

42/017025



PCT
WELTORGANISATION FÜR GEISTIGES EIGENTUM
Internationales Büro
INTERNATIONALE ANMELDUNG VERÖFFENTLICHT NACH DEM VERTRAG ÜBER DIE
INTERNATIONALE ZUSAMMENARBEIT AUF DEM GEBIET DES PATENTWESENS (PCT)

(51) Internationale Patentklassifikation ⁷ : C12N 15/00	A2	(11) Internationale Veröffentlichungsnummer: WO 00/26355 (43) Internationales Veröffentlichungsdatum: 11. Mai 2000 (11.05.00)
(21) Internationales Aktenzeichen: PCT/EP99/07952 (22) Internationales Anmeldedatum: 20. Oktober 1999 (20.10.99) (30) Prioritätsdaten: 198 50 242.7 31. Oktober 1998 (31.10.98) DE (71) Anmelder (für alle Bestimmungsstaaten ausser US): HAARMANN & REIMER GMBH [DE/DE]; D-37601 Holzmin-den (DE). (72) Erfinder; und (75) Erfinder/Anmelder (nur für US): RABENHORST, Jürgen [DE/DE]; Weserblick 13, D-37671 Höxter (DE). STEINBÜCHEL, Alexander [DE/DE]; Rönne-nthal 27, D-48341 Altenberge (DE). PRIEFERT, Horst [DE/DE]; Potthoffskamp 5, D-48291 Telgte (DE). OVERHAGE, Jörg [DE/DE]; Rotherbachstrasse 42 a, D-59132 Bergkamen (DE). (74) Anwalt: MANN, Volker, Bayer Aktiengesellschaft, D-51368 Leverkusen (DE). (81) Bestimmungsstaaten: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO Patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), eurasisches Patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), europäisches Patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI Patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG). Veröffentlicht <i>Ohne internationalen Recherchenbericht und erneut zu veröffentlichen nach Erhalt des Berichts.</i>		
(54) Title: CONSTRUCTION OF PRODUCTION STRAINS FOR PRODUCING SUBSTITUTED PHENOLS BY SPECIFICALLY INACTIVATING GENES OF THE EUGENOL AND FERULIC ACID CATABOLISM (54) Bezeichnung: KONSTRUKTION VON PRODUKTIONSSTÄMMEN FÜR DIE HERSTELLUNG VON SUBSTITUIERTEN PHENOLEN DURCH GEZIELTE INAKTIVIERUNGEN VON GENEN DES EUGENOL- UND FERULASÄURE-KATABOLISMUS (57) Abstract The invention relates to a transformed and/or mutagenated unicellular or multicellular organism which is characterized in that enzymes of the eugenol and/or ferulic acid catabolism are deactivated in such a manner that the intermediates coniferyl alcohol, coniferyl aldehyde, ferulic acid, vanillin and/or vanillic acid are accumulated. (57) Zusammenfassung Die vorliegende Erfindung betrifft einen transformierten und/oder mutagenisierten ein- oder mehrzelligen Organismus, der dadurch gekennzeichnet ist, daß Enzyme des Eugenol- und/oder Ferulasäure-Katabolismus derart inaktiviert sind, daß eine Akkumulation der Intermediate Coniferylalkohol, Coniferylaldehyd, Ferulasäure, Vanillin und/oder Vanillinsäure erfolgt.		

LEDIGLICH ZUR INFORMATION

Codes zur Identifizierung von PCT-Vertragsstaaten auf den Kopfbögen der Schriften, die internationale Anmeldungen gemäss dem PCT veröffentlichen.

