT GTCCATCCCC CGCCTCGCCC GCGCCGCCCG GCCCGACCCG GCCGCCGGGT	50520
TCAAGACCCG GGGACCGGTG CTGGTCACCG GCGGGACCGG GTCGCTCGGC GGCCTGGTCG	50580
CCCGGCACCT GGTGAGCGG CACGGCGTCC GGCAGCTGGT GCTGGCGAGT CGCCGGGGCC	50640
TGGACGCCGA AGGCGCGAAG GACCTGGTCA CCGACCTCAC CGCACTGGGG GCCGACGTCG	50700
CGGTGCCCCC TTGCGACGTC GCCGACCGGG ACCAGGTGGC GGCCCTGCTG ACCGAGCACC	50760
GGCCGTCCGC CGTGGTGCAC ACGGCCGGCG TCCCGGACGC CGGGGTGATC GGGACGGTGA	50820
CCCCGGACCG GCTGGCCGAG GTGTTCGCGC CCAAGGTCAC CGCGGCCCGG CACCTCGACG	50880
AGCTGACCCG CGACCTGGAC CTCGACAGTT TCGTCGTCTA CTCCTCGGTT TCCGCGGTGT	50940
TCATGGGGCG CCGCAGCGGC AGCTACGCCG CGGCGAACGC GTACCTGGAC GGGCTGATGG	51000
CCCACGGCG CCGGGCCGGC CTGCCGGGCC AGTCGCTGGC GTGGGGGCTG TGGGACCAGA	51060
CCACCGGCG CATGGCGGCC GGGACCGACG AGGCCGGCCG GGCCCGGATG ACCCGGCGCG	51120
GCGGCCTGGT CGCGATGAAA CCCGCCGCCG GACTGGACCT CTTGACGCT GCCATCGGGT	51180
CCGGCGAGCC GCTGCTGGTG CCCGCCAGC TCGACCTGCG GGGCCTGCGC GCCGAAGCGG	51240
CGGGCGGCAC CGAAGTGCCG CACCTGCTGC GCGGCCTGGT CCGCGCCGGA CGCCAGCAGG	51300
CCCGTGCGGC GTCCACTGTG GAGGAGAACT GGGCCGGCCG GCTGGCCGGG CTCGAGCCGG	51360
CCGAGCGGGG CCAGSTCCTC CTGGAAGTGG TGCCGCCCA GGTGGCAGGG GTCCTGGGCT	51420

ACCGCGCCGC CCACCAGGTC GACCCGGACC AGGGCCTGTT CGAGATCGGG TTCGACTCGC	51480
TCACCGCGAT CGAACTCCGC AACCGGCTGC GCGCCAGGAC CGAACGGAAG ATCTCGCCCG	51540
GTGTCGTCTT CGACCATCCC ACGCCGGCCC TGCTCGCCGC GCACTTGAAC GAGCTGCTCC	51600
GAAAGAAGGT GTGAACGTGT TCGACGTGGA GACCTACCTC CAGCGGATCG GCTGCGGCGG	51660
GGAAACCGGC GTGGACCTCG AAACGCTGGC GAAGCTGCAG AAGAGCCACC TGATGGCGAT	51720
CCCGTACAGC AGCCTCGCCT ACGAACTCCG GGACGCGGTG AACGTCGTCTG ACCTCGACGA	51780
GGACGACGTC TTCGTCACCA GCATCGCCGA AGGGCAGGGC GGCGCCTGCT ACCACCTGAA	51840
CCGGCTGTTC CACCGGCTCC TGACCGAACT CGGCTACGAC GTCACGCCGC TGGCCGGCAG	51900
CACCGCCGAA GGCCGGGAGA CCTTCGGCAC CGACGTCGAG CACATGTTCA ACCTGGTCAC	51960
CCTGGACGGC GCCGACTGGC TCGTGGACGT CGGCTACCCC GGCCCCACCT ACGTCGAGCC	52020
ACTGGCGGTC TCGCCCGCGG TGCAGACCCA GTACGGGAGC CAGTTCCGGT TGGTGGAAAC	52080
GGAAACCGGT TATGCGCTGC AACGCCGGGG TGCGGTCACC CGCTGGAGCG TCGTCTACAC	52140
GTTACAGACG CAACCGCGTC AGTGGAGTGA CTGGAAGGAA CTGGAGGACA ACTTCCGGGC	52200
CCTCGTGGGG GACACCACCC GCACCGACAC GCAGGAAACC CTGTGCGGCC GCGCGTTCCG	52260
GAACGGCCAG GTCTTCCTGC GGCAGCGCCG CTACCTGACG GTCGAGAACG GCCGCGAGCA	52320
GGTGCGCACG ATCACCGACG ACGACGAGTT CCGGGCGCTG GTGTCCCGCG TGCTGTCCGG	52380
CGACCACGGC TGAAGTGGCG AAAGGCACGA CGATGACGGA AAAAGCGGGC CTGCTGGCGA	52440
AGTTCGCCGG CCTCTGCAAA ACCGCCTACG AGCACCATA CATCCCGTAC CTGCACTTCT	52500
TCTACGGCGG CGAGTACCTC CACCACGGCA GCGAGCCGGT GTCCCGGATC GCGGACCTGC	52560

CGTACGTGAC CGTGCCGGAG CCGCGGAAGA AGGCGCCGTG AGGACGACGA TCCCGGTCCG	52620
CCTGGCGGAA CGGTCTACG ACGTGCTCGT CGGCCCCGGG GTGCGGGCGG CGCTGCCCCA	52680
GGTCGTCCGG CGGCTCGGCG CGAGACGGGC CGTGGTCGTG TCGGCCCCGC CGGCGGACTG	52740
GCTGCCCCGC ACCGGCGTCG AGACCCCTCT GCTCCAGGCG CGCGACGGCG AGCCGACCAA	52800
GCGGCTGTCC ACAGTGGAGG AACTGTGCGG TGAGTTCGCG CGGTCGGGC TCACCCGGTC	52860
CGACGTCTG GTCTCCTGCG GCGGCGGCAC GACCACGGAC GTCGTGCGGC TCGCGGCCGC	52920
GCTGTACCAC CGGGGGGTCT CCGTGGTCCA CCTGCCACG TCCCTGCTCG CCCAGGTCTA	52980
CGCCAGCGTC GGCAGGAAGA CCGCGGTGAA CCTGCCGGCG GGCAAGAACC TCGTCGGGGC	53040
GTACTGGCAG CCCAGCGCGG TGCTGTGCGA CACGGACTAC CTGACGACGC TGCCGCGGCG	53100
GGAGGTGCTG AACGGCCTCG GCGAGATCGC CCGCTGCCAC TTCATCGGCG CGCCGGACCT	53160
GCGGGGGCGC TCGCGCCCGG AGCAGATCGC CGCCAGCGTC ACCCTCAAGG CGGGCATCGT	53220
CGCGCAGGAC GAGCGGGACA CCGGCCCCG GCACCTGCTC AACTACGGCC ACACGCTGGG	53280
GCACGCGCTG GAGATCGCGA CCGGCTTCGC CCTGCGCCAC GCGAGGCGG TGGCGATCGG	53340
CACGGTCTTC GCGGGCCGGC TGGCCGGCGC GCTCGGGCGC CTCGACCAGT CCGGTGTGGA	53400
CGAACACCTC GCCGTCTGCC GCCACTACGG CCTGCCCGCC GCGCTGCCCG CGGACGTCTA	53460
CCCGGCGGTG CTCGTCCGCG AGATGTACCG GGACAAGAAG GCGATCACCG GGCTCGCCTT	53520
CGTCCTGGCC GGGCCGCGGG GCGCGGAGCT GGTGAGCGAC GTGCCGGCGC CGGTCTGTCAC	53580
CGACGTCTTG GACCGGATGC CCCGCGACAG CCTGGAAAAC CTGGTGGGGA CGACGGAAGC	53640

GGCGGCGCCG TGAAGCGGCA GCCGGACTTC GCGGCCACG GCCGGGCGGT CGACCGGGTG 53700

CTGGCCGGCC GGCTGAGCGC GCGCTGGCC CGGCCGGCCG CGCAGCAGCC GGGCTGGCCG 53760

GACGCCGAGC GGGCGGCCGA GGTGAATTC 53789

(2) Information für SEQ ID Nr: 4

(i) Sequenzeigenschaften:

(A) Länge: 4572 Aminosäuren
 (B) Typ: Aminosäure
 (C) Strangart: Einzelstrang
 (D) Topologie: linear

(ii) Molekültyp: Peptid

(xi) Sequenzbeschreibung: SEQ ID Nr: 4

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

Thr Gln Arg Asn Cys Leu Trp Ser Val Ala Ser Cys Tyr Val Pro Phe
 20 25 30

Pro Gly Leu Ser Asp Gln Asp Arg Val Leu Trp Pro Leu Pro Leu Phe
 35 40 45

His Ser Leu Ser His Ile Ala Cys Val Leu Ser Ala Thr Val Val Gly
 50 55 60

Ala Ser Val Arg Ile Ala Asp Gly Ser Ser Ala Asp Asp Val Met Arg
 65 70 75 80

Leu Ile Glu Ala Glu Ser Ser Thr Phe Leu Ala Gly Val Pro Thr Thr
 85 90 95

Tyr His His Leu Val Arg Ala Ala Arg Gln Arg Gly Phe Ser Ala Pro
 100 105 110

Ser Leu Arg Ile Gly Leu Ala Gly Gly Ala Val Leu Gly Ala Gly Leu
 115 120 125

Arg Ser Glu Phe Glu Glu Thr Phe Gly Val Pro Leu Ile Asp Ala Tyr
 130 135 140

Gly Ser Thr Glu Thr Cys Gly Ala Ile Thr Met Asn Pro Pro Asp Gly
 145 150 155 160

Ala Arg Val Glu Gly Ser Cys Gly Leu Ala Val Pro Gly Val Asp Val
 165 170 175

Arg Val Val Asp Pro Asp Thr Gly Leu Asp Val Pro Ala Gly Glu Glu
 180 185 190

Gly Glu Val Trp Val Ser Gly Pro Asn Val Met Leu Gly Tyr His Asn
 195 200 205

Ser Pro Glu Ala Thr Ala Ala Ala Met Arg Asp Gly Trp Phe Arg Thr
 210 215 220

Gly Asp Leu Ala Arg Arg Asp Asp Ala Gly Tyr Phe Thr Ile Cys Gly
 225 230 235 240

Arg Ile Lys Glu Leu Ile Ile Arg Gly Gly Ala Asn Ile His Pro Gly
 245 250 255

Glu Val Glu Ala Val Leu Arg Thr Val Asp Gly Val Ala Asp Ala Ala
 260 265 270

Val Gly Gly Val Pro His Asp Thr Leu Gly Glu Val Pro Val Ala Tyr			
275	280	285	
Val Ile Pro Gly Pro Thr Gly Phe Asp Pro Ala Ala Leu Ile Glu Lys			
290	295	300	
Cys Arg Glu Gln Leu Ser Ala Tyr Lys Val Pro Asp Arg Ile Leu Glu			
305	310	315	320
Val Ala His Ile Pro Arg Thr Ala Ser Gly Lys Ile Arg Arg Gly Leu			
	325	330	335
Leu Thr Asp Glu Pro Ala Gln Leu Arg Tyr Ala Ala Thr Glu His Glu			
	340	345	350
Glu Gln Ser Arg His Ala Asp Glu Ser Val Ala Ala Ala Leu Arg Ala			
	355	360	365
Arg Leu Ser Gly Leu Asp Glu Arg Ala Gln Cys Glu Leu Leu Glu Asp			
	370	375	380
Leu Val Arg Thr Gln Ala Ala Asp Val Leu Gly Gln Pro Val Pro Asp			
385	390	395	400
Gly Arg Ala Phe Arg Asp Leu Gly Phe Thr Ser Leu Ala Ile Val Glu			
	405	410	415
Leu Arg Asn Arg Leu Thr Glu His Thr Gly Leu Trp Leu Pro Ala Ser			
	420	425	430
Ala Val Phe Asp His Pro Thr Pro Ala Ala Leu Ala Ala Arg Val Arg			
	435	440	445
Ala Glu Leu Leu Gly Ile Thr Gln Ala Val Ala Glu Pro Val Val Ala			
	450	455	460
Ala Asp Pro Gly Glu Pro Ile Ala Ile Val Gly Met Ala Cys Arg Leu			

465	470	475	480
Pro Gly Gly Val Ala Ser Pro Glu Asp Leu Trp Arg Leu Val Ala Glu			
	485	490	495
Arg Val Asp Ala Val Ser Glu Phe Pro Gly Asp Arg Gly Trp Asp Leu			
	500	505	510
Asp Ser Leu Ile Asp Pro Asp Arg Glu Arg Ala Gly Thr Ser Tyr Val			
	515	520	525
Gly Gln Gly Gly Phe Leu His Asp Ala Gly Glu Phe Asp Ala Gly Phe			
	530	535	540
Phe Gly Ile Ser Pro Arg Glu Ala Val Ala Met Asp Pro Gln Gln Arg			
545	550	555	560
Leu Leu Leu Glu Thr Ser Trp Glu Ala Leu Glu Asn Ala Gly Val Asp			
	565	570	575
Pro Ile Ala Leu Lys Gly Thr Asp Thr Gly Val Phe Ser Gly Leu Met			
	580	585	590
Gly Gln Gly Tyr Gly Ser Gly Ala Val Ala Pro Glu Leu Glu Gly Phe			
	595	600	605
Val Thr Thr Gly Val Ala Ser Ser Val Ala Ser Gly Arg Val Ser Tyr			
	610	615	620
Val Leu Gly Leu Glu Gly Pro Ala Val Thr Val Asp Thr Ala Cys Ser			
625	630	635	640
Ser Ser Leu Val Ala Met His Leu Ala Ala Gln Ala Leu Arg Gln Gly			
	645	650	655
Glu Cys Ser Met Ala Leu Ala Gly Gly Val Thr Val Met Ala Thr Pro			
	660	665	670

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

Arg Asn Gly Arg Pro Arg Arg Ala Gly Val Ser Ser Phe Gly Val Ser
 865 870 875 880

Gly Thr Asn Ala His Leu Ile Leu Glu Glu Ala Pro Ala Glu Glu Pro
 885 890 895

Val Ala Ala Pro Glu Leu Pro Val Val Pro Leu Val Val Ser Ala Arg
 900 905 910

Ser Thr Glu Ser Leu Ser Gly Gln Ala Glu Arg Leu Ala Ser Leu Leu
 915 920 925

Glu Gly Asp Val Ser Leu Thr Glu Val Ala Gly Ala Leu Val Ser Arg
 930 935 940

Arg Ala Val Leu Asp Glu Arg Ala Val Val Val Ala Gly Ser Arg Glu
 945 950 955 960

Glu Ala Val Thr Gly Leu Arg Ala Leu Asn Thr Ala Gly Ser Gly Thr
 965 970 975

Pro Gly Lys Val Val Trp Val Phe Pro Gly Gln Gly Thr Gln Trp Ala
 980 985 990

Gly Met Gly Arg Glu Leu Leu Ala Glu Ser Pro Val Phe Ala Glu Arg
 995 1000 1005

Ile Ala Glu Cys Ala Ala Ala Leu Ala Pro Trp Ile Asp Trp Ser Leu
 1010 1015 1020

Val Asp Val Leu Arg Gly Glu Gly Asp Leu Gly Arg Val Asp Val Leu
 1025 1030 1035 1040

Gln Pro Ala Cys Phe Ala Val Met Val Gly Leu Ala Ala Val Trp Glu
 1045 1050 1055

Ser Val Gly Val Arg Pro Asp Ala Val Val Gly His Ser Gln Gly Glu

1060	1065	1070
Ile Ala Ala Ala Cys Val Ser Gly Ala Leu Ser Leu Glu Asp Ala Ala		
1075	1080	1085
Lys Val Val Ala Leu Arg Ser Gln Ala Ile Ala Ala Glu Leu Ser Gly		
1090	1095	1100
Arg Gly Gly Met Ala Ser Val Ala Leu Gly Glu Asp Asp Val Val Ser		
1105	1110	1115 1120
Arg Leu Val Asp Gly Val Glu Val Ala Ala Val Asn Gly Pro Ser Ser		
1125	1130	1135
Val Val Ile Ala Gly Asp Ala His Ala Leu Asp Ala Thr Leu Glu Ile		
1140	1145	1150
Leu Ser Gly Glu Gly Ile Arg Val Arg Arg Val Ala Val Asp Tyr Ala		
1155	1160	1165
Ser His Thr Arg His Val Glu Asp Ile Arg Asp Thr Leu Ala Glu Thr		
1170	1175	1180
Leu Ala Gly Ile Ser Ala Gln Ala Pro Ala Val Pro Phe Tyr Ser Thr		
1185	1190	1195 1200
Val Thr Ser Glu Trp Val Arg Asp Ala Gly Val Leu Asp Gly Gly Tyr		
1205	1210	1215
Trp Tyr Arg Asn Leu Arg Asn Gln Val Arg Phe Gly Ala Ala Ala Thr		
1220	1225	1230
Ala Leu Leu Glu Gln Gly His Thr Val Phe Val Glu Val Ser Ala His		
1235	1240	1245
Pro Val Thr Val Gln Pro Leu Ser Glu Leu Thr Gly Asp Ala Ile Gly		
1250	1255	1260

Thr Leu Arg Arg Glu Asp Gly Gly Leu Arg Arg Leu Leu Ala Ser Met
 1265 1270 1275 1280

Gly Glu Leu Phe Val Arg Gly Ile Asp Val Asp Trp Thr Ala Met Val
 1285 1290 1295

Pro Ala Ala Gly Trp Val Asp Leu Pro Thr Tyr Ala Phe Glu His Arg
 1300 1305 1310

His Tyr Trp Leu Glu Pro Ala Glu Pro Ala Ser Ala Gly Asp Pro Leu
 1315 1320 1325

Leu Gly Thr Val Val Ser Thr Pro Gly Ser Asp Arg Leu Thr Ala Val
 1330 1335 1340

Ala Gln Trp Ser Arg Arg Ala Gln Pro Trp Ala Val Asp Gly Leu Val
 1345 1350 1355 1360

Pro Asn Ala Ala Leu Val Glu Ala Ala Ile Arg Leu Gly Asp Leu Ala
 1365 1370 1375

Gly Thr Pro Val Val Gly Glu Leu Val Val Asp Ala Pro Val Val Leu
 1380 1385 1390

Pro Arg Arg Gly Ser Arg Glu Val Gln Leu Ile Val Gly Glu Pro Gly
 1395 1400 1405

Glu Gln Arg Arg Arg Pro Ile Glu Val Phe Ser Arg Glu Ala Asp Glu
 1410 1415 1420

Pro Trp Thr Arg His Ala His Gly Thr Leu Ala Pro Ala Ala Ala Ala
 1425 1430 1435 1440

Val Pro Glu Pro Ala Ala Ala Gly Asp Ala Thr Asp Val Thr Val Ala
 1445 1450 1455

Gly Leu Arg Asp Ala Asp Arg Tyr Gly Ile His Pro Ala Leu Leu Asp			
1460	1465	1470	
Ala Ala Val Arg Thr Val Val Gly Asp Asp Leu Leu Pro Ser Val Trp			
1475	1480	1485	
Thr Gly Val Ser Leu Leu Ala Ser Gly Ala Thr Ala Val Thr Val Thr			
1490	1495	1500	
Pro Thr Ala Thr Gly Leu Arg Leu Thr Asp Pro Ala Gly Gln Pro Val			
1505	1510	1515	1520
Leu Thr Val Glu Ser Val Arg Gly Thr Pro Phe Val Ala Glu Gln Gly			
1525	1530	1535	
Thr Thr Asp Ala Leu Phe Arg Val Asp Trp Pro Glu Ile Pro Leu Pro			
1540	1545	1550	
Thr Ala Glu Thr Ala Asp Phe Leu Pro Tyr Glu Ala Thr Ser Ala Glu			
1555	1560	1565	
Ala Thr Leu Ser Ala Leu Gln Ala Trp Leu Ala Asp Pro Ala Glu Thr			
1570	1575	1580	
Arg Leu Ala Val Val Thr Gly Asp Cys Thr Glu Pro Gly Ala Ala Ala			
1585	1590	1595	1600
Ile Trp Gly Leu Val Arg Ser Ala Gln Ser Glu His Pro Gly Arg Ile			
1605	1610	1615	
Val Leu Ala Asp Leu Asp Asp Pro Ala Val Leu Pro Ala Val Val Ala			
1620	1625	1630	
Ser Gly Glu Pro Gln Val Arg Val Arg Asn Gly Val Ala Ser Val Pro			
1635	1640	1645	
Arg Leu Thr Arg Val Thr Pro Arg Gln Asp Ala Arg Pro Leu Asp Pro			

1650

1655

1660

Glu Gly Thr Val Leu Ile Thr Gly Gly Thr Gly Thr Leu Gly Ala Leu
 1665 1670 1675 1680

Thr Ala Arg His Leu Val Thr Ala His Gly Val Arg His Leu Val Leu
 1685 1690 1695

Val Ser Arg Arg Gly Glu Ala Pro Glu Leu Gln Glu Glu Leu Thr Ala
 1700 1705 1710

Leu Gly Ala Ser Val Ala Ile Ala Ala Cys Asp Val Ala Asp Arg Ala
 1715 1720 1725

Gln Leu Glu Ala Val Leu Arg Ala Ile Pro Ala Glu His Pro Leu Thr
 1730 1735 1740

Ala Val Ile His Thr Ala Gly Val Leu Asp Asp Gly Val Val Thr Glu
 1745 1750 1755 1760

Leu Thr Pro Asp Arg Leu Ala Thr Val Arg Arg Pro Lys Val Asp Ala
 1765 1770 1775

Ala Arg Leu Leu Asp Glu Leu Thr Arg Glu Ala Asp Leu Ala Ala Phe
 1780 1785 1790

Val Leu Phe Ser Ser Ala Ala Gly Val Leu Gly Asn Pro Gly Gln Ala
 1795 1800 1805

Gly Tyr Ala Ala Ala Asn Ala Glu Leu Asp Ala Leu Ala Arg Gln Arg
 1810 1815 1820

Asn Ser Leu Asp Leu Pro Ala Val Ser Ile Ala Trp Gly Tyr Trp Ala
 1825 1830 1835 1840

Thr Val Ser Gly Met Thr Glu His Leu Gly Asp Ala Asp Leu Arg Arg
 1845 1850 1855

Asn Gln Arg Ile Gly Met Ser Gly Leu Pro Ala Asp Glu Gly Met Ala
 1860 1865 1870

Leu Leu Asp Ala Ala Ile Ala Thr Gly Gly Thr Leu Val Ala Ala Lys
 1875 1880 1885

Phe Asp Val Ala Ala Leu Arg Ala Thr Ala Lys Ala Gly Gly Pro Val
 1890 1895 1900

Pro Pro Leu Leu Arg Gly Leu Ala Pro Leu Pro Arg Arg Ala Ala Ala
 1905 1910 1915 1920

Lys Thr Ala Ser Leu Thr Glu Arg Leu Ala Gly Leu Ala Glu Thr Glu
 1925 1930 1935

Gln Ala Ala Ala Leu Leu Asp Leu Val Arg Arg His Ala Ala Glu Val
 1940 1945 1950

Leu Gly His Ser Gly Ala Glu Ser Val His Ser Gly Arg Thr Phe Lys
 1955 1960 1965

Asp Ala Gly Phe Asp Ser Leu Thr Ala Val Glu Leu Arg Asn Arg Leu
 1970 1975 1980

Ala Ala Ala Thr Gly Leu Thr Leu Ser Pro Ala Met Ile Phe Asp Tyr
 1985 1990 1995 2000

Pro Lys Pro Pro Ala Leu Ala Asp His Leu Arg Ala Lys Leu Phe Gly
 2005 2010 2015

Ser Ala Ala Asn Arg Pro Ala Glu Ile Gly Thr Ala Ala Ala Glu Glu
 2020 2025 2030

Pro Ile Ala Ile Val Ala Met Ala Cys Arg Phe Pro Gly Gly Val His
 2035 2040 2045

Ser Pro Glu Asp Leu Trp Arg Leu Val Ala Asp Gly Ala Asp Ala Val			
2050	2055	2060	
Thr Glu Phe Pro Ala Asp Arg Gly Trp Asp Thr Asp Arg Leu Tyr His			
2065	2070	2075	2080
Glu Asp Pro Asp His Glu Gly Thr Thr Tyr Val Arg His Gly Ala Phe			
	2085	2090	2095
Leu Asp Asp Ala Ala Gly Phe Asp Ala Ala Phe Phe Gly Ile Ser Pro			
	2100	2105	2110
Asn Glu Ala Leu Ala Met Asp Pro Gln Gln Arg Leu Leu Leu Glu Thr			
	2115	2120	2125
Ser Trp Glu Leu Phe Glu Arg Ala Ala Ile Asp Pro Thr Thr Leu Ala			
	2130	2135	2140
Gly Gln Asp Ile Gly Val Phe Ala Gly Val Asn Ser His Asp Tyr Ser			
2145	2150	2155	2160
Met Arg Met His Arg Ala Ala Gly Val Glu Gly Phe Arg Leu Thr Gly			
	2165	2170	2175
Gly Ser Ala Ser Val Leu Ser Gly Arg Val Ala Tyr His Phe Gly Val			
	2180	2185	2190
Glu Gly Pro Ala Val Thr Val Asp Thr Ala Cys Ser Ser Ser Leu Val			
	2195	2200	2205
Ala Leu His Met Ala Val Gln Ala Leu Gln Arg Gly Glu Cys Ser Met			
	2210	2215	2220
Ala Leu Ala Gly Gly Val Met Val Met Gly Thr Val Glu Thr Phe Val			
2225	2230	2235	2240
Glu Phe Ser Arg Gln Arg Gly Leu Ala Pro Asp Gly Arg Cys Lys Ala			

2245	2250	2255
Phe Ala Asp Gly Ala Asp Gly Thr Gly Trp Ser Glu Gly Val Gly Leu		
2260	2265	2270
Leu Leu Val Glu Arg Leu Ser Glu Ala Gln Arg Arg Gly His Gln Val		
2275	2280	2285
Leu Ala Val Val Arg Gly Ser Ala Val Asn Ser Asp Gly Ala Ser Asn		
2290	2295	2300
Gly Leu Thr Ala Pro Asn Gly Pro Ser Gln Gln Arg Val Ile Arg Lys		
2305	2310	2315
		2320
Ala Leu Ala Ala Ala Gly Leu Ser Thr Ser Asp Val Asp Ala Val Glu		
2325	2330	2335
Ala His Gly Thr Gly Thr Thr Leu Gly Asp Pro Ile Glu Ala Glu Ala		
2340	2345	2350
Leu Leu Ala Thr Tyr Gly Gln Asn Arg Glu Thr Pro Leu Trp Leu Gly		
2355	2360	2365
Ser Val Lys Ser Asn Leu Gly His Thr Gln Ala Ala Ala Gly Val Ala		
2370	2375	2380
Gly Val Ile Lys Met Val Met Ala Met Arg His Gly Val Leu Pro Arg		
2385	2390	2395
		2400
Thr Leu His Val Asp Arg Pro Ser Ser Tyr Val Asp Trp Ser Ala Gly		
2405	2410	2415
Ala Val Glu Leu Leu Thr Glu Ala Arg Asp Trp Val Ser Asn Gly His		
2420	2425	2430
Pro Arg Arg Ala Gly Val Ser Ser Phe Gly Ile Gly Gly Thr Asn Ala		
2435	2440	2445

His Val Val Leu Glu Glu Val Ala Ala Pro Ile Thr Thr Pro Gln Pro
 2450 2455 2460

Glu Pro Ala Glu Phe Leu Val Pro Val Leu Val Ser Ala Arg Thr Ala
 2465 2470 2475 2480

Ala Gly Leu Arg Gly Gln Ala Gly Arg Leu Ala Ala Phe Leu Gly Asp
 2485 2490 2495

Arg Thr Asp Val Arg Val Pro Asp Ala Ala Tyr Ala Leu Ala Thr Thr
 2500 2505 2510

Arg Ala Gln Leu Asp His Arg Ala Val Val Leu Ala Ser Asp Arg Ala
 2515 2520 2525

Gln Leu Cys Ala Asp Leu Ala Ala Phe Gly Ser Gly Val Val Thr Gly
 2530 2535 2540

Thr Pro Val Asp Gly Lys Leu Ala Val Leu Phe Thr Gly Gln Gly Ser
 2545 2550 2555 2560

Gln Trp Ala Gly Met Gly Arg Glu Leu Ala Glu Thr Phe Pro Val Phe
 2565 2570 2575

Arg Asp Ala Phe Glu Ala Ala Cys Glu Ala Val Asp Thr His Leu Arg
 2580 2585 2590

Glu Arg Pro Leu Arg Glu Val Val Phe Asp Asp Ser Ala Leu Leu Asp
 2595 2600 2605

Gln Thr Met Tyr Thr Gln Gly Ala Leu Phe Ala Val Glu Thr Ala Leu
 2610 2615 2620

Phe Arg Leu Phe Glu Ser Trp Gly Val Arg Pro Gly Leu Leu Ala Gly
 2625 2630 2635 2640

His Val Val Leu Glu Glu Val Ala Ala Pro Ile Thr Thr Pro Gln Pro			
2450	2455	2460	
Glu Pro Ala Glu Phe Leu Val Pro Val Leu Val Ser Ala Arg Thr Ala			
2465	2470	2475	2480
Ala Gly Leu Arg Gly Gln Ala Gly Arg Leu Ala Ala Phe Leu Gly Asp			
2485	2490	2495	
Arg Thr Asp Val Arg Val Pro Asp Ala Ala Tyr Ala Leu Ala Thr Thr			
2500	2505	2510	
Arg Ala Gln Leu Asp His Arg Ala Val Val Leu Ala Ser Asp Arg Ala			
2515	2520	2525	
Gln Leu Cys Ala Asp Leu Ala Ala Phe Gly Ser Gly Val Val Thr Gly			
2530	2535	2540	
Thr Pro Val Asp Gly Lys Leu Ala Val Leu Phe Thr Gly Gln Gly Ser			
2545	2550	2555	2560
Gln Trp Ala Gly Met Gly Arg Glu Leu Ala Glu Thr Phe Pro Val Phe			
2565	2570	2575	
Arg Asp Ala Phe Glu Ala Ala Cys Glu Ala Val Asp Thr His Leu Arg			
2580	2585	2590	
Glu Arg Pro Leu Arg Glu Val Val Phe Asp Asp Ser Ala Leu Leu Asp			
2595	2600	2605	
Gln Thr Met Tyr Thr Gln Gly Ala Leu Phe Ala Val Glu Thr Ala Leu			
2610	2615	2620	
Phe Arg Leu Phe Glu Ser Trp Gly Val Arg Pro Gly Leu Leu Ala Gly			
2625	2630	2635	2640

His Ser Ile Gly Glu Leu Ala Ala Ala His Val Ser Gly Val Leu Asp			
2645	2650	2655	
Leu Ala Asp Ala Gly Glu Leu Val Ala Ala Arg Gly Arg Leu Met Gln			
2660	2665	2670	
Ala Leu Pro Ala Gly Gly Ala Met Val Ala Val Gln Ala Thr Glu Asp			
2675	2680	2685	
Glu Val Ala Pro Leu Leu Asp Gly Thr Val Cys Val Ala Ala Val Asn			
2690	2695	2700	
Gly Pro Asp Ser Val Val Leu Ser Gly Thr Glu Ala Ala Val Leu Ala			
2705	2710	2715	2720
Val Ala Asp Glu Leu Ala Gly Arg Gly Arg Lys Thr Arg Arg Leu Ala			
2725	2730	2735	
Val Ser His Ala Phe His Ser Pro Leu Met Glu Pro Met Leu Asp Asp			
2740	2745	2750	
Phe Arg Ala Val Ala Glu Arg Leu Thr Tyr Arg Ala Gly Ser Leu Pro			
2755	2760	2765	
Val Val Ser Thr Leu Thr Gly Glu Leu Ala Ala Leu Asp Ser Pro Asp			
2770	2775	2780	
Tyr Trp Val Gly Gln Val Arg Asn Ala Val Arg Phe Ser Asp Ala Val			
2785	2790	2795	2800
Thr Ala Leu Gly Ala Gln Gly Ala Ser Thr Phe Leu Glu Leu Gly Pro			
2805	2810	2815	
Gly Gly Ala Leu Ala Ala Met Ala Leu Gly Thr Leu Gly Gly Pro Glu			
2820	2825	2830	
Gln Ser Cys Val Ala Thr Leu Arg Lys Asn Gly Ala Glu Val Pro Asp			

2835

2840

2845

Val Leu Thr Ala Leu Ala Glu Leu His Val Arg Gly Val Gly Val Asp
 2850 2855 2860

Trp Thr Thr Val Leu Asp Glu Pro Ala Thr Ala Val Gly Thr Val Leu
 2865 2870 2875 2880

Pro Thr Tyr Ala Phe Gln His Gln Arg Phe Trp Val Asp Val Asp Glu
 2885 2890 2895

Thr Ala Ala Val Ser Val Thr Pro Pro Pro Ala Glu Pro Ile Val Asp
 2900 2905 2910

Arg Pro Val Gln Asp Val Leu Glu Leu Val Arg Glu Ser Ala Ala Val
 2915 2920 2925

Val Leu Gly His Arg Asp Ala Gly Ser Phe Asp Leu Asp Arg Ser Phe
 2930 2935 2940

Lys Asp His Gly Phe Asp Ser Leu Ser Ala Val Lys Leu Arg Asn Arg
 2945 2950 2955 2960

Leu Arg Asp Phe Thr Gly Val Glu Leu Pro Ser Thr Leu Ile Phe Asp
 2965 2970 2975

Tyr Pro Asn Pro Ala Val Leu Ala Asp His Leu Arg Ala Glu Leu Leu
 2980 2985 2990

Gly Glu Arg Pro Ala Ala Pro Ala Pro Val Thr Arg Asp Val Ser Asp
 2995 3000 3005

Glu Pro Ile Ala Ile Val Gly Met Ser Thr Arg Leu Pro Gly Gly Ala
 3010 3015 3020

Asp Ser Pro Glu Glu Leu Trp Lys Leu Val Ala Glu Gly Arg Asp Ala
 3025 3030 3035 3040

Val Ser Gly Phe Pro Val Asp Arg Gly Trp Asp Leu Asp Gly Leu Tyr
 3045 3050 3055

His Pro Asp Pro Ala His Ala Gly Thr Ser Tyr Thr Arg Ser Gly Gly
 3060 3065 3070

Phe Leu His Asp Ala Ala Gln Phe Asp Ala Gly Leu Phe Gly Ile Ser
 3075 3080 3085

Pro Arg Glu Ala Leu Ala Met Asp Pro Gln Gln Arg Leu Leu Leu Glu
 3090 3095 3100

Thr Ser Trp Glu Ala Leu Glu Arg Ala Gly Val Asp Pro Leu Ser Ala
 3105 3110 3115 3120

Arg Gly Ser Asp Val Gly Val Phe Thr Gly Ile Val His His Asp Tyr
 3125 3130 3135

Val Thr Arg Leu Arg Glu Val Pro Glu Asp Val Gln Gly Tyr Thr Met
 3140 3145 3150

Thr Gly Thr Ala Ser Ser Val Ala Ser Gly Arg Val Ala Tyr Val Phe
 3155 3160 3165

Gly Phe Glu Gly Pro Ala Val Thr Val Asp Thr Ala Cys Ser Ser Ser
 3170 3175 3180

Leu Val Ala Met His Leu Ala Ala Gln Ala Leu Arg Gln Gly Glu Cys
 3185 3190 3195 3200

Ser Met Ala Leu Ala Gly Gly Ala Thr Val Met Ala Ser Pro Asp Ala
 3205 3210 3215

Phe Leu Glu Phe Ser Arg Gln Arg Gly Leu Ser Ala Asp Gly Arg Cys
 3220 3225 3230

Lys Ala Tyr Ala Glu Gly Ala Asp Gly Thr Gly Trp Ala Glu Gly Val			
3235	3240	3245	
Gly Val Val Val Leu Glu Arg Leu Ser Val Ala Arg Glu Arg Gly His			
3250	3255	3260	
Arg Val Leu Ala Val Leu Arg Gly Ser Ala Val Asn Gln Asp Gly Ala			
3265	3270	3275	3280
Ser Asn Gly Leu Thr Ala Pro Asn Gly Pro Ser Gln Gln Arg Val Ile			
3285	3290	3295	
Arg Gly Ala Leu Ala Ser Ala Gly Leu Ala Pro Ser Asp Val Asp Val			
3300	3305	3310	
Val Glu Gly His Gly Thr Gly Thr Ala Leu Gly Asp Pro Ile Glu Val			
3315	3320	3325	
Gln Ala Leu Leu Ala Thr Tyr Gly Gln Glu Arg Glu Gln Pro Leu Trp			
3330	3335	3340	
Leu Gly Ser Leu Lys Ser Asn Leu Gly His Thr Gln Ala Ala Ala Gly			
3345	3350	3355	3360
Val Val Gly Val Ile Lys Met Ile Met Ala Met Arg His Gly Val Met			
3365	3370	3375	
Pro Ala Thr Leu His Val Asp Glu Arg Thr Ser Gln Val Asp Trp Ser			
3380	3385	3390	
Ala Gly Ala Ile Glu Val Leu Thr Glu Ala Arg Glu Trp Pro Arg Thr			
3395	3400	3405	
Gly Arg Pro Arg Arg Ala Gly Val Ser Ser Phe Gly Ala Ser Gly Thr			
3410	3415	3420	
Asn Ala His Leu Ile Ile Glu Glu Gly Pro Ala Glu Glu Ala Val Asp			

3425	3430	3435	3440
Glu Glu Val Ala Ser Val Val Pro Leu Val Val Ser Ala Arg Ser Ala			
3445	3450	3455	
Gly Ser Leu Ala Gly Gln Ala Gly Arg Leu Ala Ala Val Leu Glu Asn			
3460	3465	3470	
Glu Ser Leu Ala Gly Val Ala Gly Ala Leu Val Ser Gly Arg Ala Thr			
3475	3480	3485	
Leu Asn Glu Arg Ala Val Val Ile Ala Gly Ser Arg Asp Glu Ala Gln			
3490	3495	3500	
Asp Gly Leu Gln Ala Leu Ala Arg Gly Glu Asn Ala Pro Gly Val Val			
3505	3510	3515	3520
Thr Gly Thr Ala Gly Lys Pro Gly Lys Val Val Trp Val Phe Pro Gly			
3525	3530	3535	
Gln Gly Ser Gln Trp Met Gly Met Gly Arg Asp Leu Leu Asp Ser Ser			
3540	3545	3550	
Pro Val Phe Ala Ala Arg Ile Lys Glu Cys Ala Ala Ala Leu Glu Gln			
3555	3560	3565	
Trp Thr Asp Trp Ser Leu Leu Asp Val Leu Arg Gly Asp Ala Asp Leu			
3570	3575	3580	
Leu Asp Arg Val Asp Val Val Gln Pro Ala Ser Phe Ala Met Met Val			
3585	3590	3595	3600
Gly Leu Ala Ala Val Trp Thr Ser Leu Gly Val Thr Pro Asp Ala Val			
3605	3610	3615	
Leu Gly His Ser Gln Gly Glu Ile Ala Ala Ala Cys Val Ser Gly Ala			
3620	3625	3630	

Leu Ser Leu Asp Asp Ala Ala Lys Val Val Ala Leu Arg Ser Gln Ala			
3635	3640	3645	
Ile Ala Gly Glu Leu Ala Gly Arg Gly Gly Met Ala Ser Val Ala Leu			
3650	3655	3660	
Ser Glu Glu Asp Ala Val Ala Arg Leu Thr Pro Trp Ala Asn Arg Val			
3665	3670	3675	3680
Glu Val Ala Ala Val Asn Ser Pro Ser Ser Val Val Ile Ala Gly Asp			
3685	3690	3695	
Ala Gln Ala Leu Asp Glu Ala Leu Glu Ala Leu Ala Gly Asp Gly Val			
3700	3705	3710	
Arg Val Arg Arg Val Ala Val Asp Tyr Ala Ser His Thr Arg His Val			
3715	3720	3725	
Glu Ala Ile Ala Glu Thr Leu Ala Lys Thr Leu Ala Gly Ile Asp Ala			
3730	3735	3740	
Arg Val Pro Ala Ile Pro Phe Tyr Ser Thr Val Leu Gly Thr Trp Ile			
3745	3750	3755	3760
Glu Gln Ala Val Val Asp Ala Gly Tyr Trp Tyr Arg Asn Leu Arg Gln			
3765	3770	3775	
Gln Val Arg Phe Gly Pro Ser Val Ala Asp Leu Ala Gly Leu Gly His			
3780	3785	3790	
Thr Val Phe Val Glu Ile Ser Ala His Pro Val Leu Val Gln Pro Leu			
3795	3800	3805	
Ser Glu Ile Ser Asp Asp Ala Val Val Thr Gly Ser Leu Arg Arg Asp			
3810	3815	3820	