AL	Albanien	ES	Spanien	LS	Lesotho	SI	Slowenien
AM	Armenien	FI	Finnland	LT	Litauen	SK	Slowakei
AT	Österreich	FR	Frankreich	LU	Luxemburg	SN	Senegal
AU	Australien	GA	Gabun	LV	Lettland	SZ	Swasiland
AZ	Aserbaidschan	GB	Vereinigtes Königreich	MC	Monaco	TD	Tschad
BA	Bosnien-Herzegowina	GE	Georgien	MD	Republik Moldau	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagaskar	TJ	Tadschikistan
BE	Belgien	GN	Guinea	MK	Die ehemalige jugoslawische Republik Mazedonien	TM	Turkmenistan
BF	Burkina Faso	GR	Griechenland	ML	Mali	TR	Türkei
BG	Bulgarien	HU	Ungarn	MN	Mongolei	TT	Trinidad und Tobago
BJ	Benin	IE	Irland	MR	Mauretanien	UA	Ukraine
BR	Brasilien	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Island	MX	Mexiko	US	Vereinigte Staaten von Amerika
CA	Kanada	IT	Italien	NE	Niger	UZ	Usbekistan
CF	Zentralafrikanische Republik	JP	Japan	NL	Niederlande	VN	Vietnam
CG	Kongo	KE	Kenia	NO	Norwegen	YU	Jugoslawien
CH	Schweiz	KG	Kirgisistan	NZ	Neuseeland	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Demokratische Volksrepublik Korea	PL	Polen		
CM	Kamerun	KR	Republik Korea	PT	Portugal		
CN	China	KZ	Kasachstan	RO	Rumänien		
CU	Kuba	LC	St. Lucia	RU	Russische Föderation		
CZ	Tschechische Republik	LI	Liechtenstein	SD	Sudan		
DE	Deutschland	LK	Sri Lanka	SE	Schweden		
DK	Dänemark	LR	Liberia	SG	Singapur		
EE	Estland						

Konstruktion von Produktionsstämmen für die Herstellung von substituierten Phenolen durch gezielte Inaktivierungen von Genen des Eugenol- und Ferulasäure-Katabolismus

5

Die vorliegende Erfindung betrifft die Konstruktion von Produktionsstämmen und ein Verfahren für die Herstellung substituierter Methoxyphenole, insbesondere Vanillin.

10

Die DE-A 4 227 076 (Verfahren zur Herstellung substituierter Methoxyphenole und dafür geeigneter Mikroorganismus) beschreibt die Herstellung substituierter Methoxyphenole mit einer neuen *Pseudomonas* sp.. Ausgangsmaterial ist hier Eugenol und die Produkte sind Ferulasäure, Vanillinsäure, Coniferylalkohol und Coniferylaldehyd.

15

Ebenfalls 1995 erscheint ein umfangreiches Review über die Biotransformationsmöglichkeiten mit Ferulasäure von Rosazza et al. (Biocatalytic transformation of ferulic acid: an abundant aromatic natural product; J. Ind. Microbiol. 15:457-471).

20

Die Gene und Enzyme zur Synthese von Coniferylalkohol, Coniferylaldehyd, Ferulasäure, Vanillin und Vanillinsäure aus *Pseudomonas* sp. wurden in EP-A 0 845 532 beschrieben.

25

Die Enzyme zur Umsetzung von *trans*-Ferulasäure zu *trans*-Feruloyl-SCoA Ester und weiter zum Vanillin, sowie das Gen für die Spaltung des Esters wurden vom Institute of Food Research, Norwich, GB, in WO 97/35999 beschrieben. 1998 erscheint der Inhalt des Patents auch als wissenschaftliche Publikationen (Gasson et al. 1998. Metabolism of ferulic acid to vanillin. J. Biol. Chem. 273:4163-4170; Narbad and Gasson 1998. Metabolism of ferulic acid via vanillin using a novel CoA-dependent pathway in a newly isolated strain of *Pseudomonas fluorescens*. Microbiology 144:1397 - 1405).

30

Die DE-A 195 32 317 beschreibt die fermentative Gewinnung von Vanillin aus Ferulasäure mit *Amycolatopsis* sp. in hohen Ausbeuten.

- 5 Die bekannten Verfahren haben den Nachteil, daß entweder nur sehr geringe Ausbeuten an Vanillin erzielt werden, oder von relativ teuren Edukten ausgegangen wird. Bei dem letztgenannten Verfahren (DE-A 195 32 317) werden zwar hohe Ausbeuten erzielt, jedoch bedingt der Einsatz von *Pseudomonas* sp. HR199 und *Amycolatopsis* sp. HR167 für die Biotransformation von Eugenol zu Vanillin eine zweistufige
10 Fermentationsführung und somit einen erheblichen Kosten- und Zeitaufwand.