Asp Gly Gly Leu Arg Arg Leu Leu Ala Ser Ala Ala Glu Leu Tyr Val
 3825 3830 3835 3840

Arg Gly Val Ala Val Asp Trp Thr Ala Ala Val Pro Ala Ala Gly Trp
 3845 3850 3855

Val Asp Leu Pro Thr Tyr Ala Phe Asp Arg Arg His Phe Trp Leu His
 3860 3865 3870

Glu Ala Glu Thr Ala Glu Ala Ala Glu Gly Met Asp Gly Glu Phe Trp
 3875 3880 3885

Thr Ala Ile Glu Gln Ser Asp Val Asp Ser Leu Ala Glu Leu Leu Glu
 3890 3895 3900

Leu Val Pro Glu Gln Arg Gly Ala Leu Ser Thr Val Val Pro Val Leu
 3905 3910 3915 3920

Ala Gln Trp Arg Asp Arg Arg Arg Glu Arg Ser Thr Ala Glu Lys Leu
 3925 3930 3935

Arg Tyr Gln Val Thr Trp Gln Pro Leu Glu Arg Glu Ala Ala Gly Val
 3940 3945 3950

Pro Gly Gly Arg Trp Leu Ala Val Val Pro Ala Gly Thr Thr Asp Ala
 3955 3960 3965

Leu Leu Lys Glu Leu Thr Gly Gln Gly Leu Asp Ile Val Arg Leu Glu
 3970 3975 3980

Ile Glu Glu Ala Ser Arg Ala Gln Leu Ala Glu Gln Leu Arg Asn Val
 3985 3990 3995 4000

Leu Ala Glu His Asp Leu Thr Gly Val Leu Ser Leu Leu Ala Leu Asp
 4005 4010 4015

Gly Gly Pro Ala Asp Ala Ala Glu Ile Thr Ala Ser Thr Leu Ala Leu

Thr Ser Ser Val Asp Asp Thr Gly Val Ala Asp Leu Asp Glu Val Phe
 4225 4230 4235 4240

Ala Ala Lys Val Thr Thr Ala Val Trp Leu Asp Gln Leu Phe Glu Asp
 4245 4250 4255

Thr Pro Leu Asp Ala Phe Val Val Phe Ser Ser Ile Ala Gly Ile Trp
 4260 4265 4270

Gly Gly Gly Gly Gln Gly Pro Ala Gly Ala Ala Asn Ala Val Leu Asp
 4275 4280 4285

Ala Leu Val Glu Trp Arg Arg Ala Arg Gly Leu Lys Ala Thr Ser Ile
 4290 4295 4300

Ala Trp Gly Ala Leu Asp Gln Ile Gly Ile Gly Met Asp Glu Ala Ala
 4305 4310 4315 4320

Leu Ala Gln Leu Arg Arg Arg Gly Val Ile Pro Met Ala Pro Pro Leu
 4325 4330 4335

Ala Val Thr Ala Met Val Gln Ala Val Ala Gly Asn Glu Lys Ala Val
 4340 4345 4350

Ala Val Ala Asp Met Asp Trp Ala Ala Phe Ile Pro Ala Phe Thr Ser
 4355 4360 4365

Val Arg Pro Ser Pro Leu Phe Ala Asp Leu Pro Glu Ala Lys Ala Ile
 4370 4375 4380

Leu Arg Ala Ala Gln Asp Asp Gly Glu Asp Gly Asp Thr Ala Ser Ser
 4385 4390 4395 4400

Leu Ala Asp Ser Leu Arg Ala Val Pro Asp Ala Glu Gln Asn Arg Ile
 4405 4410 4415

Leu Leu Lys Leu Val Arg Gly His Ala Ser Thr Val Leu Gly His Ser			
4420	4425	4430	
Gly Ala Glu Gly Ile Gly Pro Arg Gln Ala Phe Gln Glu Val Gly Phe			
4435	4440	4445	
Asp Ser Leu Ala Ala Val Asn Leu Arg Asn Ser Leu His Ala Ala Thr			
4450	4455	4460	
Gly Leu Arg Leu Pro Ala Thr Leu Ile Phe Asp Tyr Pro Thr Pro Glu			
4465	4470	4475	4480
Ala Leu Val Gly Tyr Leu Arg Val Glu Leu Leu Arg Glu Ala Asp Asp			
4485	4490	4495	
Gly Leu Asp Gly Arg Glu Asp Asp Leu Arg Arg Val Leu Ala Ala Val			
4500	4505	4510	
Pro Phe Ala Arg Phe Lys Glu Ala Gly Val Leu Asp Thr Leu Leu Gly			
4515	4520	4525	
Leu Ala Asp Thr Gly Thr Glu Pro Gly Thr Asp Ala Glu Thr Thr Glu			
4530	4535	4540	
Ala Ala Pro Ala Ala Asp Asp Ala Glu Leu Ile Asp Ala Leu Asp Ile			
4545	4550	4555	4560
Ser Gly Leu Val Gln Arg Ala Leu Gly Gln Thr Ser			
4565	4570		

(2) Information für SEQ ID Nr: 5

(i) Sequenzeigenschaften:

- | | |
|----------------|------------------|
| (A) Länge: | 5069 Aminosäuren |
| (B) Typ: | Aminosäure |
| (C) Strangart: | Einzelstrang |
| (D) Topologie: | linear |

(ii) Molekültyp:

Peptid

(xi) Sequenzbeschreibung: SEQ ID Nr: 5

Met Ala Asn Gln Ser Trp Arg Lys Asn Met Ser Ala Pro Asn Glu Gln
 1 5 10 15

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

Gln Glu Asn Ser Ala Leu Ala Ala Ala Ala Glu Pro Val Ala Ile
 35 40 45

Val Ser Met Ala Cys Arg Tyr Ala Gly Gly Ile Arg Gly Pro Glu Asp
 50 55 60

Phe Trp Arg Val Val Ser Glu Gly Ala Asp Val Tyr Thr Gly Phe Pro
 65 70 75 80

Glu Asp Arg Gly Trp Asp Val Glu Gly Leu Tyr His Pro Asp Pro Asp
 85 90 95

Asn Pro Gly Thr Thr Tyr Val Arg Glu Gly Ala Phe Leu Gln Asp Ala
 100 105 110

Ala Gln Phe Asp Ala Gly Phe Phe Gly Ile Ser Pro Arg Glu Ala Leu
 115 120 125

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

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

Gly Val Tyr Ala Gly Val Val His Gln Asp Tyr Ala Pro Asp Leu Ser
 165 170 175

Gly Phe Glu Gly Phe Met Ser Leu Glu Arg Ala Leu Gly Thr Ala Gly
 180 185 190

Gly Val Ala Ser Gly Arg Val Ala Tyr Thr Leu Gly Leu Glu Gly Pro
 195 200 205

Ala Val Thr Val Asp Thr Met Cys Ser Ser Ser Leu Val Ala Ile His
 210 215 220

Leu Ala Ala Gln Ala Leu Arg Arg Gly Glu Cys Ser Met Ala Leu Ala
 225 230 235 240

Gly Gly Ser Thr Val Met Ala Thr Pro Gly Gly Phe Val Gly Phe Ala
 245 250 255

Arg Gln Arg Ala Leu Ala Phe Asp Gly Arg Cys Lys Ser Tyr Ala Ala
 260 265 270

Ala Ala Asp Gly Ser Gly Trp Ala Glu Gly Val Gly Val Leu Leu Leu
 275 280 285

Glu Arg Leu Ser Val Ala Arg Glu Arg Gly His Gln Val Leu Ala Val
 290 295 300

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

Ala Pro Asn Gly Pro Ala Gln Gln Arg Val Ile Arg Lys Ala Leu Ala
 325 330 335

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

Thr Gly Thr Val Leu Gly Asp Pro Ile Glu Val Gln Ala Leu Leu Ala
 355 360 365

Thr Tyr Gly Gln Gly Arg Asp Pro Gln Gln Pro Leu Trp Leu Gly Ser
 370 375 380

Val Lys Ser Val Val Gly His Thr Gln Ala Ala Ser Gly Val Ala Gly
 385 390 395 400

Val Ile Lys Met Val Gln Ser Leu Arg His Gly Gln Leu Pro Ala Thr
 405 410 415

Gln His Val Asp Ala Pro Thr Pro Gln Val Asp Trp Ser Ala Gly Ala
 420 425 430

Ile Glu Leu Leu Ala Glu Gly Arg Glu Trp Pro Arg Asn Gly His Pro
 435 440 445

Arg Arg Gly Gly Ile Ser Ser Phe Gly Ala Ser Gly Thr Asn Ala His
 450 455 460

Met Ile Leu Glu Glu Ala Pro Glu Asp Glu Pro Val Thr Glu Ala Pro
 465 470 475 480

Ala Pro Thr Gly Val Val Pro Leu Val Val Ser Ala Ala Thr Ala Ala
 485 490 495

Ser Leu Ala Ala Gln Ala Gly Arg Leu Ala Glu Val Gly Asp Val Ser
 500 505 510

Leu Ala Asp Val Ala Gly Thr Leu Val Ser Gly Arg Ala Met Leu Ser
 515 520 525

Glu Arg Ala Val Val Val Ala Gly Ser His Glu Glu Ala Val Thr Gly
 530 535 540

Leu Arg Ala Leu Ala Arg Gly Glu Ser Ala Pro Gly Leu Leu Ser Gly

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

Ile Arg Arg Val Ala Val Asp Tyr Ala Ser His Thr Arg His Val Glu
755 760 765

Ala Ala Arg Asp Ala Leu Ala Glu Met Leu Gly Gly Ile Arg Ala Gln
770 775 780

Ala Pro Glu Val Pro Phe Tyr Ser Thr Val Thr Gly Gly Trp Val Glu
785 790 795 800

Asp Ala Gly Val Leu Asp Gly Gly Tyr Trp Tyr Arg Asn Leu Arg Arg
805 810 815

Gln Val Arg Phe Gly Pro Ala Val Ala Glu Leu Ile Glu Gln Gly His
820 825 830

Arg Val Phe Val Glu Val Ser Ala His Pro Val Leu Val Gln Pro Ile
835 840 845

Asn Glu Leu Val Asp Asp Thr Glu Ala Val Val Thr Gly Thr Leu Arg
850 855 860

Arg Glu Asp Gly Gly Leu Arg Arg Leu Leu Ala Ser Ala Ala Glu Leu
865 870 875 880

Phe Val Arg Gly Val Thr Val Asp Trp Ser Gly Val Leu Pro Pro Ser
885 890 895

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

Leu Gln Met Gly Gly Ser Ala Thr Asp Ala Val Ser Leu Gly Leu Ala
915 920 925

Gly Ala Asp His Pro Leu Leu Gly Ala Val Val Pro Leu Pro Gln Ser
930 935 940

Asp Gly Leu Val Phe Thr Ser Arg Leu Ser Leu Lys Ser His Pro Trp
 945 950 955 960

Leu Ala Gly His Ala Ile Gly Gly Val Val Leu Ile Pro Gly Thr Val
 965 970 975

Tyr Val Asp Leu Ala Leu Arg Ala Gly Asp Glu Leu Gly Phe Gly Val
 980 985 990

Leu Glu Glu Leu Val Ile Glu Ala Pro Leu Val Leu Gly Glu Arg Gly
 995 1000 1005

Gly Val Arg Val Gln Val Ala Val Ser Gly Pro Asn Glu Thr Gly Ser
 1010 1015 1020

Arg Ala Val Asp Val Phe Ser Met Arg Glu Asp Gly Asp Glu Trp Thr
 1025 1030 1035 1040

Arg His Ala Thr Gly Leu Leu Gly Ala Ser Thr Ser Arg Glu Pro Ser
 1045 1050 1055

Arg Phe Asp Phe Ala Ala Trp Pro Pro Ala Gly Ala Glu Pro Ile Asp
 1060 1065 1070

Val Glu Asn Phe Tyr Thr Asp Leu Thr Glu Arg Gly Tyr Ala Tyr Ser
 1075 1080 1085

Gly Ala Phe Gln Gly Met Arg Ala Val Trp Arg Arg Gly Asp Glu Val
 1090 1095 1100

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

Phe Gly Leu His Pro Ala Leu Leu Asp Ala Ala Leu His Thr Asn Ala
 1125 1130 1135

Phe Ala Asn Pro Asp Asp Asp Arg Ser Val Leu Pro Phe Ala Trp Asn

1140	1145	1150
Gly Leu Val Leu His Ala Val Gly Ala Ser Ala Leu Arg Val Arg Val		
1155	1160	1165
Ala Pro Gly Gly Pro Asp Ala Leu Thr Phe Gln Ala Ala Asp Glu Thr		
1170	1175	1180
Gly Gly Leu Val Val Thr Met Asp Ser Leu Val Ser Arg Glu Val Ser		
1185	1190	1195
		1200
Ala Ala Gln Leu Glu Thr Ala Ala Gly Glu Glu Arg Asp Ser Leu Phe		
1205	1210	1215
Gln Val Asp Trp Ile Glu Val Pro Ala Thr Glu Thr Ala Ala Thr Glu		
1220	1225	1230
His Ala Glu Val Leu Glu Ala Phe Gly Glu Ala Ala Pro Leu Glu Leu		
1235	1240	1245
Thr Ser Arg Val Leu Glu Ala Val Gln Ser Trp Leu Ala Asp Ala Ala		
1250	1255	1260
Asp Glu Ala Arg Leu Val Val Val Thr Arg Gly Ala Val Arg Glu Val		
1265	1270	1275
		1280
Thr Asp Pro Ala Gly Ala Ala Val Trp Gly Leu Val Arg Ala Ala Gln		
1285	1290	1295
Ala Glu Asn Pro Gly Arg Ile Ile Leu Val Asp Thr Asp Gly Asp Val		
1300	1305	1310
Pro Leu Gly Ala Val Leu Ala Ser Gly Glu Pro Gln Leu Ala Val Arg		
1315	1320	1325
Gly Asn Ala Phe Ser Val Pro Arg Leu Ala Arg Ala Thr Gly Glu Val		
1330	1335	1340

Pro Glu Ala Pro Ala Val Phe Ser Pro Glu Gly Thr Val Leu Leu Thr
 1345 1350 1355 1360

Gly Gly Thr Gly Ser Leu Gly Gly Leu Val Ala Lys His Leu Val Ala
 1365 1370 1375

Arg His Gly Val Arg Arg Leu Val Leu Ala Ser Arg Arg Gly Val Ala
 1380 1385 1390

Ala Glu Asp Leu Val Thr Glu Leu Thr Glu Gln Gly Ala Thr Val Ser
 1395 1400 1405

Val Val Ala Cys Asp Val Ser Asp Arg Asp Gln Val Ala Ala Leu Leu
 1410 1415 1420

Ala Glu His Arg Pro Thr Gly Ile Val His Leu Ala Gly Leu Leu Asp
 1425 1430 1435 1440

Asp Gly Val Ile Gly Ala Leu Asn Arg Glu Arg Leu Ala Gly Val Phe
 1445 1450 1455

Ala Pro Lys Val Asp Ala Val Gln His Leu Asp Glu Leu Thr Arg Asp
 1460 1465 1470

Leu Gly Leu Asp Ala Phe Val Val Phe Ser Ser Ala Ala Ala Leu Met
 1475 1480 1485

Gly Ser Ala Gly Gln Gly Asn Tyr Ala Ala Ala Asn Ala Phe Leu Asp
 1490 1495 1500

Gly Leu Met Ala Gly Arg Arg Ala Ala Gly Leu Pro Gly Val Ser Leu
 1505 1510 1515 1520

Ala Trp Gly Leu Trp Glu Gln Ala Asp Gly Leu Thr Ala Asn Leu Ser
 1525 1530 1535

Ala Thr Asp Gln Ala Arg Met Ser Arg Gly Gly Val Leu Pro Met Thr
 1540 1545 1550

Pro Ala Glu Ala Leu Asp Ile Phe Asp Ile Gly Leu Ala Ala Glu Gln
 1555 1560 1565

Ala Leu Leu Val Pro Ile Lys Leu Asp Leu Arg Thr Leu Arg Gly Gln
 1570 1575 1580

Ala Thr Ala Gly Gly Glu Val Pro His Leu Leu Arg Gly Leu Val Arg
 1585 1590 1595 1600

Ala Ser Arg Arg Val Thr Arg Thr Ala Ala Ala Ser Gly Gly Gly Gly
 1605 1610 1615

Leu Val His Lys Leu Ala Gly Arg Pro Ala Glu Glu Gln Glu Ala Val
 1620 1625 1630

Leu Leu Gly Ile Val Gln Ala Glu Ala Ala Ala Val Leu Gly Phe Asn
 1635 1640 1645

Ala Pro Glu Leu Ala Gln Gly Thr Arg Gly Phe Ser Asp Leu Gly Phe
 1650 1655 1660

Asp Ser Leu Thr Ala Val Glu Leu Arg Asn Arg Leu Ser Ala Ala Thr
 1665 1670 1675 1680

Gly Val Lys Leu Pro Ala Thr Leu Val Phe Asp Tyr Pro Thr Pro Val
 1685 1690 1695

Ala Leu Ala Arg His Leu Arg Glu Glu Leu Gly Glu Thr Val Ala Gly
 1700 1705 1710

Ala Pro Ala Thr Pro Val Thr Thr Val Ala Asp Ala Gly Glu Pro Ile
 1715 1720 1725

Ala Ile Val Gly Met Ala Cys Arg Leu Pro Gly Gly Val Met Ser Pro

1730

1735

1740

Asp Asp Leu Trp Arg Met Val Ala Glu Gly Arg Asp Gly Met Ser Pro
 1745 1750 1755 1760

Phe Pro Gly Asp Arg Gly Trp Asp Leu Asp Gly Leu Phe Asp Ser Asp
 1765 1770 1775

Pro Glu Arg Pro Gly Thr Ala Tyr Ile Arg Gln Gly Gly Phe Leu His
 1780 1785 1790

Glu Ala Ala Leu Phe Asp Pro Gly Phe Phe Gly Ile Ser Pro Arg Glu
 1795 1800 1805

Ala Leu Ala Met Asp Pro Gln Gln Arg Leu Leu Leu Glu Ala Ser Trp
 1810 1815 1820

Glu Ala Leu Glu Arg Ala Gly Ile Asp Pro Thr Lys Ala Arg Gly Asp
 1825 1830 1835 1840

Ala Val Gly Val Phe Ser Gly Val Ser Ile His Asp Tyr Leu Glu Ser
 1845 1850 1855

Leu Ser Asn Met Pro Ala Glu Leu Glu Gly Phe Val Thr Thr Ala Thr
 1860 1865 1870

Ala Gly Ser Val Ala Ser Gly Arg Val Ser Tyr Thr Phe Gly Phe Glu
 1875 1880 1885

Gly Pro Ala Val Thr Val Asp Thr Ala Cys Ser Ser Ser Leu Val Ala
 1890 1895 1900

Ile His Leu Ala Ala Gln Ala Leu Arg Gln Gly Glu Cys Thr Met Ala
 1905 1910 1915 1920

Leu Ala Gly Gly Val Ala Val Met Gly Ser Pro Ile Gly Val Ile Gly
 1925 1930 1935

Met Ser Arg Gln Arg Gly Met Ala Glu Asp Gly Arg Val Lys Ala Phe
1940 1945 1950

Ala Asp Gly Ala Asp Gly Thr Val Leu Ser Glu Gly Val Gly Ile Val
1955 1960 1965

Val Leu Glu Arg Leu Ser Val Ala Arg Glu Arg Gly His Arg Val Leu
1970 1975 1980

Ala Val Leu Arg Gly Ser Ala Val Asn Gln Asp Gly Ala Ser Asn Gly
1985 1990 1995 2000

Leu Thr Ala Pro Asn Gly Pro Ser Gln Gln Arg Val Ile Arg Ser Ala
2005 2010 2015

Leu Ala Gly Ala Gly Leu Gln Pro Ser Glu Val Asp Val Val Glu Ala
2020 2025 2030

His Gly Thr Gly Thr Ala Leu Gly Glu Pro Ile Glu Ala Gln Ala Leu
2035 2040 2045

Leu Ala Thr Tyr Gly Lys Ser Arg Glu Thr Pro Leu Trp Leu Gly Ser
2050 2055 2060

Leu Lys Ser Asn Ile Gly His Thr Gln Ala Ala Ala Gly Val Ala Ala
2065 2070 2075 2080

Val Ile Lys Met Val Gln Ala Leu Arg Gln Asp Thr Leu Pro Pro Thr
2085 2090 2095

Leu His Val Gln Glu Pro Thr Lys Gln Val Asp Trp Ser Ala Gly Ala
2100 2105 2110

Val Glu Leu Leu Thr Glu Gly Arg Glu Trp Ala Arg Asn Gly His Pro
2115 2120 2125

Arg Arg Ala Gly Val Ser Ser Phe Gly Ile Ser Gly Thr Asn Ala His			
2130	2135	2140	
Leu Ile Leu Glu Glu Ala Pro Ala Asp Asp Thr Ala Glu Ala Asp Val			
2145	2150	2155	2160
Pro Asp Ala Val Val Pro Val Val Ile Ser Ala Arg Ser Thr Gly Ser			
	2165	2170	2175
Leu Ala Gly Gln Ala Gly Arg Leu Ala Ala Phe Leu Asp Gly Asp Val			
	2180	2185	2190
Pro Leu Thr Arg Val Ala Gly Ala Leu Leu Ser Thr Arg Ala Thr Leu			
	2195	2200	2205
Thr Asp Arg Ala Val Val Val Ala Gly Ser Ala Glu Glu Ala Arg Ala			
	2210	2215	2220
Gly Leu Thr Ala Leu Ala Arg Gly Glu Ser Ala Ser Gly Leu Val Thr			
2225	2230	2235	2240
Gly Thr Ala Gly Met Pro Gly Lys Thr Val Trp Val Phe Pro Gly Gln			
	2245	2250	2255
Gly Thr Gln Trp Ala Gly Met Gly Arg Glu Leu Leu Glu Ala Ser Pro			
	2260	2265	2270
Val Phe Ala Glu Arg Ile Glu Glu Cys Ala Ala Ala Leu Gln Pro Trp			
	2275	2280	2285
Ile Asp Trp Ser Leu Leu Asp Val Leu Arg Gly Glu Gly Glu Leu Asp			
	2290	2295	2300
Arg Val Asp Val Leu Gln Pro Ala Cys Phe Ala Val Met Val Gly Leu			
2305	2310	2315	2320
Ala Ala Val Trp Ala Ser Val Gly Val Val Pro Asp Ala Val Leu Gly			

2325	2330	2335	
His Ser Gln Gly Glu Ile Ala Ala Ala Cys Val Ser Gly Ala Leu Ser			
2340	2345	2350	
Leu Glu Asp Ala Ala Lys Val Val Ala Leu Arg Ser Gln Ala Ile Ala			
2355	2360	2365	
Ala Glu Leu Ser Gly Arg Gly Gly Met Ala Ser Ile Gln Leu Ser His			
2370	2375	2380	
Asp Glu Val Ala Ala Arg Leu Ala Pro Trp Ala Gly Arg Val Glu Ile			
2385	2390	2395	2400
Ala Ala Val Asn Gly Pro Ala Ser Val Val Ile Ala Gly Asp Ala Glu			
2405	2410	2415	
Ala Leu Thr Glu Ala Val Glu Val Leu Gly Gly Arg Arg Val Ala Val			
2420	2425	2430	
Asp Tyr Ala Ser His Thr Arg His Val Glu Asp Ile Gln Asp Thr Leu			
2435	2440	2445	
Ala Glu Thr Leu Ala Gly Ile Asp Ala Gln Ala Pro Val Val Pro Phe			
2450	2455	2460	
Tyr Ser Thr Val Ala Gly Glu Trp Ile Thr Asp Ala Gly Val Val Asp			
2465	2470	2475	2480
Gly Gly Tyr Trp Tyr Arg Asn Leu Arg Asn Gln Val Gly Phe Gly Pro			
2485	2490	2495	
Ala Val Ala Glu Leu Ile Glu Gln Gly His Gly Val Phe Val Glu Val			
2500	2505	2510	
Ser Ala His Pro Val Leu Val Gln Pro Ile Ser Glu Leu Thr Asp Ala			
2515	2520	2525	

Val Val Thr Gly Thr Leu Arg Arg Asp Asp Gly Gly Val Arg Arg Leu
 2530 2535 2540

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

Ala Thr Met Ala Pro Pro Ala Arg Val Glu Leu Pro Thr Tyr Ala Phe
 2565 2570 2575

Asp His Gln His Phe Trp Leu Ser Pro Pro Ala Val Ala Asp Ala Pro
 2580 2585 2590

Ala Leu Gly Leu Ala Gly Ala Asp His Pro Leu Leu Gly Ala Val Leu
 2595 2600 2605

Pro Leu Pro Gln Ser Asp Gly Leu Val Phe Thr Ser Arg Leu Ser Val
 2610 2615 2620

Arg Thr His Pro Trp Leu Ala Asp Gly Val Pro Ala Ala Ala Leu Val
 2625 2630 2635 2640

Glu Leu Ala Val Arg Ala Gly Asp Glu Ala Gly Cys Pro Val Leu Ala
 2645 2650 2655

Asp Leu Thr Val Glu Lys Leu Leu Val Leu Pro Glu Ser Gly Gly Leu
 2660 2665 2670

Arg Val Gln Val Ile Val Ser Gly Glu Arg Thr Val Glu Val Tyr Ser
 2675 2680 2685

Gln Leu Glu Gly Ala Glu Asp Trp Ile Arg Asn Ala Thr Gly His Leu
 2690 2695 2700

Ser Ala Thr Ala Pro Ala His Glu Ala Phe Asp Phe Thr Ala Trp Pro
 2705 2710 2715 2720

Pro Ala Gly Ala Gln Gln Val Asp Gly Leu Trp Arg Arg Gly Asp Glu
 2725 2730 2735

Ile Phe Ala Glu Val Ala Leu Pro Glu Glu Leu Asp Ala Gly Ala Phe
 2740 2745 2750

Gly Ile His Pro Phe Leu Leu Asp Ala Ala Val Gln Pro Val Leu Ala
 2755 2760 2765

Asp Asp Glu Gln Pro Ala Glu Trp Arg Ser Leu Val Leu His Ala Ala
 2770 2775 2780

Gly Ala Ser Ala Leu Arg Val Arg Leu Val Pro Gly Gly Ala Leu Gln
 2785 2790 2795 2800

Ala Ala Asp Glu Thr Gly Gly Leu Val Leu Thr Ala Asp Ser Val Ala
 2805 2810 2815

Gly Arg Glu Leu Ser Ala Gly Lys Thr Arg Ala Gly Ser Leu Tyr Arg
 2820 2825 2830

Val Asp Trp Thr Glu Val Ser Ile Ala Asp Ser Ala Val Pro Ala Asn
 2835 2840 2845

Ile Glu Val Val Glu Ala Phe Gly Glu Glu Pro Leu Glu Leu Thr Gly
 2850 2855 2860

Arg Val Leu Glu Ala Val Gln Thr Trp Leu Val Thr Ala Ala Asp Asp
 2865 2870 2875 2880

Ala Arg Leu Val Val Val Thr Arg Gly Ala Val Arg Glu Val Thr Asp
 2885 2890 2895

Pro Ala Gly Ala Ala Val Trp Gly Leu Val Arg Ala Ala Gln Ala Glu
 2900 2905 2910

Asn Pro Gly Arg Ile Phe Leu Ile Asp Thr Asp Gly Glu Ile Pro Ala

2915	2920	2925
Leu Thr Gly Asp Glu Pro Glu Ile Ala Val Arg Gly Gly Lys Phe Phe		
2930	2935	2940
Val Pro Arg Ile Thr Arg Ala Glu Pro Ser Gly Ala Ala Val Phe Arg		
2945	2950	2955
		2960
Pro Asp Gly Thr Val Leu Ile Ser Gly Ala Gly Ala Leu Gly Gly Leu		
2965	2970	2975
Val Ala Arg Arg Leu Val Glu Arg His Gly Val Arg Lys Leu Val Leu		
2980	2985	2990
Ala Ser Arg Arg Gly Arg Asp Ala Asp Gly Val Ala Asp Leu Val Ala		
2995	3000	3005
Asp Leu Ala Ala Asp Val Ser Val Val Ala Cys Asp Val Ser Asp Arg		
3010	3015	3020
Ala Gln Val Ala Ala Leu Leu Asp Glu His Arg Pro Thr Ala Val Val		
3025	3030	3035
		3040
His Thr Ala Gly Val Ile Asp Ala Gly Val Ile Glu Thr Leu Asp Arg		
3045	3050	3055
Asp Arg Leu Ala Thr Val Phe Ala Pro Lys Val Asp Ala Val Arg His		
3060	3065	3070
Leu Asp Glu Leu Thr Arg Asp Arg Asp Leu Asp Ala Phe Val Val Tyr		
3075	3080	3085
Ser Ser Val Ser Ala Val Phe Met Gly Ala Gly Ser Gly Ser Tyr Ala		
3090	3095	3100
Ala Ala Asn Ala Phe Leu Asp Gly Leu Met Ala Asn Arg Arg Ala Ala		
3105	3110	3115
		3120

Gly Leu Pro Gly Leu Ser Leu Ala Trp Gly Leu Trp Asp Gln Ser Thr
 3125 3130 3135

Gly Met Ala Ala Gly Thr Asp Glu Ala Thr Arg Ala Arg Met Ser Arg
 3140 3145 3150

Arg Gly Gly Leu Gln Ile Met Thr Gln Ala Glu Gly Met Asp Leu Phe
 3155 3160 3165

Asp Ala Ala Leu Ser Ser Ala Glu Ser Leu Leu Val Pro Ala Lys Leu
 3170 3175 3180

Asp Leu Arg Gly Val Arg Ala Asp Ala Ala Ala Gly Gly Val Val Pro
 3185 3190 3195 3200

His Met Leu Arg Gly Leu Val Arg Ala Gly Arg Ala Gln Ala Arg Ala
 3205 3210 3215

Ala Ser Thr Val Asp Asn Gly Leu Ala Gly Arg Leu Ala Gly Leu Ala
 3220 3225 3230

Pro Ala Asp Gln Leu Thr Leu Leu Leu Asp Leu Val Arg Ala Gln Val
 3235 3240 3245

Ala Ala Val Leu Gly His Ala Asp Ala Ser Ala Val Arg Val Asp Thr
 3250 3255 3260

Ala Phe Lys Asp Ala Gly Phe Asp Ser Leu Thr Ala Val Glu Leu Arg
 3265 3270 3275 3280

Asn Arg Met Arg Thr Ala Thr Gly Leu Lys Leu Pro Ala Thr Leu Val
 3285 3290 3295

Phe Asp Tyr Pro Asn Pro Gln Ala Leu Ala Arg His Leu Arg Asp Glu
 3300 3305 3310

Leu Gly Gly Ala Ala Gln Thr Pro Val Thr Thr Ala Ala Ala Lys Ala
 3315 3320 3325

Asp Leu Asp Glu Pro Ile Ala Ile Val Gly Met Ala Cys Arg Leu Pro
 3330 3335 3340

Gly Gly Val Ala Gly Pro Glu Asp Leu Trp Arg Leu Val Ala Glu Gly
 3345 3350 3355 3360

Arg Asp Ala Val Ser Ser Phe Pro Thr Asp Arg Gly Trp Asp Thr Asp
 3365 3370 3375

Ser Leu Tyr Asp Pro Asp Pro Ala Arg Pro Gly Lys Thr Tyr Thr Arg
 3380 3385 3390

His Gly Gly Phe Leu His Glu Ala Gly Leu Phe Asp Ala Gly Phe Phe
 3395 3400 3405

Gly Ile Ser Pro Arg Glu Ala Val Ala Met Asp Pro Gln Gln Arg Leu
 3410 3415 3420

Leu Leu Glu Ala Ser Trp Glu Ala Met Glu Asp Ala Gly Val Asp Pro
 3425 3430 3435 3440

Leu Ser Leu Lys Gly Asn Asp Val Gly Val Phe Thr Gly Met Phe Gly
 3445 3450 3455

Gln Gly Tyr Val Ala Pro Gly Asp Ser Val Val Thr Pro Glu Leu Glu
 3460 3465 3470

Gly Phe Ala Gly Thr Gly Gly Ser Ser Ser Val Ala Ser Gly Arg Val
 3475 3480 3485

Ser Tyr Val Phe Gly Phe Glu Gly Pro Ala Val Thr Ile Asp Ser Ala
 3490 3495 3500

Cys Ser Ser Ser Leu Val Ala Met His Leu Ala Ala Gln Ser Leu Arg

3505	3510	3515	3520
Gln Gly Glu Cys Ser Met Ala Leu Ala Gly Gly Ala Thr Val Met Ala			
3525	3530	3535	
Asn Pro Gly Ala Phe Val Glu Phe Ser Arg Gln Arg Gly Leu Ala Val			
3540	3545	3550	
Asp Gly Arg Cys Lys Ala Phe Ala Ala Ala Ala Asp Gly Thr Gly Trp			
3555	3560	3565	
Ala Glu Gly Val Gly Val Val Ile Leu Glu Arg Leu Ser Val Ala Arg			
3570	3575	3580	
Glu Arg Gly His Arg Ile Leu Ala Val Leu Arg Gly Ser Ala Val Asn			
3585	3590	3595	3600
Gln Asp Gly Ala Ser Asn Gly Leu Thr Ala Pro Asn Gly Pro Ser Gln			
3605	3610	3615	
Gln Arg Val Ile Arg Arg Ala Leu Val Ser Ala Gly Leu Ala Pro Ser			
3620	3625	3630	
Asp Val Asp Val Val Glu Ala His Gly Thr Gly Thr Thr Leu Gly Asp			
3635	3640	3645	
Pro Ile Glu Ala Gln Ala Leu Leu Ala Thr Tyr Gly Lys Asp Arg Glu			
3650	3655	3660	
Ser Pro Leu Trp Leu Gly Ser Leu Lys Ser Asn Ile Gly His Ala Gln			
3665	3670	3675	3680
Ala Ala Ala Gly Val Ala Gly Val Ile Lys Met Val Gln Ala Leu Arg			
3685	3690	3695	
His Glu Val Leu Pro Pro Thr Leu His Val Asp Arg Pro Thr Pro Glu			
3700	3705	3710	

Val Asp Trp Ser Ala Gly Ala Val Glu Leu Leu Thr Glu Ala Arg Glu
 3715 3720 3725

Trp Pro Arg Asn Gly Arg Pro Arg Arg Ala Gly Val Ser Ala Phe Gly
 3730 3735 3740

Val Ser Gly Thr Asn Ala His Leu Ile Leu Glu Glu Ala Pro Ala Glu
 3745 3750 3755 3760

Glu Pro Val Pro Thr Pro Glu Val Pro Leu Val Pro Val Val Val Ser
 3765 3770 3775

Ala Arg Ser Arg Ala Ser Leu Ala Gly Gln Ala Gly Arg Leu Ala Gly
 3780 3785 3790

Phe Val Ala Gly Asp Ala Ser Leu Ala Gly Val Ala Arg Ala Leu Val
 3795 3800 3805

Thr Asn Arg Ala Ala Leu Thr Glu Arg Ala Val Met Val Val Gly Ser
 3810 3815 3820

Arg Glu Glu Ala Val Thr Asn Leu Glu Ala Leu Ala Arg Gly Glu Asp
 3825 3830 3835 3840

Pro Ala Ala Val Val Thr Gly Arg Ala Gly Ser Pro Gly Lys Leu Val
 3845 3850 3855

Trp Val Phe Pro Gly Gln Gly Ser Gln Trp Ile Gly Met Gly Arg Glu
 3860 3865 3870

Leu Leu Asp Ser Ser Pro Val Phe Ala Glu Arg Val Ala Glu Cys Ala
 3875 3890 3885

Ala Ala Leu Glu Pro Trp Ile Asp Trp Ser Leu Leu Asp Val Leu Arg
 3890 3895 3900

Gly Glu Ser Asp Leu Leu Asp Arg Val Asp Val Val Gln Pro Ala Ser
 3905 3910 3915 3920

Phe Ala Met Met Val Gly Leu Ala Ala Val Trp Gln Ser Val Gly Val
 3925 3930 3935

Arg Pro Asp Ala Val Val Gly His Ser Gln Gly Glu Ile Ala Ala Ala
 3940 3945 3950

Cys Val Ser Gly Ala Leu Ser Leu Gln Asp Ala Ala Lys Val Val Ala
 3955 3960 3965

Leu Arg Ser Gln Ala Ile Ala Thr Arg Leu Ala Gly Arg Gly Gly Met
 3970 3975 3980

Ala Ser Val Ala Leu Ser Glu Glu Asp Ala Thr Ala Trp Leu Ala Pro
 3985 3990 3995 4000

Trp Ala Asp Arg Val Gln Val Ala Ala Val Asn Ser Pro Ala Ser Val
 4005 4010 4015

Val Ile Ala Gly Glu Ala Gln Ala Leu Asp Glu Val Val Asp Ala Leu
 4020 4025 4030

Ser Gly Gln Glu Val Arg Val Arg Arg Val Ala Val Asp Tyr Gly Ser
 4035 4040 4045

His Thr Asn Gln Val Glu Ala Ile Glu Asp Leu Leu Ala Glu Thr Leu
 4050 4055 4060

Ala Gly Ile Glu Ala Gln Ala Pro Lys Val Pro Phe Tyr Ser Thr Leu
 4065 4070 4075 4080

Ile Gly Asp Trp Ile Arg Asp Ala Gly Ile Val Asp Gly Gly Tyr Trp
 4085 4090 4095

Tyr Arg Asn Leu Arg Asn Gln Val Gly Phe Gly Pro Ala Val Ala Glu

4100

4105

4110

Leu Val Arg Gln Gly His Gly Val Phe Val Glu Val Ser Ala His Pro
 4115 4120 4125

Val Leu Val Gln Pro Leu Ser Glu Leu Ser Asp Asp Ala Val Val Thr
 4130 4135 4140

Gly Ser Leu Arg Arg Glu Asp Gly Gly Leu Arg Arg Leu Leu Thr Ser
 4145 4150 4155 4160

Met Ala Glu Leu Tyr Val Gln Gly Val Pro Leu Asp Trp Thr Ala Val
 4165 4170 4175

Leu Pro Arg Thr Gly Arg Val Asp Leu Pro Lys Tyr Ala Phe Asp His
 4180 4185 4190

Arg His Tyr Trp Leu Arg Pro Ala Glu Ser Ala Thr Asp Ala Ala Ser
 4195 4200 4205

Leu Gly Gln Ala Ala Ala Asp His Pro Leu Leu Gly Ala Val Val Glu
 4210 4215 4220

Leu Pro Gln Ser Asp Gly Leu Val Phe Thr Ser Arg Leu Ser Val Arg
 4225 4230 4235 4240

Thr His Pro Trp Leu Ala Asp His Ala Val Gly Gly Val Val Ile Leu
 4245 4250 4255

Pro Gly Ser Gly Leu Ala Glu Leu Ala Val Arg Ala Gly Asp Glu Ala
 4260 4265 4270

Gly Cys Thr Ala Leu Asp Glu Leu Ile Ile Glu Ala Pro Leu Val Val
 4275 4280 4285

Pro Ala Gln Gly Ala Val Arg Val Gln Val Ala Leu Ser Gly Pro Asp
 4290 4295 4300

Glu Thr Gly Ser Arg Thr Val Asp Leu Tyr Ser Gln Arg Asp Gly Gly
 4305 4310 4315 4320

Ala Gly Thr Trp Thr Arg His Ala Thr Gly Val Leu Ser Thr Ala Pro
 4325 4330 4335

Ala Gln Glu Pro Glu Phe Asp Phe His Ala Trp Pro Pro Ala Asp Ala
 4340 4345 4350

Glu Arg Ile Asp Val Glu Thr Phe Tyr Thr Asp Leu Ala Glu Arg Gly
 4355 4360 4365

Tyr Gly Tyr Gly Pro Ala Phe Gln Gly Leu Gln Ala Val Trp Arg Arg
 4370 4375 4380

Asp Gly Asp Val Phe Ala Glu Val Ala Leu Pro Glu Asp Leu Arg Lys
 4385 4390 4395 4400

Asp Ala Gly Arg Phe Gly Val His Pro Ala Leu Leu Asp Ala Ala Leu
 4405 4410 4415

Gln Ala Ala Thr Ala Val Gly Gly Asp Glu Pro Gly Gln Pro Val Leu
 4420 4425 4430

Ala Phe Ala Trp Asn Gly Leu Val Leu His Ala Ala Gly Ala Ser Ala
 4435 4440 4445

Leu Arg Val Arg Leu Ala Pro Ser Gly Pro Asp Thr Leu Ser Val Ala
 4450 4455 4460

Ala Ala Asp Glu Thr Gly Gly Leu Val Leu Thr Met Glu Ser Leu Val
 4465 4470 4475 4480