Aufgabe der vorliegenden Erfindung ist es daher, Organismen zu konstruieren, die in der Lage sind den preiswerten Rohstoff Eugenol in einem einstufigen Prozeß zu Vanillin umzusetzen.

- 15 Diese Aufgabe wird durch die Konstruktion von Produktionsstämmen ein- oder mehrzelliger Organismen gelöst, die dadurch gekennzeichnet sind, daß Enzyme des Eugenol- und/oder Ferulasäure-Katabolismus derart inaktiviert sind, daß eine Akkumulation der Intermediate Coniferylalkohol, Coniferylaldehyd, Ferulasäure, Vanillin und/oder Vanillinsäure erfolgt.
20

- Der Produktionsstamm kann einzellig oder mehrzellig sein. Demgemäß können Gegenstand der Erfindung Mikroorganismen, Pflanzen oder Tiere sein. Darüber hinaus können auch Extrakte die aus dem Produktionsstamm gewonnen werden zum
25 Einsatz kommen. Erfindungsgemäß werden vorzugsweise einzellige Organismen eingesetzt. Hierbei kann es sich um Mikroorganismen, tierische oder pflanzliche Zellen handeln. Besonders bevorzugt ist erfindungsgemäß der Einsatz von Pilzen und Bakterien. Höchst bevorzugt sind Bakterienarten. Unter den Bakterien können insbesondere *Rhodococcus*-, *Pseudomonas*- und *Escherichia*-Arten nach Veränderung des
30 Eugenol- und/oder Ferulasäure-Katabolismus zum Einsatz kommen.

- Die Gewinnung der erfindungsgemäß einsetzbaren Organismen kann im einfachsten Fall mittels bekannter, konventioneller mikrobiologischer Methoden erfolgen. So kann die Enzymaktivität der am Eugenol- und/oder Ferulasäure-Katabolismus beteiligten Proteine durch den Einsatz von Enzym-Hemmstoffen verändert werden.
- 5 Darüber hinaus kann die Enzymaktivität der am Eugenol- und/oder Ferulasäure-Katabolismus beteiligten Proteine durch Mutation der für diese Proteine kodierenden Gene verändert werden. Derartige Mutationen können nach klassischen Methoden ungerichtet erzeugt werden, wie beispielsweise durch UV-Bestrahlung oder mutationsauslösende Chemikalien.
- 10 Ebenso sind gentechnische Methoden zur Gewinnung der erfindungsgemäßen Organismen geeignet, wie Deletionen, Insertionen und/oder Nukleotid-Austausche. So können beispielsweise die Gene der Organismen mit Hilfe von anderen DNA-Elementen (Ω -Elemente) inaktiviert werden. Ebenso können mittels geeigneter Vektoren
- 15 Austausche der intakten Gene gegen veränderte und/oder inaktivierte Gen-Strukturen durchgeführt werden. Die zu inaktivierenden Gene und die für die Inaktivierung eingesetzten DNA-Elemente können dabei durch klassische Klonierungstechniken oder durch Polymerase-Kettenreaktionen (PCR) gewonnen werden.
- 20 In einer möglichen Ausgestaltung der Erfindung kann beispielsweise der Eugenol- sowie der Ferulasäure-Katabolismus durch Ω -Element-Insertion oder Einführen von Deletionen in entsprechende Gene verändert werden. Hierbei können die Funktionen der Gene, die für Dehydrogenasen, Synthetasen, Hydratasen-Aldolasen, Thiolasen oder Demethylasen kodieren, mittels der oben genannten gentechnischen Methoden
- 25 inaktiviert werden, so daß die Erzeugung der betreffenden Enzyme blockiert ist. Vorzugsweise handelt es sich um die Gene, die für Coniferylalkohol-Dehydrogenasen, Coniferylaldehyd-Dehydrogenasen, Ferulasäure-CoA-Synthetasen, Enoyl-CoA-Hydratasen-Aldolasen, beta-Ketothiolasen, Vanillin-Dehydrogenasen oder Vanillinsäure-Demethylasen kodieren. Ganz besonders bevorzugt sind Gene, die die in der
- 30 EP-A 0845532 angegebenen Aminosäuresequenzen kodieren und/oder deren Allelvariationen kodierenden Nukleotidsequenzen.