Ser Arg Pro Val Ser Ala Glu Gln Leu Gly Ala Ala Ala Asp Ala Gly
 4485 4490 4495

His Asp Ala Met Phe Arg Val Asp Trp Thr Glu Leu Pro Ala Val Pro
 4500 4505 4510

Arg Ala Glu Leu Pro Pro Trp Val Arg Ile Asp Thr Ala Asp Asp Val
 4515 4520 4525

Ala Ala Leu Ala Glu Lys Ala Asp Ala Pro Pro Val Val Val Trp Glu
 4530 4535 4540

Ala Ala Gly Gly Asp Pro Ala Leu Ala Val Ser Ser Arg Val Leu Glu
 4545 4550 4555 4560

Ile Met Gln Ala Trp Leu Ala Ala Pro Ala Phe Glu Glu Ala Arg Leu
 4565 4570 4575

Val Val Thr Thr Arg Gly Ala Val Pro Ala Gly Gly Asp His Thr Leu
 4580 4585 4590

Thr Asp Pro Ala Ala Ala Ala Val Trp Gly Leu Val Arg Ser Ala Gln
 4595 4600 4605

Ala Glu His Pro Asp Arg Val Val Leu Leu Asp Thr Asp Gly Glu Val
 4610 4615 4620

Pro Leu Gly Ala Val Leu Ala Ser Gly Glu Pro Gln Leu Ala Val Arg
 4625 4630 4635 4640

Gly Thr Thr Phe Phe Val Pro Arg Leu Ala Arg Ala Thr Arg Leu Ser
 4645 4650 4655

Asp Ala Pro Pro Ala Phe Asp Pro Asp Gly Thr Val Leu Val Ser Gly
 4660 4665 4670

Ala Gly Ser Leu Gly Thr Leu Val Ala Arg His Leu Val Thr Arg His
 4675 4680 4685

Gly Val Arg Arg Val Val Leu Ala Ser Arg Gln Gly Arg Asp Ala Glu

4690

4695

4700

Gly Ala Gln Asp Leu Ile Thr Glu Leu Thr Gly Glu Gly Ala Asp Val
 4705 4710 4715 4720

Ser Phe Val Ala Cys Asp Val Ser Asp Arg Asp Gln Val Ala Ala Leu
 4725 4730 4735

Leu Ala Gly Leu Pro Asp Leu Thr Gly Val Val His Thr Ala Gly Val
 4740 4745 4750

Phe Glu Asp Gly Val Ile Glu Ala Leu Thr Pro Asp Gln Leu Ala Asn
 4755 4760 4765

Val Tyr Ala Ala Lys Val Thr Ala Ala Met His Leu Asp Glu Leu Thr
 4770 4775 4780

Arg Asp Arg Asp Leu Gly Ala Phe Val Val Phe Ser Ser Val Ala Gly
 4785 4790 4795 4800

Val Met Gly Gly Gly Gly Gln Gly Pro Tyr Ala Ala Ala Asn Ala Phe
 4805 4810 4815

Leu Asp Ala Ala Met Ala Ser Arg Gln Ala Ala Gly Leu Pro Gly Leu
 4820 4825 4830

Ser Leu Ala Trp Gly Leu Trp Glu Arg Ser Ser Gly Met Ala Ala His
 4835 4840 4845

Leu Ser Glu Val Asp His Ala Arg Ala Ser Arg Asn Gly Val Leu Glu
 4850 4855 4860

Leu Thr Arg Ala Glu Gly Leu Ala Leu Phe Asp Leu Gly Leu Arg Met
 4865 4870 4875 4880

Ala Glu Ser Leu Leu Val Pro Ile Lys Leu Asp Leu Ala Ala Met Arg
 4885 4890 4895

Ala Ser Thr Val Pro Val Leu Phe Arg Gly Leu Val Arg Pro Ser Arg
 4900 4905 4910

Thr Gln Ala Arg Thr Ala Ser Thr Val Asp Arg Gly Leu Ala Gly Arg
 4915 4920 4925

Leu Ala Gly Leu Pro Val Ala Glu Arg Ala Ala Val Leu Val Asp Leu
 4930 4935 4940

Val Arg Gly Gln Val Ala Val Val Leu Gly Tyr Asp Gly Pro Glu Ala
 4945 4950 4955 4960

Val Arg Pro Asp Thr Ala Phe Lys Asp Thr Gly Phe Asp Ser Leu Thr
 4965 4970 4975

Ser Val Glu Leu Arg Asn Arg Leu Arg Glu Ala Thr Gly Leu Lys Leu
 4980 4985 4990

Pro Ala Thr Leu Val Phe Asp Tyr Pro Asn Pro Leu Ala Val Ala Arg
 4995 5000 5005

Tyr Leu Gly Ala Arg Leu Val Pro Asp Gly Thr Ala Asn Gly Asn Gly
 5010 5015 5020

Asn Gly Asn Gly His Ser Glu Asp Asp Arg Leu Arg His Ala Leu Ala
 5025 5030 5035 5040

Ala Ile Ala Ala Glu Asp Ala Gly Glu Glu Arg Ser Ile Ala Asp Leu
 5045 5050 5055

Gly Val Asp Asp Leu Val Gln Leu Ala Phe Gly Asp Glu
 5060 5065

(2) Information für SEQ ID Nr: 6

(i) Sequenzeigenschaften:

(A) Länge: 1721 Aminosäuren

(B) Typ: Aminosäure

(C) Strangart: Einzelstrang

(D) Topologie: linear

(ii) Molekültyp: Peptid

(xi) Sequenzbeschreibung: SEQ ID Nr: 6

Met Ala Cys Arg Leu Pro Gly Gly Val Thr Gly Pro Gly Asp Leu Trp
 1 5 10 15

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

Arg Cys Trp Asp Leu Asp Thr Leu Phe Asp Pro Asp Pro Asp His Ala
 35 40 45

Gly Thr Ser Tyr Thr Asp Gln Gly Gly Phe Leu His Asp Ala Ala Leu
 50 55 60

Phe Asp Pro Gly Phe Phe Gly Ile Ser Pro Arg Glu Ala Leu Ala Met
 65 70 75 80

Asp Pro Gln Gln Arg Leu Leu Leu Glu Ala Ser Trp Glu Ala Leu Glu
 85 90 95

Gly Val Gly Leu Asp Pro Ala Ser Leu Gln Gly Thr Asp Val Gly Val
 100 105 110

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

Glu Met Gln Ser Phe Ala Gly Thr Gly Leu Ala Ser Ser Val Ala Ser

130	135	140	
Gly Arg Val Ser Tyr Val Phe Gly Phe Glu Gly Pro Ala Val Thr Ile			
145	150	155	160
Asp Thr Ala Cys Ser Ser Ser Leu Val Ala Met His Leu Ala Ala Gln			
	165	170	175
Ala Leu Arg Gln Gly Asp Cys Ser Met Ala Leu Ala Gly Gly Ala Met			
	180	185	190
Val Met Ser Gly Pro Asp Ser Phe Val Val Phe Ser Arg Gln Arg Gly			
	195	200	205
Leu Ala Thr Asp Gly Arg Cys Lys Ala Phe Ala Ser Gly Ala Asp Gly			
	210	215	220
Met Val Leu Ala Glu Gly Ile Ser Val Val Val Leu Glu Arg Leu Ser			
225	230	235	240
Val Ala Arg Glu Arg Gly His Arg Val Leu Ala Val Leu Arg Gly Ser			
	245	250	255
Ala Val Asn Gln Asp Gly Ala Ser Asn Gly Leu Thr Ala Pro Asn Gly			
	260	265	270
Pro Ser Gln Gln Arg Val Ile Arg Ala Ala Leu Ala Asn Ala Gly Ile			
	275	280	285
Gly Pro Ser Asp Val Asp Leu Val Glu Ala His Gly Thr Gly Thr Ser			
	290	295	300
Leu Gly Asp Pro Ile Glu Ala Gln Ala Leu Leu Ala Thr Tyr Gly Gln			
305	310	315	320
Asp Arg Glu Thr Pro Leu Trp Leu Gly Ser Leu Lys Ser Asn Ile Gly			
	325	330	335

His Thr Gln Ala Ala Ala Gly Val Ala Ser Val Ile Lys Val Val Gln
 340 345 350

Ala Leu Arg His Gly Val Met Pro Pro Thr Leu His Val Asp Glu Pro
 355 360 365

Ser Ser Gln Val Asp Trp Ser Glu Gly Ala Val Glu Leu Leu Thr Gly
 370 375 380

Ser Arg Asp Trp Pro Arg Gly Asp Arg Pro Arg Arg Ala Gly Val Ser
 385 390 395 400

Ser Phe Gly Val Ser Gly Thr Asn Val His Leu Ile Ile Glu Glu Ala
 405 410 415

Pro Glu Glu Pro Ala Ala Ala Val Pro Thr Ser Ala Asp Val Val Pro
 420 425 430

Leu Val Val Ser Ala Arg Ser Thr Gly Ser Leu Ala Gly Gln Ala Asp
 435 440 445

Arg Leu Thr Glu Val Asp Val Pro Leu Gly His Leu Ala Gly Ala Leu
 450 455 460

Val Ala Gly Arg Ala Val Leu Glu Glu Arg Ala Val Val Val Ala Gly
 465 470 475 480

Ser Ala Glu Glu Ala Arg Ala Gly Leu Gly Ala Leu Ala Arg Gly Glu
 485 490 495

Ala Ala Pro Gly Val Val Thr Gly Thr Ala Gly Lys Pro Gly Lys Val
 500 505 510

Val Trp Val Phe Pro Gly Gln Gly Thr Gln Trp Val Gly Met Gly Arg
 515 520 525

Glu Leu Leu Asp Ala Ser Pro Val Phe Ala Glu Arg Ile Lys Glu Cys			
530	535	540	
Ala Ala Ala Leu Asp Gln Trp Thr Asp Trp Ser Leu Leu Asp Val Leu			
545	550	555	560
Arg Gly Asp Gly Asp Leu Asp Ser Val Glu Val Leu Gln Pro Ala Cys			
	565	570	575
Phe Ala Val Met Val Gly Leu Ala Ala Val Trp Glu Ser Ala Gly Val			
	580	585	590
Arg Pro Asp Ala Val Val Gly His Ser Gln Gly Glu Ile Ala Ala Ala			
	595	600	605
Cys Val Ser Gly Ala Leu Thr Leu Asp Asp Ala Ala Lys Val Val Ala			
	610	615	620
Leu Arg Ser Gln Ala Ile Ala Ala Arg Leu Ser Gly Arg Gly Gly Met			
625	630	635	640
Ala Ser Val Ala Leu Ser Glu Asp Glu Ala Asn Ala Arg Leu Gly Leu			
	645	650	655
Trp Asp Gly Arg Ile Glu Val Ala Ala Val Asn Gly Pro Ala Ser Val			
	660	665	670
Val Ile Ala Gly Asp Ala Gln Ala Leu Asp Glu Ala Leu Glu Val Leu			
	675	680	685
Ala Gly Asp Gly Val Arg Val Arg Gln Val Ala Val Asp Tyr Ala Ser			
	690	695	700
His Thr Arg His Val Glu Asp Ile Arg Asp Thr Leu Ala Glu Thr Leu			
705	710	715	720
Ala Gly Ile Thr Ala Gln Ala Pro Asp Val Pro Phe Arg Ser Thr Val			

725

730

735

Thr Gly Gly Trp Val Arg Asp Ala Asp Val Leu Asp Gly Gly Tyr Trp
 740 745 750

Tyr Arg Asn Leu Arg Asn Gln Val Arg Phe Gly Pro Ala Val Ala Glu
 755 760 765

Leu Leu Glu Gln Gly His Gly Val Phe Val Glu Val Ser Ala His Pro
 770 775 780

Val Leu Val Gln Pro Ile Ser Glu Leu Thr Asp Ala Val Val Thr Gly
 785 790 795 800

Thr Leu Arg Arg Asp Asp Gly Gly Leu Arg Arg Leu Leu Thr Ser Met
 805 810 815

Ala Glu Leu Phe Val Arg Gly Val Arg Val Asp Trp Ala Thr Leu Val
 820 825 830

Pro Pro Ala Arg Val Asp Leu Pro Thr Tyr Ala Phe Asp His Gln His
 835 840 845

Phe Trp Leu Arg Pro Ala Ala Gln Ala Asp Ala Val Ser Leu Gly Gln
 850 855 860

Ala Ala Ala Glu His Pro Leu Leu Gly Ala Val Val Arg Leu Pro Gln
 865 870 875 880

Ser Asp Gly Leu Val Phe Thr Ser Arg Leu Ser Leu Arg Thr His Pro
 885 890 895

Trp Leu Ala Asp His Thr Ile Gly Gly Val Val Leu Phe Pro Gly Thr
 900 905 910

Gly Leu Val Glu Leu Ala Val Arg Ala Gly Asp Glu Ala Gly Cys Pro
 915 920 925

Val Leu Asp Glu Leu Val Thr Glu Ala Pro Leu Val Val Pro Gly Gln
 930 935 940

Gly Gly Val Asn Val Gln Val Thr Val Ser Gly Pro Asp Gln Asn Gly
 945 950 955 960

Leu Arg Thr Val Asp Ile His Ser Gln Arg Asp Asp Val Trp Thr Arg
 965 970 975

His Ala Thr Gly Thr Val Ser Ala Thr Pro Ala Ser Ser Pro Gly Phe
 980 985 990

Asp Phe Thr Ala Trp Pro Pro Pro Asp Gly Gln Arg Val Glu Ile Gly
 995 1000 1005

Asp Phe Tyr Ala Asp Leu Ala Glu Arg Gly Tyr Ala Tyr Gly Pro Leu
 1010 1015 1020

Phe Gln Gly Val Arg Ala Val Trp Gln Arg Gly Glu Asp Val Phe Ala
 1025 1030 1035 1040

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

Leu His Pro Ala Leu Leu Asp Ala Ala Leu Gln Thr Gly Thr Ile Ala
 1060 1065 1070

Ala Ala Ala Ser Gly Gln Pro Gly Lys Ser Val Met Pro Phe Ser Trp
 1075 1080 1085

Asn Arg Leu Ala Leu His Ala Val Gly Ala Ala Gly Leu Arg Val Arg
 1090 1095 1100

Val Ala Pro Gly Gly Pro Asp Ala Leu Thr Val Glu Ala Ala Asp Glu
 1105 1110 1115 1120

Thr Gly Ala Pro Val Leu Thr Met Asp Ser Leu Ile Leu Arg Glu Val			
1125	1130	1135	
Ala Leu Asp Gln Leu Asp Thr Ala Arg Ala Gly Ser Leu Tyr Arg Val			
1140	1145	1150	
Asp Trp Thr Pro Leu Pro Thr Val Asp Ser Ala Val Pro Ala Gly Arg			
1155	1160	1165	
Ala Glu Val Leu Glu Ala Phe Gly Glu Glu Pro Leu Asp Leu Thr Gly			
1170	1175	1180	
Arg Val Leu Ala Ala Leu Gln Ala Trp Leu Ser Asp Ala Ala Glu Glu			
1185	1190	1195	1200
Ala Arg Leu Val Val Val Thr Arg Gly Ala Val Pro Ala Gly Asp Gly			
1205	1210	1215	
Val Val Ser Asp Pro Ala Gly Ala Ala Val Trp Gly Leu Val Arg Ala			
1220	1225	1230	
Ala Gln Ala Glu Asn Pro Asp Arg Phe Val Leu Leu Asp Thr Asp Gly			
1235	1240	1245	
Glu Val Pro Leu Glu Ala Val Leu Ala Thr Gly Glu Pro Gln Leu Ala			
1250	1255	1260	
Leu Arg Gly Thr Thr Phe Ser Val Pro Arg Leu Ala Arg Val Thr Glu			
1265	1270	1275	1280
Pro Ala Glu Ala Pro Leu Thr Phe Arg Pro Asp Gly Thr Val Leu Val			
1285	1290	1295	
Ser Gly Ala Gly Thr Leu Gly Ala Leu Ala Ala Arg Asp Leu Val Thr			
1300	1305	1310	
Arg His Gly Val Arg Arg Leu Val Leu Ala Ser Arg Arg Gly Arg Ala			

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

Leu Arg Glu Val Arg Ala Asp Ala Ala Gly Gly Gly Thr Val Pro His
 1525 1530 1535

Leu Leu Arg Gly Leu Val Arg Ala Gly Arg Gln Ala Ala Arg Thr Ala
 1540 1545 1550

Ala Thr Glu Asp Gly Gly Leu Glu Arg Arg Leu Ala Gly Leu Thr Val
 1555 1560 1565

Ala Glu Gln Glu Ala Leu Leu Leu Asp Leu Val Arg Gly Gln Val Ala
 1570 1575 1580

Val Val Leu Gly His Ala Asp Ser Ser Gly Val Arg Ala Asp Ala Ala
 1585 1590 1595 1600

Phe Lys Asp Ala Gly Phe Asp Ser Leu Thr Ser Val Glu Leu Arg Asn
 1605 1610 1615

Arg Leu Arg Glu Thr Thr Gly Leu Lys Leu Pro Ala Thr Leu Val Phe
 1620 1625 1630

Asp His Pro Asn Pro Leu Ala Leu Ala Arg His Leu Arg Ala Glu Leu
 1635 1640 1645

Ala Val Asp Glu Ala Ser Pro Ala Asp Ala Val Leu Ala Gly Leu Ala
 1650 1655 1660

Gly Leu Glu Ala Ala Ile Ala Ala Ala Gly Ala Pro Asp Gly Asp Arg
 1665 1670 1675 1680

Ile Thr Ala Arg Leu Arg Glu Leu Leu Lys Ala Ala Glu Ala Ala Glu
 1685 1690 1695

Ala Arg Pro Gly Thr Ser Gly Asp Leu Asp Thr Ala Ser Asp Glu Glu
 1700 1705 1710

Leu Phe Ala Leu Val Asp Gly Leu Asp

1715

1720

(2)Information für SEQ ID Nr: 7

(i) Sequenzeigenschaften:

(A) Länge: 1688 Aminosäuren

(B) Typ: Aminosäure

(C) Strangart: Einzelstrang

(D) Topologie: linear

(ii) Molekültyp: Peptid

(xi) Sequenzbeschreibung: SEQ ID Nr: 7

Met Ala Cys Arg Tyr Pro Gly Gly Val Ser Ser Pro Glu Asp Leu Trp

1

5

10

15

Arg Leu Val Ala Glu Gly Thr Asp Ala Val Ser Ala Phe Pro Gly Asp

20

25

30

Arg Gly Trp Asp Val Asp Gly Leu Val Asp Pro Asp Pro Asp Arg Pro

35

40

45

Gly Thr Thr Tyr Thr Asp Gln Gly Gly Phe Leu His Glu Ala Gly Leu

50

55

60

Phe Asp Ala Gly Phe Phe Gly Ile Ser Pro Arg Glu Ala Val Ala Met

65

70

75

80

Asp Pro Gln Gln Arg Leu Leu Leu Glu Thr Ser Trp Glu Ala Ile Glu

85

90

95

Arg Thr Gly Thr Asp Pro Leu Ser Leu Lys Gly Ser Asp Ile Gly Val

100	105	110
Phe Thr Gly Val Ala Ser Met Gly Tyr Gly Ala Gly Gly Gly Val Val		
115	120	125
Ala Pro Glu Leu Glu Gly Phe Val Gly Thr Gly Ala Ala Pro Cys Ile		
130	135	140
Ala Ser Gly Arg Val Ser Tyr Val Leu Gly Phe Glu Gly Pro Ala Val		
145	150	155
Thr Val Asp Thr Gly Cys Ser Ser Ser Leu Val Ala Met His Leu Ala		
165	170	175
Ala Gln Ala Leu Arg Arg Gly Glu Cys Ser Met Ala Leu Ala Gly Gly		
180	185	190
Ala Met Val Met Ala Gln Pro Gly Ser Phe Val Ser Phe Ser Arg Gln		
195	200	205
Arg Gly Leu Ala Leu Asp Gly Arg Cys Lys Ala Phe Ser Asp Ser Ala		
210	215	220
Asp Gly Met Gly Leu Ala Glu Gly Val Gly Val Ile Ala Leu Glu Arg		
225	230	235
Leu Ser Val Ala Arg Glu Arg Gly His Arg Val Leu Ala Val Leu Arg		
245	250	255
Gly Ile Ala Val Asn Gln Asp Gly Ala Ser Asn Gly Leu Thr Ala Pro		
260	265	270
Asn Gly Pro Ser Gln Gln Arg Val Ile Arg Ala Ala Leu Ala Glu Ala		
275	280	285
Gly Leu Ser Pro Ser Asp Val Asp Ala Val Glu Gly His Gly Thr Gly		
290	295	300

Thr Thr Leu Gly Asp Pro Ile Glu Ala Gln Ala Leu Leu Ala Thr Tyr
 305 310 315 320

Gly Lys Gly Arg Asp Pro Glu Lys Pro Leu Trp Leu Gly Ser Val Lys
 325 330 335

Ser Asn Leu Gly His Thr Gln Ala Ala Ala Gly Val Ala Ser Val Ile
 340 345 350

Lys Met Val Gln Ala Leu Arg His Gly Val Leu Pro Pro Thr Leu His
 355 360 365

Val Asp Arg Pro Ser Thr Glu Val Asp Trp Ser Ala Gly Ala Val Ser
 370 375 380

Leu Leu Thr Glu Ala Arg Glu Trp Pro Arg Glu Gly Arg Pro Arg Arg
 385 390 395 400

Ala Gly Val Ser Ser Phe Gly Ile Ser Gly Thr Asn Ala His Leu Ile
 405 410 415

Leu Glu Glu Ala Pro Glu Glu Glu Pro Pro Val Ala Glu Ala Pro Ser
 420 425 430

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

Gln Ala Gly Arg Leu Ala Ala Phe Leu Glu Ala Ser Asp Glu Pro Leu
 450 455 460

Val Thr Val Ala Gly Ala Leu Ile Cys Gly Arg Ser Arg Phe Gly Asp
 465 470 475 480

Arg Ala Val Val Val Ala Gly Thr Arg Ala Glu Ala Thr Ala Gly Leu
 485 490 495

Ala Ala Leu Ala Arg Gly Glu Ser Ala Ala Asp Val Val Thr Gly Thr			
500	505	510	
Val Ala Ala Ser Gly Val Pro Gly Lys Leu Val Trp Val Phe Pro Gly			
515	520	525	
Gln Gly Ser Gln Trp Val Gly Met Gly Arg Glu Leu Leu Glu Ala Ser			
530	535	540	
Pro Val Phe Ala Ala Arg Ile Ala Glu Cys Ala Ala Ala Leu Glu Pro			
545	550	555	560
Trp Ile Asp Trp Ser Leu Leu Asp Val Leu Arg Gly Glu Gly Asp Leu			
565	570	575	
Asp Arg Val Asp Val Val Gln Pro Ala Ser Phe Ala Val Met Val Gly			
580	585	590	
Leu Ala Ala Val Trp Ser Ser Val Gly Val Val Pro Asp Ala Val Leu			
595	600	605	
Gly His Ser Gln Gly Glu Ile Ala Ala Ala Cys Val Ser Gly Ala Leu			
610	615	620	
Ser Leu Gln Asp Ala Ala Lys Val Val Ala Leu Arg Ser Gln Ala Ile			
625	630	635	640
Ala Ala Lys Leu Ala Gly Arg Gly Gly Met Ala Ser Val Ala Leu Ser			
645	650	655	
Glu Glu Asp Ala Val Ala Arg Leu Arg His Trp Ala Asp Arg Val Glu			
660	665	670	
Val Ala Ala Val Asn Ser Pro Ser Ser Val Val Ile Ala Gly Asp Ala			
675	680	685	
Glu Ala Leu Asp Gln Ala Leu Glu Ala Leu Thr Gly Gln Asp Ile Arg			

690

695

700

Val Arg Arg Val Ala Val Asp Tyr Ala Ser His Thr Arg His Val Glu
705 710 715 720

Asp Ile Gln Glu Pro Leu Ala Glu Ala Leu Ala Gly Ile Glu Ala His
725 730 735

Ala Pro Thr Leu Pro Phe Phe Ser Thr Leu Thr Gly Asp Trp Ile Arg
740 745 750

Glu Ala Gly Val Val Asp Gly Gly Tyr Trp Tyr Arg Asn Leu Arg Asn
755 760 765

Gln Val Gly Phe Gly Pro Ala Val Ala Glu Leu Leu Gly Leu Gly His
770 775 780

Arg Val Phe Val Glu Val Ser Ala His Pro Val Leu Val Gln Ala Ile
785 790 795 800

Ser Ala Ile Ala Asp Asp Thr Asp Ala Val Val Thr Gly Ser Leu Arg
805 810 815

Arg Glu Glu Gly Gly Leu Arg Arg Leu Leu Thr Ser Met Ala Glu Leu
820 825 830

Phe Val Arg Gly Val Asp Val Asp Trp Ala Thr Met Val Pro Pro Ala
835 840 845

Arg Val Asp Leu Pro Thr Tyr Ala Phe Asp His Gln His Tyr Trp Leu
850 855 860

Arg Tyr Val Glu Thr Ala Thr Asp Ala Ala Gly Pro Val Val Arg Leu
865 870 875 880

Pro Gln Thr Gly Gly Leu Val Phe Thr Thr Glu Trp Ser Leu Lys Ser
885 890 895

Gln Pro Trp Leu Ala Glu His Thr Leu Glu Asp Leu Val Val Val Pro

900

905

910

Gly Ala Ala Leu Val Glu Leu Ala Val Arg Ala Gly Asp Glu Ala Gly

915

920

925

Thr Pro Val Leu Asp Glu Leu Val Ile Glu Thr Pro Leu Val Val Pro

930

935

940

Glu Arg Gly Ala Ile Arg Val Gln Val Thr Val Ser Gly Pro Asp Asp

945

950

955

960

Gly Thr Arg Thr Leu Glu Val His Ser Gln Pro Glu Asp Ala Thr Asp

965

970

975

Glu Trp Thr Arg His Ala Thr Gly Thr Leu Ser Ala Thr Pro Asp Glu

980

985

990

Ser Ser Gly Phe Asp Phe Thr Ala Trp Pro Pro Pro Gly Ala Arg Gln

995

1000

1005

Leu Asp Gly Val Pro Ala Ile Trp Arg Ala Gly Asp Glu Ile Phe Ala

1010

1015

1020

Glu Val Ser Leu Pro Asp Asp Ala Asp Ala Glu Ala Phe Gly Ile His

1025

1030

1035

1040

Pro Ala Leu Leu Asp Ala Ala Leu His Pro Ala Leu Pro Gly Asp Asp

1045

1050

1055

Gly Leu Thr Gln Pro Met Glu Trp Arg Gly Leu Thr Leu His Ala Ala

1060

1065

1070

Gly Ala Ser Thr Leu Arg Val Arg Leu Val Pro Gly Gly Phe Leu Glu

1075

1080

1085

Ala Ala Asp Gly Ala Gly Ser Leu Val Val Thr Ala Lys Glu Val Ala
 1090 1095 1100

Leu Arg Pro Val Thr Ile Ala Arg Ser Arg Thr Thr Thr Arg Asp Ser
 1105 1110 1115 1120

Leu Phe Gln Leu Asn Trp Ile Glu Leu Pro Glu Ser Gly Val Val Ala
 1125 1130 1135

Ala Ala Asp Asp Thr Glu Val Leu Glu Val Pro Ala Gly Asp Ser Pro
 1140 1145 1150

Leu Ala Ala Thr Ser Arg Val Leu Glu Arg Leu Gln Thr Trp Leu Thr
 1155 1160 1165

Glu Pro Glu Ala Glu Gln Leu Val Val Val Thr Arg Gly Ala Val Pro
 1170 1175 1180

Ala Gly Asp Thr Pro Val Thr Asp Pro Ala Ala Ala Val Trp Gly
 1185 1190 1195 1200

Leu Val Arg Ser Ala Gln Ala Glu Asn Pro Asp Arg Ile Val Leu Leu
 1205 1210 1215

Asp Thr Asp Gly Glu Val Pro Leu Gly Ala Val Leu Ala Gly Gly Glu
 1220 1225 1230

Pro Gln Val Ala Val Arg Gly Thr Ala Leu Tyr Val Pro Arg Leu Ala
 1235 1240 1245

Arg Ala Asp Ala Ala Pro Val Ser Gly Leu His Gly Thr Val Leu Val
 1250 1255 1260

Ser Gly Ala Gly Val Leu Gly Glu Ile Val Ala Arg His Leu Val Thr
 1265 1270 1275 1280

Arg His Gly Val Arg Lys Leu Val Leu Ala Ser Arg Arg Gly Leu Asp

1285

1290

1295

Ala Asp Gly Ala Lys Asp Leu Val Thr Asp Leu Thr Gly Glu Gly Ala
 1300 1305 1310

Asp Val Ser Val Val Ala Cys Asp Leu Ala Asp Arg Asn Gln Val Ala
 1315 1320 1325

Ala Leu Leu Ala Asp His Arg Pro Ala Ser Val Ile His Thr Ala Gly
 1330 1335 1340

Val Leu Asp Asp Gly Val Ile Gly Thr Leu Thr Pro Glu Arg Leu Ala
 1345 1350 1355 1360

Lys Val Phe Ala Pro Lys Val Asp Ala Val Arg His Leu Asp Glu Leu
 1365 1370 1375

Thr Arg Asp Leu Asp Leu Asp Ala Phe Val Val Phe Ser Ser Gly Ser
 1380 1385 1390

Gly Val Phe Gly Ser Pro Gly Gln Gly Asn Tyr Ala Ala Ala Asn Ala
 1395 1400 1405

Phe Leu Asp Ala Ala Met Ala Ser Arg Arg Ala Ala Gly Leu Pro Gly
 1410 1415 1420

Leu Ser Leu Ala Trp Gly Leu Trp Glu Gln Ala Thr Gly Met Thr Ala
 1425 1430 1435 1440

His Leu Gly Gly Thr Asp Gln Ala Arg Met Ser Arg Gly Gly Val Arg
 1445 1450 1455

Pro Ile Thr Ala Glu Glu Gly Met Ala Leu Phe Asp Thr Ala Leu Gly
 1460 1465 1470

Ala Gln Pro Ala Leu Leu Val Pro Val Lys Leu Asp Leu Arg Glu Val
 1475 1480 1485

Arg Ala Gly Gly Ala Val	Pro His Leu Leu Arg	Gly Leu Val Arg Ala
1490	1495	1500
Gly Arg Arg Gln Ala Gln Ala Ala	Ser Thr Val Asp Asn Gln Leu Leu	
1505	1510	1515 1520
Gly Arg Leu Ala Gly Leu Gly Ala	Pro Glu Gln Glu Ala Leu Leu Val	
	1525	1530 1535
Asp Leu Val Arg Gly Gln Val Ala Ala	Val Leu Gly His Ala Gly Pro	
	1540	1545 1550
Asp Ala Val Arg Ala Asp Thr Ala Phe	Lys Asp Ala Gly Phe Asp Ser	
	1555	1560 1565
Leu Thr Ser Val Asp Leu Arg Asn Arg	Leu Arg Glu Ser Thr Gly Leu	
	1570	1575 1580
Lys Leu Pro Ala Thr Leu Ala Phe Asp	Tyr Pro Thr Pro Leu Val Leu	
1585	1590	1595 1600
Ala Arg His Leu Arg Asp Glu Leu Gly	Ala Gly Asp Asp Ala Leu Ser	
	1605	1610 1615
Val Val His Ala Arg Leu Glu Asp	Val Glu Ala Leu Leu Gly Gly Leu	
	1620	1625 1630
Arg Leu Asp Glu Ser Thr Lys Thr	Gly Leu Thr Leu Arg Leu Gln Gly	
	1635	1640 1645
Leu Val Ala Arg Cys Asn Gly Val Asn	Asp Gln Thr Gly Gly Glu Thr	
	1650	1655 1660
Leu Ala Asp Arg Leu Glu Ala Ala Ser	Ala Asp Glu Val Leu Asp Phe	
1665	1670	1675 1680

Ile Asp Glu Glu Leu Gly Leu Thr

1685

(2) Information für SEQ ID Nr: 8

(i) Sequenzeigenschaften:

(A) Länge: 3413 Aminosäuren
 (B) Typ: Aminosäure
 (C) Strangart: Einzelstrang
 (D) Topologie: linear

(ii) Molekültyp: Peptid

(xi) Sequenzbeschreibung: SEQ ID Nr: 8

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

Glu Leu His Ser Leu Arg Lys Gln Gly Ala Arg His Ala Asp Glu Pro
 20 25 30

Leu Ala Val Val Gly Met Ala Cys Arg Phe Pro Gly Gly Val Ser Ser
 35 40 45

Pro Glu Asp Leu Trp Gln Leu Val Ala Gly Gly Val Asp Ala Leu Ser
 50 55 60

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

Asp Pro Asp His Pro Gly Thr Ser Tyr Thr Ser Gln Gly Gly Phe Leu
 85 90 95

Arg Gly Ala Gly Leu Phe Asp Ala Gly Leu Phe Gly Ile Ser Pro Arg

100	105	110
Glu Ala Leu Val Met Asp Pro Gln Gln Arg Val Leu Leu Glu Thr Ser		
115	120	125
Trp Glu Ala Leu Glu Asp Ala Gly Val Asp Pro Leu Ser Leu Lys Gly		
130	135	140
Ser Asp Val Gly Val Phe Ser Gly Val Phe Thr Gln Gly Tyr Gly Ala		
145	150	155
Gly Ala Ile Thr Pro Asp Leu Glu Ala Phe Ala Gly Ile Gly Ala Ala		
165	170	175
Ser Ser Val Ala Ser Gly Arg Val Ser Tyr Val Phe Gly Leu Glu Gly		
180	185	190
Pro Ala Val Thr Ile Asp Thr Ala Cys Ser Ser Ser Leu Val Ala Ile		
195	200	205
His Leu Ala Ala Gln Ala Leu Arg Ala Gly Glu Cys Ser Met Ala Leu		
210	215	220
Ala Gly Gly Ala Thr Val Met Pro Thr Pro Gly Thr Phe Val Ala Phe		
225	230	235
Ser Arg Gln Arg Val Leu Ala Ala Asp Gly Arg Ser Lys Ala Phe Ser		
245	250	255
Ser Thr Ala Asp Gly Thr Gly Trp Ala Glu Gly Ala Gly Val Leu Val		
260	265	270
Leu Glu Arg Leu Ser Val Ala Gln Glu Arg Gly His Arg Ile Leu Ala		
275	280	285
Val Leu Arg Gly Ser Ala Val Asn Gln Asp Gly Ala Ser Asn Gly Leu		
290	295	300

Thr Ala Pro Asn Gly Pro Ser Gln Gln Arg Val Ile Arg Lys Ala Leu
305 310 315 320

Ala Gly Ala Gly Leu Val Ala Ser Asp Val Asp Val Val Glu Ala His
325 330 335

Gly Thr Gly Thr Ala Leu Gly Asp Pro Ile Glu Ala Gln Ala Leu Leu
340 345 350

Ala Thr Tyr Gly Gln Gly Arg Glu Arg Pro Leu Trp Leu Gly Ser Val
355 360 365

Lys Ser Asn Phe Gly His Thr Gln Ala Ala Ala Gly Val Ala Gly Val
370 375 380

Ile Lys Met Val Gln Ala Leu Arg His Gly Ala Met Pro Pro Thr Leu
385 390 395 400

His Val Ala Glu Pro Thr Pro Glu Val Asp Trp Ser Ala Gly Ala Val
405 410 415

Glu Leu Leu Thr Glu Pro Arg Glu Trp Pro Ala Gly Asp Arg Pro Arg
420 425 430

Arg Ala Gly Val Ser Ala Phe Gly Ile Ser Gly Thr Asn Ala His Leu
435 440 445

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

Phe Lys Gly Pro Val Pro Leu Val Val Ser Ala Gly Ser Pro Thr Ser
465 470 475 480

Leu Ala Ala Gln Ala Gly Arg Leu Ala Glu Val Leu Ala Ser Gly Gly
485 490 495

Val Ser Arg Ala Arg Leu Ala Ser Gly Leu Leu Ser Gly Arg Ala Leu
 500 505 510

Leu Gly Asp Arg Ala Val Val Val Ala Gly Thr Asp Glu Asp Ala Val
 515 520 525

Ala Gly Leu Arg Ala Leu Ala Arg Gly Asp Arg Ala Pro Gly Val Leu
 530 535 540

Thr Gly Ser Ala Lys His Gly Lys Val Val Tyr Val Phe Pro Gly Gln
 545 550 555 560

Gly Ser Gln Arg Leu Gly Met Gly Arg Glu Leu Tyr Asp Arg Tyr Pro
 565 570 575

Val Phe Ala Thr Ala Phe Asp Glu Ala Cys Glu Gln Leu Asp Val Cys
 580 585 590

Leu Ala Gly Arg Ala Gly His Arg Val Arg Asp Val Val Leu Gly Glu
 595 600 605

Val Pro Ala Glu Thr Gly Leu Leu Asn Gln Thr Val Phe Thr Gln Ala
 610 615 620

Gly Leu Phe Ala Val Glu Ser Ala Leu Phe Arg Leu Ala Glu Ser Trp
 625 630 635 640

Gly Val Arg Pro Asp Val Val Leu Gly His Ser Ile Gly Glu Ile Thr
 645 650 655

Ala Ala Tyr Ala Ala Gly Val Phe Ser Leu Pro Asp Ala Ala Arg Ile
 660 665 670

Val Ala Ala Arg Gly Arg Leu Met Gln Ala Leu Ala Pro Gly Gly Ala
 675 680 685

Met Val Ala Val Ala Ala Ser Glu Ala Glu Val Ala Glu Leu Leu Gly

690

695

700

Asp Gly Val Glu Leu Ala Ala Val Asn Gly Pro Ser Ala Val Val Leu
 705 710 715 720

Ser Gly Asp Ala Asp Ala Val Val Ala Ala Ala Ala Arg Met Arg Glu
 725 730 735

Arg Gly His Lys Thr Lys Gln Leu Lys Val Ser His Ala Phe His Ser
 740 745 750

Ala Arg Met Ala Pro Met Leu Ala Glu Phe Ala Ala Glu Leu Ala Gly
 755 760 765

Val Thr Trp Arg Glu Pro Glu Ile Pro Val Val Ser Asn Val Thr Gly
 770 775 780

Arg Phe Ala Glu Pro Gly Glu Leu Thr Glu Pro Gly Tyr Trp Ala Glu
 785 790 795 800

His Val Arg Arg Pro Val Arg Phe Ala Glu Gly Val Ala Ala Ala Thr
 805 810 815

Glu Ser Gly Gly Ser Leu Phe Val Glu Leu Gly Pro Gly Ala Ala Leu
 820 825 830

Thr Ala Leu Val Glu Glu Thr Ala Glu Val Thr Cys Val Ala Ala Leu
 835 840 845

Arg Asp Asp Arg Pro Glu Val Thr Ala Leu Ile Thr Ala Val Ala Glu
 850 855 860

Leu Phe Val Arg Gly Val Ala Val Asp Trp Pro Ala Leu Leu Pro Pro
 865 870 875 880

Val Thr Gly Phe Val Asp Leu Pro Lys Tyr Ala Phe Asp Gln Gln His
 885 890 895

Tyr Trp Leu Gln Pro Ala Ala Gln Ala Thr Asp Ala Ala Ser Leu Gly
 900 905 910

Gln Val Ala Ala Asp His Pro Leu Leu Gly Ala Val Val Arg Leu Pro
 915 920 925

Gln Ser Asp Gly Leu Val Phe Thr Ser Arg Leu Ser Leu Lys Ser His
 930 935 940

Pro Trp Leu Ala Asp His Val Ile Gly Gly Val Val Leu Val Ala Gly
 945 950 955 960

Thr Gly Leu Val Glu Leu Ala Val Arg Ala Gly Asp Glu Ala Gly Cys
 965 970 975

Pro Val Leu Glu Glu Leu Val Ile Glu Ala Pro Leu Val Val Pro Asp
 980 985 990

His Gly Gly Val Arg Ile Gln Val Val Val Gly Ala Pro Gly Glu Thr
 995 1000 1005

Gly Ser Arg Ala Val Glu Val Tyr Ser Leu Arg Glu Asp Ala Gly Ala
 1010 1015 1020

Glu Val Trp Ala Arg His Ala Thr Gly Phe Leu Ala Ala Thr Pro Ser
 1025 1030 1035 1040

Gln His Lys Pro Phe Asp Phe Thr Ala Trp Pro Pro Pro Gly Val Glu
 1045 1050 1055

Arg Val Asp Val Glu Asp Phe Tyr Asp Gly Leu Val Asp Arg Gly Tyr
 1060 1065 1070

Ala Tyr Gly Pro Ser Phe Arg Gly Leu Arg Ala Val Trp Arg Arg Gly
 1075 1080 1085

Asp Glu Val Phe Ala Glu Val Ala Leu Ala Glu Asp Asp Arg Ala Asp
 1090 1095 1100

Ala Ala Arg Phe Gly Ile His Pro Gly Leu Leu Asp Ala Ala Leu His
 1105 1110 1115 1120