Gegenstand der Erfindung sind demgemäß auch Gen-Strukturen zur Herstellung transformierter Organismen und Mutanten.

5 Vorzugsweise werden Gen-Strukturen zur Gewinnung der Organismen und Mutanten eingesetzt, bei denen die für Dehydrogenasen, Synthetasen, Hydratasen-Aldolasen, Thiolasen oder Demethylasen kodierenden Nukleotidsequenzen inaktiviert sind. Besonders bevorzugt sind Gen-Strukturen, bei denen die für Coniferylalkohol-Dehydrogenasen, Coniferylaldehyd-Dehydrogenasen, Ferulasäure-CoA-Synthetasen,
10 Enoyl-CoA-Hydratasen-Aldolasen, beta-Ketothiolasen, Vanillin-Dehydrogenasen oder Vanillinsäure-Demethylasen kodierenden Nukleotidsequenzen inaktiviert sind. Ganz besonders bevorzugt sind Gen-Strukturen, die die in den Figuren 1a bis 1r angegebenen Strukturen mit den in den Figuren 2a bis 2r wiedergegebenen Nukleotidsequenzen und/oder deren Allelvariationen kodierenden Nukleotidsequenzen aufweisen. Besonders bevorzugt sind hierbei Nukleotidsequenzen von 1 bis 18.

Die Erfindung schließt auch die Teilsequenzen dieser Gen-Strukturen sowie funktionelle Äquivalente ein. Unter funktionellen Äquivalenten sind solche Derivate der DNA zu verstehen, bei denen einzelne Nukleobasen ausgetauscht worden sind
20 (Wobbelaustausche), ohne die Funktion zu ändern. Auch auf Proteinebene können Aminosäuren ausgetauscht werden, ohne daß eine Veränderung der Funktion die Folge ist.

Den Gen-Strukturen können ein oder mehrere DNA-Sequenzen vor- und/oder nachgeschaltet sein. Durch Klonierung der Gen-Strukturen sind Plasmide bzw. Vektoren
25 erhältlich, die zur Transformation und/oder Transfektion eines Organismus und/oder zur konjugativen Übertragung in einen Organismus geeignet sind.

Gegenstand der Erfindung sind ferner Plasmide und/oder Vektoren zur Herstellung
30 der erfindungsgemäßen transformierten Organismen und Mutanten. Diese enthalten demgemäß die beschriebenen Gen-Strukturen. Die vorliegende Erfindung betrifft

demgemäß auch Organismen, die die genannten Plasmide und/oder Vektoren enthalten.

Die Art der Plasmide und/oder Vektoren hängt von deren Einsatzzweck ab. Um z. B. die intakten Gene des Eugenol- und/oder Ferulasäure-Katabolismus in *Pseudomonaden* gegen die durch Omega-Elemente inaktivierten Gene austauschen zu können, benötigt man Vektoren, die einerseits in *Pseudomonaden* übertragen werden können (konjugativ übertragbare Plasmide), andererseits dort jedoch nicht repliziert werden können und somit in *Pseudomonaden* instabil sind (sogenannte Suizid-Plasmide). DNA-Abschnitte, die mit Hilfe eines solchen Plasmidsystems in *Pseudomonaden* übertragen werden, können nur erhalten bleiben, wenn sie durch homologe Rekombination in das Genom der Bakterienzelle integriert werden.