Ala Gly Met Ala Gly Ala Thr Thr Thr Glu Glu Pro Gly Arg Pro Val
 1125 1130 1135

Leu Pro Phe Ala Trp Asn Gly Leu Val Leu His Ala Ala Gly Ala Ser
 1140 1145 1150

Ala Leu Arg Val Arg Leu Ala Pro Ser Gly Pro Asp Ala Leu Ser Val
 1155 1160 1165

Glu Ala Ala Asp Glu Ala Gly Gly Leu Val Val Thr Ala Asp Ser Leu
 1170 1175 1180

Val Ser Arg Pro Val Ser Ala Glu Gln Leu Gly Ala Ala Ala Asn His
 1185 1190 1195 1200

Asp Ala Leu Phe Arg Val Glu Trp Thr Glu Ile Ser Ser Ala Gly Asp
 1205 1210 1215

Val Pro Ala Asp His Val Glu Val Leu Glu Ala Val Gly Glu Asp Pro
 1220 1225 1230

Leu Glu Leu Thr Gly Arg Val Leu Glu Ala Val Gln Thr Trp Leu Ala
 1235 1240 1245

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

His Glu Val Thr Asp Pro Ala Gly Ala Ala Val Trp Gly Leu Ile Arg
 1265 1270 1275 1280

Ala Ala Gln Ala Glu Asn Pro Asp Arg Ile Val Leu Leu Asp Thr Asp

1285

1290

1295

Gly Glu Val Pro Leu Gly Arg Val Leu Ala Thr Gly Glu Pro Gln Thr
 1300 1305 1310

Ala Val Arg Gly Ala Thr Leu Phe Ala Pro Arg Leu Ala Arg Ala Glu
 1315 1320 1325

Ala Ala Glu Ala Pro Ala Val Thr Gly Gly Thr Val Leu Ile Ser Gly
 1330 1335 1340

Ala Gly Ser Leu Gly Ala Leu Thr Ala Arg His Leu Val Ala Arg His
 1345 1350 1355 1360

Gly Val Arg Arg Leu Val Leu Val Ser Arg Arg Gly Pro Asp Ala Asp
 1365 1370 1375

Gly Met Ala Glu Leu Thr Ala Glu Leu Ile Ala Gln Gly Ala Glu Val
 1380 1385 1390

Ala Val Val Ala Cys Asp Leu Ala Asp Arg Asp Gln Val Arg Val Leu
 1395 1400 1405

Leu Ala Glu His Arg Pro Asn Ala Val Val His Thr Ala Gly Val Leu
 1410 1415 1420

Asp Asp Gly Val Phe Glu Ser Leu Thr Arg Glu Arg Leu Ala Lys Val
 1425 1430 1435 1440

Phe Ala Pro Lys Val Thr Ala Ala Asn His Leu Asp Glu Leu Thr Arg
 1445 1450 1455

Glu Leu Asp Leu Arg Ala Phe Val Val Phe Ser Ser Ala Ser Gly Val
 1460 1465 1470

Phe Gly Ser Ala Gly Gln Gly Asn Tyr Ala Ala Ala Asn Ala Tyr Leu
 1475 1480 1485

Asp Ala Val Val Ala Asn Arg Arg Ala Ala Gly Leu Pro Gly Thr Ser
 1490 1495 1500

Leu Ala Trp Gly Leu Trp Glu Gln Thr Asp Gly Met Thr Ala His Leu
 1505 1510 1515 1520

Gly Asp Ala Asp Gln Ala Arg Ala Ser Arg Gly Gly Val Leu Ala Ile
 1525 1530 1535

Ser Pro Ala Glu Gly Met Glu Leu Phe Asp Ala Ala Pro Asp Gly Leu
 1540 1545 1550

Val Val Pro Val Lys Leu Asp Leu Arg Lys Thr Arg Ala Gly Gly Thr
 1555 1560 1565

Val Pro His Leu Leu Arg Gly Leu Val Arg Pro Gly Arg Gln Gln Ala
 1570 1575 1580

Arg Pro Ala Ser Thr Val Asp Asn Gly Leu Ala Gly Arg Leu Ala Gly
 1585 1590 1595 1600

Leu Ala Pro Ala Glu Gln Glu Ala Leu Leu Leu Asp Val Val Arg Thr
 1605 1610 1615

Gln Val Ala Leu Val Leu Gly His Ala Gly Pro Glu Ala Val Arg Ala
 1620 1625 1630

Asp Thr Ala Phe Lys Asp Thr Gly Phe Asp Ser Leu Thr Ser Val Glu
 1635 1640 1645

Leu Arg Asn Arg Leu Arg Glu Ala Ser Gly Leu Lys Leu Pro Ala Thr
 1650 1655 1660

Leu Val Phe Asp Tyr Pro Thr Pro Val Ala Leu Ala Arg Tyr Leu Arg
 1665 1670 1675 1680

Asp Glu Leu Gly Asp Thr Val Ala Thr Thr Pro Val Ala Thr Ala Ala	1685	1690	1695	
Ala Ala Asp Ala Gly Glu Pro Ile Ala Ile Val Gly Met Ala Cys Arg	1700	1705	1710	
Leu Pro Gly Gly Val Thr Asp Pro Glu Gly Leu Trp Arg Leu Val Arg	1715	1720	1725	
Asp Gly Leu Glu Gly Leu Ser Pro Phe Pro Glu Asp Arg Gly Trp Asp	1730	1735	1740	
Leu Glu Asn Leu Phe Asp Asp Asp Pro Asp Arg Ser Gly Thr Thr Tyr	1745	1750	1755	1760
Thr Ser Arg Gly Gly Phe Leu Asp Gly Ala Gly Leu Phe Asp Ala Gly	1765	1770	1775	
Phe Phe Gly Ile Ser Pro Arg Glu Ala Leu Ala Met Asp Pro Gln Gln	1780	1785	1790	
Arg Leu Leu Leu Glu Ala Ala Trp Glu Ala Leu Glu Gly Thr Gly Val	1795	1800	1805	
Asp Pro Gly Ser Leu Lys Gly Ala Asp Val Gly Val Phe Ala Gly Val	1810	1815	1820	
Ser Asn Gln Gly Tyr Gly Met Gly Ala Asp Pro Ala Glu Leu Ala Gly	1825	1830	1835	1840
Tyr Ala Ser Thr Ala Gly Ala Ser Ser Val Val Ser Gly Arg Val Ser	1845	1850	1855	
Tyr Val Phe Gly Phe Glu Gly Pro Ala Val Thr Ile Asp Thr Ala Cys	1860	1865	1870	
Ser Ser Ser Leu Val Ala Met His Leu Ala Gly Gln Ala Leu Arg Gln				

1875

1880

1885

Gly Glu Cys Ser Met Ala Leu Ala Gly Gly Val Thr Val Met Gly Thr
 1890 1895 1900

Pro Gly Thr Phe Val Glu Phe Ala Lys Gln Arg Gly Leu Ala Gly Asp
 1905 1910 1915 1920

Gly Arg Cys Lys Ala Tyr Ala Glu Gly Ala Asp Gly Thr Gly Trp Ala
 1925 1930 1935

Glu Gly Val Gly Val Val Val Leu Glu Arg Leu Ser Val Ala Arg Glu
 1940 1945 1950

Arg Gly His Arg Val Leu Ala Val Leu Arg Gly Ser Ala Val Asn Ser
 1955 1960 1965

Asp Gly Ala Ser Asn Gly Leu Thr Ala Pro Asn Gly Pro Ser Gln Gln
 1970 1975 1980

Arg Val Ile Arg Arg Ala Leu Ala Gly Ala Gly Leu Glu Pro Ser Asp
 1985 1990 1995 2000

Val Asp Ile Val Glu Gly His Gly Thr Gly Thr Ala Leu Gly Asp Pro
 2005 2010 2015

Ile Glu Ala Gln Ala Leu Leu Ala Thr Tyr Gly Lys Asp Arg Asp Pro
 2020 2025 2030

Glu Thr Pro Leu Trp Leu Gly Ser Val Lys Ser Asn Phe Gly His Thr
 2035 2040 2045

Gln Ser Ala Ala Gly Val Ala Gly Val Ile Lys Met Val Gln Ala Leu
 2050 2055 2060

Arg His Gly Val Met Pro Pro Thr Leu His Val Asp Arg Pro Thr Ser
 2065 2070 2075 2080

Gln Val Asp Trp Ser Ala Gly Ala Val Glu Val Leu Thr Glu Ala Arg			
2085	2090	2095	
Glu Trp Pro Arg Asn Gly Arg Pro Arg Arg Ala Gly Val Ser Ser Phe			
2100	2105	2110	
Gly Ile Ser Gly Thr Asn Ala His Leu Ile Ile Glu Glu Ala Pro Ala			
2115	2120	2125	
Glu Pro Gln Leu Ala Gly Pro Pro Pro Asp Gly Gly Val Val Pro Leu			
2130	2135	2140	
Val Val Ser Ala Arg Ser Pro Gly Ala Leu Ala Gly Gln Ala Arg Arg			
2145	2150	2155	2160
Leu Ala Thr Phe Leu Gly Asp Gly Pro Leu Ser Asp Val Ala Gly Ala			
2165	2170	2175	
Leu Thr Ser Arg Ala Leu Phe Gly Glu Arg Ala Val Val Val Ala Asp			
2180	2185	2190	
Ser Ala Glu Glu Ala Arg Ala Gly Leu Gly Ala Leu Ala Arg Gly Glu			
2195	2200	2205	
Asp Ala Pro Gly Leu Val Arg Gly Arg Val Pro Ala Ser Gly Leu Pro			
2210	2215	2220	
Gly Lys Leu Val Trp Val Phe Pro Gly Gln Gly Thr Gln Trp Val Gly			
2225	2230	2235	2240
Met Gly Arg Glu Leu Leu Glu Glu Ser Pro Val Phe Ala Glu Arg Ile			
2245	2250	2255	
Ala Glu Cys Ala Ala Ala Leu Glu Pro Trp Ile Gly Trp Ser Leu Phe			
2260	2265	2270	

Asp Val Leu Arg Gly Asp Gly Asp Leu Asp Arg Val Asp Val Leu Gln			
2275	2280	2285	
Pro Ala Cys Phe Ala Val Met Val Gly Leu Ala Ala Val Trp Ser Ser			
2290	2295	2300	
Ala Gly Val Val Pro Asp Ala Val Leu Gly His Ser Gln Gly Glu Ile			
2305	2310	2315	2320
Ala Ala Ala Cys Val Ser Gly Ala Leu Ser Leu Glu Asp Ala Ala Lys			
2325	2330	2335	
Val Val Ala Leu Arg Ser Gln Ala Ile Ala Ala Lys Leu Ser Gly Arg			
2340	2345	2350	
Gly Gly Met Ala Ser Val Ala Leu Gly Glu Ala Asp Val Val Ser Arg			
2355	2360	2365	
Leu Ala Asp Gly Val Glu Val Ala Ala Val Asn Gly Pro Ala Ser Val			
2370	2375	2380	
Val Ile Ala Gly Asp Ala Gln Ala Leu Asp Glu Thr Leu Glu Ala Leu			
2385	2390	2395	2400
Ser Gly Ala Gly Ile Arg Ala Arg Arg Val Ala Val Asp Tyr Ala Ser			
2405	2410	2415	
His Thr Arg His Val Glu Asp Ile Glu Asp Thr Leu Ala Glu Ala Leu			
2420	2425	2430	
Ala Gly Ile Asp Ala Arg Ala Pro Leu Val Pro Phe Leu Ser Thr Leu			
2435	2440	2445	
Thr Gly Glu Trp Ile Arg Asp Glu Gly Val Val Asp Gly Gly Tyr Trp			
2450	2455	2460	
Tyr Arg Asn Leu Arg Gly Arg Val Arg-Phe Gly Pro Ala Val Glu Ala			

2465	2470	2475	2480
Leu Leu Ala Gln Gly His Gly Val Phe Val Glu Leu Ser Ala His Pro			
2485	2490	2495	
Val Leu Val Gln Pro Ile Thr Glu Leu Thr Asp Glu Thr Ala Ala Val			
2500	2505	2510	
Val Thr Gly Ser Leu Arg Arg Asp Asp Gly Gly Leu Arg Arg Leu Leu			
2515	2520	2525	
Thr Ser Met Ala Glu Leu Phe Val Arg Gly Val Glu Val Asp Trp Thr			
2530	2535	2540	
Ser Leu Val Pro Pro Ala Arg Ala Asp Leu Pro Thr Tyr Ala Phe Asp			
2545	2550	2555	2560
His Glu His Tyr Trp Leu Arg Ala Ala Asp Thr Ala Ser Asp Ala Val			
2565	2570	2575	
Ser Leu Gly Leu Ala Gly Ala Asp His Pro Leu Leu Gly Ala Val Val			
2580	2585	2590	
Gln Leu Pro Gln Ser Asp Gly Leu Val Phe Thr Ser Arg Leu Ser Leu			
2595	2600	2605	
Arg Ser His Pro Trp Leu Ala Asp His Ala Val Arg Asp Val Val Ile			
2610	2615	2620	
Val Pro Gly Thr Gly Leu Val Glu Leu Ala Val Arg Ala Gly Asp Glu			
2625	2630	2635	2640
Ala Gly Cys Pro Val Leu Asp Glu Leu Val Ile Glu Ala Pro Leu Val			
2645	2650	2655	
Val Pro Arg Arg Gly Gly Val Arg Val Gln Val Ala Leu Gly Gly Pro			
2660	2665	2670	

Ala Asp Asp Gly Ser Arg Thr Val Asp Val Phe Ser Leu Arg Glu Asp
 2675 2680 2685

Ala Asp Ser Trp Leu Arg His Ala Thr Gly Val Leu Val Pro Glu Asn
 2690 2695 2700

Arg Pro Arg Gly Thr Ala Ala Phe Asp Phe Ala Ala Trp Pro Pro Pro
 2705 2710 2715 2720

Glu Ala Lys Pro Val Asp Leu Thr Gly Ala Tyr Asp Val Leu Ala Asp
 2725 2730 2735

Val Gly Tyr Gly Tyr Gly Pro Thr Phe Arg Ala Val Arg Ala Val Trp
 2740 2745 2750

Arg Arg Gly Ser Gly Asn Thr Thr Glu Thr Phe Ala Glu Ile Ala Leu
 2755 2760 2765

Pro Glu Asp Ala Arg Ala Glu Ala Gly Arg Phe Gly Ile His Pro Ala
 2770 2775 2780

Leu Leu Asp Ala Ala Leu His Ser Thr Met Val Ser Ala Ala Ala Asp
 2785 2790 2795 2800

Thr Glu Ser Tyr Gly Asp Glu Val Arg Leu Pro Phe Ala Trp Asn Gly
 2805 2810 2815

Leu Arg Leu His Ala Ala Gly Ala Ser Val Leu Arg Val Arg Val Ala
 2820 2825 2830

Lys Pro Glu Arg Asp Ser Leu Ser Leu Glu Ala Val Asp Glu Ser Gly
 2835 2840 2845

Gly Leu Val Val Thr Leu Asp Ser Leu Val Gly Arg Pro Val Ser Asn
 2850 2855 2860

Asp Gln Leu Thr Thr Ala Ala Gly Pro Ala Gly Ala Gly Ser Leu Tyr
 2865 2870 2875 2880

Arg Val Asp Trp Thr Pro Leu Ser Ser Val Asp Thr Ser Gly Arg Val
 2885 2890 2895

Pro Ser Trp Leu Pro Val Ala Thr Ala Glu Glu Val Ala Thr Leu Ala
 2900 2905 2910

Asp Asp Val Leu Thr Gly Ala Thr Glu Ala Pro Ala Val Ala Val Met
 2915 2920 2925

Glu Ala Val Ala Asp Glu Gly Ser Val Leu Ala Leu Thr Val Arg Val
 2930 2935 2940

Leu Asp Val Val Gln Cys Trp Leu Ala Gly Gly Gly Leu Glu Gly Thr
 2945 2950 2955 2960

Lys Leu Ala Ile Val Thr Arg Gly Ala Val Pro Ala Gly Asp Gly Val
 2965 2970 2975

Val His Asp Pro Ala Ala Ala Ala Val Trp Gly Leu Val Arg Ala Ala
 2980 2985 2990

Gln Ala Glu Asn Pro Asp Arg Ile Val Leu Leu Asp Val Glu Pro Glu
 2995 3000 3005

Ala Asp Val Pro Pro Leu Leu Gly Ser Val Leu Ala Asp Gly Glu Pro
 3010 3015 3020

Gln Val Ala Val Arg Gly Thr Thr Leu Ser Ile Pro Arg Leu Ala Arg
 3025 3030 3035 3040

Ala Ala Arg Pro Asp Pro Ala Ala Gly Phe Lys Thr Arg Gly Pro Val
 3045 3050 3055

Leu Val Thr Gly Gly Thr Gly Ser Leu Gly Gly Leu Val Ala Arg His

3060

3065

3070

Leu Val Glu Arg His Gly Val Arg Gln Leu Val Leu Ala Ser Arg Arg
 3075 3080 3085

Gly Leu Asp Ala Glu Gly Ala Lys Asp Leu Val Thr Asp Leu Thr Ala
 3090 3095 3100

Leu Gly Ala Asp Val Ala Val Ala Ala Cys Asp Val Ala Asp Arg Asp
 3105 3110 3115 3120

Gln Val Ala Ala Leu Leu Thr Glu His Arg Pro Ser Ala Val Val His
 3125 3130 3135

Thr Ala Gly Val Pro Asp Ala Gly Val Ile Gly Thr Val Thr Pro Asp
 3140 3145 3150

Arg Leu Ala Glu Val Phe Ala Pro Lys Val Thr Ala Ala Arg His Leu
 3155 3160 3165

Asp Glu Leu Thr Arg Asp Leu Asp Leu Asp Ser Phe Val Val Tyr Ser
 3170 3175 3180

Ser Val Ser Ala Val Phe Met Gly Ala Gly Ser Gly Ser Tyr Ala Ala
 3185 3190 3195 3200

Ala Asn Ala Tyr Leu Asp Gly Leu Met Ala His Arg Arg Ala Ala Gly
 3205 3210 3215

Leu Pro Gly Gln Ser Leu Ala Trp Gly Leu Trp Asp Gln Thr Thr Gly
 3220 3225 3230

Gly Met Ala Ala Gly Thr Asp Glu Ala Gly Arg Ala Arg Met Thr Arg
 3235 3240 3245

Arg Gly Gly Leu Val Ala Met Lys Pro Ala Ala Gly Leu Asp Leu Phe
 3250 3255 3260

Asp Ala Ala Ile Gly Ser Gly Glu Pro Leu Leu Val Pro Ala Gln Leu
 3265 3270 3275 3280

Asp Leu Arg Gly Leu Arg Ala Glu Ala Ala Gly Gly Thr Glu Val Pro
 3285 3290 3295

His Leu Leu Arg Gly Leu Val Arg Ala Gly Arg Gln Gln Ala Arg Ala
 3300 3305 3310

Ala Ser Thr Val Glu Glu Asn Trp Ala Gly Arg Leu Ala Gly Leu Glu
 3315 3320 3325

Pro Ala Glu Arg Gly Gln Val Leu Leu Glu Leu Val Arg Ala Gln Val
 3330 3335 3340

Ala Gly Val Leu Gly Tyr Arg Ala Ala His Gln Val Asp Pro Asp Gln
 3345 3350 3355 3360

Gly Leu Phe Glu Ile Gly Phe Asp Ser Leu Thr Ala Ile Glu Leu Arg
 3365 3370 3375

Asn Arg Leu Arg Ala Arg Thr Glu Arg Lys Ile Ser Pro Gly Val Val
 3380 3385 3390

Phe Asp His Pro Thr Pro Ala Leu Leu Ala Ala His Leu Asn Glu Leu
 3395 3400 3405

Leu Arg Lys Lys Val
 3410

(2) Information für SEQ ID Nr: 9

(i) Sequenzeigenschaften:

(A) Länge:	226 Aminosäuren
(B) Typ:	Aminosäure
(C) Strangart:	Einzelstrang
(D) Topologie:	linear

(ii) Molekültyp:

Peptid

(xi) Sequenzbeschreibung: SEQ ID Nr: 9

Met Ala Ile Pro Tyr Ser Ser Leu Ala Tyr Glu Leu Arg Asp Ala Val
 1 5 10 15

Asn Val Val Asp Leu Asp Glu Asp Asp Val Phe Val Thr Ser Ile Ala
 20 25 30

Glu Gly Gln Gly Gly Ala Cys Tyr His Leu Asn Arg Leu Phe His Arg
 35 40 45

Leu Leu Thr Glu Leu Gly Tyr Asp Val Thr Pro Leu Ala Gly Ser Thr
 50 55 60

Ala Glu Gly Arg Glu Thr Phe Gly Thr Asp Val Glu His Met Phe Asn
 65 70 75 80

Leu Val Thr Leu Asp Gly Ala Asp Trp Leu Val Asp Val Gly Tyr Pro
 85 90 95

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

Gln Tyr Gly Ser Gln Phe Arg Leu Val Glu Gln Glu Thr Gly Tyr Ala
 115 120 125

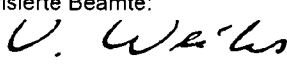
Leu Gln Arg Arg Gly Ala Val Thr Arg Trp Ser Val Val Tyr Thr Phe
 130 135 140

Thr Thr Gln Pro Arg Gln Trp Ser Asp Trp Lys Glu Leu Glu Asp Asn

145	150	155	160
Phe Arg Ala Leu Val Gly Asp Thr Thr Arg Thr Asp Thr Gln Glu Thr			
165	170	175	
Leu Cys Gly Arg Ala Phe Ala Asn Gly Gln Val Phe Leu Arg Gln Arg			
180	185	190	
Arg Tyr Leu Thr Val Glu Asn Gly Arg Glu Gln Val Arg Thr Ile Thr			
195	200	205	
Asp Asp Asp Glu Phe Arg Ala Leu Val Ser Arg Val Leu Ser Gly Asp			
210	215	220	
His Gly			
225			

I. Identifizierung des Mikroorganismus	
Identifizierungsbezeichnung vom Hinterleger: pRi7-3	Hinterlegungsnummer vergeben von der Internationalen Hinterlegungsbehörde: DSM 11114
II. Wissenschaftliche Beschreibung und/oder vorgeschlagene taxonomische Bezeichnung	
Der unter I. oben identifizierte Mikroorganismus wurde begleitet von ☒ einer wissenschaftlichen Beschreibung ☒ einer vorgeschlagenen taxonomischen Bezeichnung (Markierung mit einem Kreuz falls zutreffend)	
III. Erhalt und Annahme	
Die Internationale Hinterlegungsbehörde akzeptiert den unter I. oben identifizierten Mikroorganismus, der am 10.8.1996 erhalten wurde (Datum der Originalhinterlegung) (1)	
IV. Erhalt und Antrag auf Umwandlung	
Der oben unter I. identifizierte Mikroorganismus wurde von der Internationalen Hinterlegungsbehörde am (Datum der Originalhinterlegung) erhalten und ein Antrag, die Originalhinterlegung in eine Hinterlegung gemäß Budapester Vertrag umzuwandeln, wurde erhalten am (Datum des Umwandlungsantrags)	
V. Internationale Hinterlegungsbehörde	
Name: DSMZ-DEUTSCHE SAMMLUNG VON MIKROORGANISMEN UND ZELLKULTUREN GmbH Adresse: Mascheroder Weg 1b D-38124 Braunschweig	Unterschriften von Personen, die die Befugnis haben, die Internationale Hinterlegungsbehörde zu vertreten oder autorisierte Beamte:  Datum: 14.8.1996

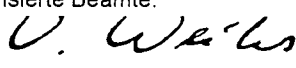
(1) Wenn Regel 6.4 (d) zutrifft, das Datum ist das Datum, an dem der Status der internationalen Hinterlegung erhalten wurde Formblatt DSMZ-BP/4 (eine Seite) 0196

I. Hinterleger	II. Identifizierung des Mikroorganismus
Name: Ciba-Geigy AG Adresse: CH-4002 Basel	Hinterlegungsnummer vergeben von der Internationalen Hinterlegungsbehörde DSM 11114 Datum der Hinterlegung oder des Transfers(1): 10.8.1996
III. Lebensfähigkeitsbestätigung	
Die Lebensfähigkeit des oben unter II identifizierten Mikroorganismus wurde getestet am 12.8.1996 (2): An diesem Datum war der Mikroorganismus ☒ lebensfähig ☐ nicht länger lebensfähig	
IV. Bedingungen, unter denen die Lebensfähigkeit getestet wurde (4)	
V. Internationale Hinterlegungsbehörde	
Name: DSMZ-DEUTSCHE SAMMLUNG VON MIKROORGANISMEN UND ZELLKULTUREN GmbH Adresse: Mascheroder Weg 1b D-38124 Braunschweig	Unterschriften von Personen, die die Befugnis haben, die Internationale Hinterlegungsbehörde zu vertreten oder autorisierte Beamte:  Datum: 14.8.1996

- (1) Angabe des Originalhinterlegungsdatums oder falls eine neue Hinterlegung oder ein Transfer durchgeführt wurde, das relevante jüngste Datum (Datum der neuen Hinterlegung oder Datum des Transfers)
- (2) In den Fällen, die sich auf Regel 10.2 (a) (ii) und (iii) beziehen, Angabe der jüngsten Lebensfähigkeitstests
- (3) Markierung mit einem Kreuz im zutreffenden Kästchen
- (4) Ausfüllen, falls die Information angefordert wurde oder falls die Ergebnisse des Tests negativ waren. Formblatt DSMZ-BP/9 (einzelne Seite) 0196

I. Identifizierung des Mikroorganismus	
Identifizierungsbezeichnung vom Hinterleger: pRi44-2	Hinterlegungsnummer vergeben von der Internationalen Hinterlegungsbehörde: DSM 11655
II. Wissenschaftliche Beschreibung und/oder vorgeschlagene taxonomische Bezeichnung	
Der unter I. oben identifizierte Mikroorganismus wurde begleitet von ☒ einer wissenschaftlichen Beschreibung ☒ einer vorgeschlagenen taxonomischen Bezeichnung (Markierung mit einem Kreuz falls zutreffend)	
III. Erhalt und Annahme	
Die Internationale Hinterlegungsbehörde akzeptiert den unter I. oben identifizierten Mikroorganismus, der am 14.7.1997 erhalten wurde (Datum der Originalhinterlegung) (1)	
IV. Erhalt und Antrag auf Umwandlung	
Der oben unter I. identifizierte Mikroorganismus wurde von der Internationalen Hinterlegungsbehörde am (Datum der Originalhinterlegung) erhalten und ein Antrag, die Originalhinterlegung in eine Hinterlegung gemäß Budapester Vertrag umzuwandeln, wurde erhalten am (Datum des Umwandlungsantrags)	
V. Internationale Hinterlegungsbehörde	
Name: DSMZ-DEUTSCHE SAMMLUNG VON MIKROORGANISMEN UND ZELLKULTUREN GmbH Adresse: Mascheroder Weg 1b D-38124 Braunschweig	Unterschriften von Personen, die die Befugnis haben, die Internationale Hinterlegungsbehörde zu vertreten oder autorisierte Beamte:  Datum: 15.7.1997

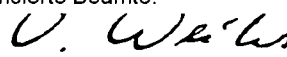
(1) Wenn Regel 6.4 (d) zutrifft, das Datum ist das Datum, an dem der Status der internationalen Hinterlegung erhalten wurde Formblatt DSMZ-BP/4 (eine Seite) 0196

I. Hinterleger	II. Identifizierung des Mikroorganismus
Name: Novartis AG Adresse: CH-4002 Basel	Hinterlegungsnummer vergeben von der Internationalen Hinterlegungsbehörde DSM 11655 Datum der Hinterlegung oder des Transfers(1): 14.7.1997
III. Lebensfähigkeitsbestätigung	
Die Lebensfähigkeit des oben unter II identifizierten Mikroorganismus wurde getestet am 14.7.1997 (2): An diesem Datum war der Mikroorganismus ☒ lebensfähig ☐ nicht länger lebensfähig	
IV. Bedingungen, unter denen die Lebensfähigkeit getestet wurde (4)	
V. Internationale Hinterlegungsbehörde	
Name: DSMZ-DEUTSCHE SAMMLUNG VON MIKROORGANISMEN UND ZELLKULTUREN GmbH Adresse: Mascheroder Weg 1b D-38124 Braunschweig	Unterschriften von Personen, die die Befugnis haben, die Internationale Hinterlegungsbehörde zu vertreten oder autorisierte Beamte:  Datum: 15.7.1997

- (1) Angabe des Originalhinterlegungsdatums oder falls eine neue Hinterlegung oder ein Transfer durchgeführt wurde, das relevante jüngste Datum (Datum der neuen Hinterlegung oder Datum des Transfers)
- (2) In den Fällen, die sich auf Regel 10.2 (a) (ii) und (iii) beziehen, Angabe der jüngsten Lebensfähigkeitstests
- (3) Markierung mit einem Kreuz im zutreffenden Kästchen
- (4) Ausfüllen, falls die Information angefordert wurde oder falls die Ergebnisse des Tests negativ waren. Formblatt DSMZ-BP/9 (einzelne Seite) 0196

I. Identifizierung des Mikroorganismus	
Identifizierungsbezeichnung vom Hinterleger: pNE95	Hinterlegungsnummer vergeben von der Internationalen Hinterlegungsbehörde: DSM 11656
II. Wissenschaftliche Beschreibung und/oder vorgeschlagene taxonomische Bezeichnung	
Der unter I. oben identifizierte Mikroorganismus wurde begleitet von ☒ einer wissenschaftlichen Beschreibung ☒ einer vorgeschlagenen taxonomischen Bezeichnung (Markierung mit einem Kreuz falls zutreffend)	
III. Erhalt und Annahme	
Die Internationale Hinterlegungsbehörde akzeptiert den unter I. oben identifizierten Mikroorganismus, der am 14.7.1997 erhalten wurde (Datum der Originalhinterlegung) (1)	
IV. Erhalt und Antrag auf Umwandlung	
Der oben unter I. identifizierte Mikroorganismus wurde von der Internationalen Hinterlegungsbehörde am (Datum der Originalhinterlegung) erhalten und ein Antrag, die Originalhinterlegung in eine Hinterlegung gemäß Budapester Vertrag umzuwandeln, wurde erhalten am (Datum des Umwandlungsantrags)	
V. Internationale Hinterlegungsbehörde	
Name: DSMZ-DEUTSCHE SAMMLUNG VON MIKROORGANISMEN UND ZELLKULTUREN GmbH Adresse: Mascheroder Weg 1b D-38124 Braunschweig	Unterschriften von Personen, die die Befugnis haben, die Internationale Hinterlegungsbehörde zu vertreten oder autorisierte Beamte:  Datum: 15.7.1997

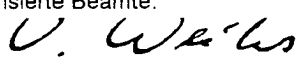
(1) Wenn Regel 6.4 (d) zutrifft, das Datum ist das Datum, an dem der Status der internationalen Hinterlegung erhalten wurde Formblatt DSMZ-BP/4 (eine Seite) 0196

I. Hinterleger	II. Identifizierung des Mikroorganismus
Name: Novartis AG Adresse: CH-4002 Basel	Hinterlegungsnummer vergeben von der Internationalen Hinterlegungsbehörde DSM 11656 Datum der Hinterlegung oder des Transfers(1): 14.7.1997
III. Lebensfähigkeitsbestätigung	
Die Lebensfähigkeit des oben unter II identifizierten Mikroorganismus wurde getestet am 14.7.1997 (2): An diesem Datum war der Mikroorganismus ☒ lebensfähig ☐ nicht länger lebensfähig	
IV. Bedingungen, unter denen die Lebensfähigkeit getestet wurde (4)	
V. Internationale Hinterlegungsbehörde	
Name: DSMZ-DEUTSCHE SAMMLUNG VON MIKROORGANISMEN UND ZELLKULTUREN GmbH Adresse: Mascheroder Weg 1b D-38124 Braunschweig	Unterschriften von Personen, die die Befugnis haben, die Internationale Hinterlegungsbehörde zu vertreten oder autorisierte Beamte:  Datum: 15.7.1997

- (1) Angabe des Originalhinterlegungsdatums oder falls eine neue Hinterlegung oder ein Transfer durchgeführt wurde, das relevante jüngste Datum (Datum der neuen Hinterlegung oder Datum des Transfers)
- (2) In den Fällen, die sich auf Regel 10.2 (a) (ii) und (iii) beziehen, Angabe der jüngsten Lebensfähigkeitstests
- (3) Markierung mit einem Kreuz im zutreffenden Kästchen
- (4) Ausfüllen, falls die Information angefordert wurde oder falls die Ergebnisse des Tests negativ waren. Formblatt DSMZ-BP/9 (einzelne Seite) 0196

I. Identifizierung des Mikroorganismus	
Identifizierungsbezeichnung vom Hinterleger: pNE112	Hinterlegungsnummer vergeben von der Internationalen Hinterlegungsbehörde: DSM 11657
II. Wissenschaftliche Beschreibung und/oder vorgeschlagene taxonomische Bezeichnung	
Der unter I. oben identifizierte Mikroorganismus wurde begleitet von ☒ einer wissenschaftlichen Beschreibung ☒ einer vorgeschlagenen taxonomischen Bezeichnung (Markierung mit einem Kreuz falls zutreffend)	
III. Erhalt und Annahme	
Die Internationale Hinterlegungsbehörde akzeptiert den unter I. oben identifizierten Mikroorganismus, der am 14.7.1997 erhalten wurde (Datum der Originalhinterlegung) (1)	
IV. Erhalt und Antrag auf Umwandlung	
Der oben unter I. identifizierte Mikroorganismus wurde von der Internationalen Hinterlegungsbehörde am (Datum der Originalhinterlegung) erhalten und ein Antrag, die Originalhinterlegung in eine Hinterlegung gemäß Budapester Vertrag umzuwandeln, wurde erhalten am (Datum des Umwandlungsantrags)	
V. Internationale Hinterlegungsbehörde	
Name: DSMZ-DEUTSCHE SAMMLUNG VON MIKROORGANISMEN UND ZELLKULTUREN GmbH Adresse: Mascheroder Weg 1b D-38124 Braunschweig	Unterschriften von Personen, die die Befugnis haben, die Internationale Hinterlegungsbehörde zu vertreten oder autorisierte Beamte:  Datum: 15.7.1997

(1) Wenn Regel 6.4 (d) zutrifft, das Datum ist das Datum, an dem der Status der internationalen Hinterlegung erhalten wurde Formblatt DSMZ-BP/4 (eine Seite) 0196

I. Hinterleger	II. Identifizierung des Mikroorganismus
Name: Novartis AG Adresse: CH-4002 Basel	Hinterlegungsnummer vergeben von der Internationalen Hinterlegungsbehörde DSM 11657 Datum der Hinterlegung oder des Transfers(1): 14.7.1997
III. Lebensfähigkeitsbestätigung	
Die Lebensfähigkeit des oben unter II identifizierten Mikroorganismus wurde getestet am 14.7.1997 (2): An diesem Datum war der Mikroorganismus ☒ lebensfähig ☐ nicht länger lebensfähig	
IV. Bedingungen, unter denen die Lebensfähigkeit getestet wurde (4)	
V. Internationale Hinterlegungsbehörde	
Name: DSMZ-DEUTSCHE SAMMLUNG VON MIKROORGANISMEN UND ZELLKULTUREN GmbH Adresse: Mascheroder Weg 1b D-38124 Braunschweig	Unterschriften von Personen, die die Befugnis haben, die internationale Hinterlegungsbehörde zu vertreten oder autorisierte Beamte:  Datum: 15.7.1997

- (1) Angabe des Originalhinterlegungsdatums oder falls eine neue Hinterlegung oder ein Transfer durchgeführt wurde, das relevante jüngste Datum (Datum der neuen Hinterlegung oder Datum des Transfers)
- (2) In den Fällen, die sich auf Regel 10.2 (a) (ii) und (iii) beziehen, Angabe der jüngsten Lebensfähigkeitstests
- (3) Markierung mit einem Kreuz im zutreffenden Kästchen
- (4) Ausfüllen, falls die Information angefordert wurde oder falls die Ergebnisse des Tests negativ waren. Formblatt DSMZ-BP/9 (einzelne Seite) 0196

Patentansprüche

1. Isolierte DNA, die entweder ist

- (i) das für die Rifamycinbiosynthese verantwortliche Gencluster in *Amycolatopsis mediterranei*, das die SEQ ID Nr. 3 umfasst,
- (ii) ein DNA Teil, der (i) für eine Polyketidsynthese oder eine enzymatisch aktive Domäne hiervon kodiert,
- (iii) ein DNA Teil der SEQ ID Nr. 3, der zumindest 15 aufeinanderfolgende Nukleotide hiervon aufweist und der als Hybridisierungssonde innerhalb einer Genbank des Rifamycin-bildenden Organismus zum Auffinden von Bestandteilen des entsprechenden Genclusters verwendet werden kann,
- (iv) eine DNA, die zumindest 70 % Identität zur DNA von (ii) aufweist und die für ein Polypeptid kodiert, das dieselbe Polyketidsyntheseaktivität aufweist, wie sie von der DNA von (ii) kodiert wird.

2. Isolierte DNA nach Anspruch 1, worin die DNA entweder

- (i) eine Nukleotidsequenz umfasst, die aus der Gruppe ausgewählt ist, welche besteht aus
 - ORF A, der aus einem Fragment von SEQ ID Nr. 3 besteht und von den Nukleotidpositionen 1825 bis 15543 reicht,
 - ORF B, der aus einem Fragment von SEQ ID Nr. 3 besteht und von den Nukleotidpositionen 15550 bis 30759

reicht,

– ORF C, der aus einem Fragment von SEQ ID Nr. 3 besteht und von den Nukleotidpositionen 30895 bis 36060 reicht,

– ORF D, der aus einem Fragment von SEQ ID Nr. 3 besteht und von den Nukleotidpositionen 36259 bis 41325 reicht,

– ORF E, der aus einem Fragment von SEQ ID Nr. 3 besteht und von den Nukleotidpositionen 41373 bis 51614 reicht, und

– ORF F, der aus einem Fragment von SEQ ID Nr. 3 besteht und von den Nukleotidpositionen 51713 bis 52393 reicht,

oder

(ii) für ein oder mehrere der Proteine oder Polypeptide der SEQ ID Nr. 4 bis 9 kodiert.

3. Verfahren zur Identifizierung, Isolierung und Klonierung einer isolierten DNA nach Anspruch 1 aus einem Organismus, der Rifamycin synthetisiert, wobei das Verfahren folgende Schritte umfasst:

– Erstellen einer Genbank aus dem Organismus, der Rifamycin synthetisiert,

– Screenen dieser Genbank mit Hilfe der DNA von Anspruch 1,

– Isolierung der als positiv identifizierten Klone.

4. Verwendung der DNA nach Anspruch 1, zur Herstellung von Ansamycinen oder Vorläufern hiervon, einschließlich derer, worin die aliphatische Brücke nur an einem Ende des aromatischen Kerns gebunden ist.

5. Verwendung nach Anspruch 4, worin das Ansamycin Rifamycin ist.

6. Verwendung der DNA nach Anspruch 1 zur Inaktivierung oder Modifizierung von Genen der Ansamycinbiosynthese.

7. Verwendung nach Anspruch 6, worin das Ansamycin Rifamycin ist.

8. Verwendung der DNA nach Anspruch 1 zur Konstruktion mutierter Actinomycetes Stämme, aus denen das natürliche Rifamycin- oder Ansamycinbiosynthesegencluster im Chromosom teilweise oder vollständig deletiert wurde.

9. Verwendung der DNA nach Anspruch 1 zur Zusammenstellung einer Genbank aus Polyketidsynthasen.

10. Polyketidsynthase aus *Amocolatopsis mediterranei*, die bei der Rifamycinsynthese oder einer enzymatisch aktiven Domäne hiervon beteiligt ist, worin die Polyketidsynthase durch eine DNA nach Anspruch 1 kodiert wird.

11. Verwendung der Polyketidsynthase nach Anspruch 10 zur Synthese von Ansamycinen.

12. Verwendung nach Anspruch 11 zur Synthese einer Genbank von Ansamycinen.

13. Hybridvektor, der die DNA nach Anspruch 1 umfasst.

14. Hybridvektor, der einen Expressionsvektor umfasst, welcher die DNA nach Anspruch 1 enthält.

15. Wirtsorganismus, der den Hybridvektor nach Anspruch 14 umfasst.

16. Hybridisierungssonde, die die DNA nach Anspruch 1 umfasst.

17. Verwendung der Hybridisierungssonde nach Anspruch 16 zur Identifizierung von DNA Fragmenten, die bei der Biosynthese von Ansamycinen beteiligt sind.

Es folgt kein Blatt Zeichnungen