Die beschriebenen Gen-Strukturen, Vektoren und Plasmide können zur Herstellung verschiedener transformierter Organismen oder Mutanten verwendet werden. Mittels der genannten Gen-Strukturen können intakte Nukleinsäuresequenzen gegen veränderte und/oder inaktivierte Gen-Strukturen ausgetauscht werden. In den durch Transformation oder Transfektion oder Konjugation erhältlichen Zellen erfolgt durch homologe Rekombination ein Austausch des intakten Gens gegen die veränderte und/oder inaktivierte Gen-Struktur, wodurch die resultierenden Zellen nur noch über die veränderte und/oder inaktivierte Gen-Struktur im Genom verfügen. So können erfindungsgemäß vorzugsweise Gene derart verändert und/oder inaktiviert werden, daß die betreffenden Organismen Coniferylalkohol, Coniferylaldehyd, Ferulasäure, Vanillin und/oder Vanillinsäure zu erzeugen vermögen.

Erfindungsgemäß derart konstruierte Produktionsstämme sind beispielsweise Mutanten des Stammes *Pseudomonas* sp. HR199 (DSM 7063) der in der DE-A 4 227 076 und der EP-A 0845532 genau beschrieben wurde, wobei sich unter anderem die entsprechenden Genstrukturen aus den Figuren 1a bis 1r in Verbindung mit den Figuren 2a bis 2r ergeben:

1. *Pseudomonas* sp. HR199*calA*ΩKm, enthaltend das durch ΩKm inaktivierte *calA*-Gen an Stelle des intakten *calA*-Gens kodierend für Coniferylalkohol-Dehydrogenase (Fig. 1a; Fig. 2a).
2. *Pseudomonas* sp. HR199*calA*ΩGm, enthaltend das durch ΩGm inaktivierte *calA*-Gen an Stelle des intakten *calA*-Gens kodierend für Coniferylalkohol-Dehydrogenase (Fig. 1b; Fig. 2b).
3. *Pseudomonas* sp. HR199*calA*Δ, enthaltend das durch Deletion inaktivierte *calA*-Gen an Stelle des intakten *calA*-Gens kodierend für Coniferylalkohol-Dehydrogenase (Fig. 1c; Fig. 2c).
4. *Pseudomonas* sp. HR199*calB*ΩKm, enthaltend das durch ΩKm inaktivierte *calB*-Gen an Stelle des intakten *calB*-Gens kodierend für Coniferylaldehyd-Dehydrogenase (Fig. 1d; Fig. 2d)
5. *Pseudomonas* sp. HR199*calB*ΩGm, enthaltend das durch ΩGm inaktivierte *calB*-Gen an Stelle des intakten *calB*-Gens kodierend für Coniferylaldehyd-Dehydrogenase (Fig. 1e; Fig. 2e).
6. *Pseudomonas* sp. HR199*calB*Δ, enthaltend das durch Deletion inaktivierte *calB*-Gen an Stelle des intakten *calB*-Gens kodierend für Coniferylaldehyd-Dehydrogenase (Fig. 1f; Fig. 2f).
7. *Pseudomonas* sp. HR199*fcs*ΩKm, enthaltend das durch ΩKm inaktivierte *fcs*-Gen an Stelle des intakten *fcs*-Gens kodierend für Ferulasäure-CoA-Synthetase (Fig. 1g; Fig. 2g).
8. *Pseudomonas* sp. HR199*fcs*ΩGm, enthaltend das durch ΩGm inaktivierte *fcs*-Gen an Stelle des intakten *fcs*-Gens kodierend für Ferulasäure-CoA-Synthetase (Fig. 1h; Fig. 2h).
9. *Pseudomonas* sp. HR199*fcs*Δ, enthaltend das durch Deletion inaktivierte *fcs*-Gen an Stelle des intakten *fcs*-Gens kodierend für Ferulasäure-CoA-Synthetase (Fig. 1i; Fig. 2i).
10. *Pseudomonas* sp. HR199*ech*ΩKm, enthaltend das durch ΩKm inaktivierte *ech*-Gen an Stelle des intakten *ech*-Gens kodierend für Enoyl-CoA-Hydratase-Aldolase (Fig. 1j; Fig. 2j).

11. *Pseudomonas* sp. HR199 $\text{ech}\Omega\text{Gm}$, enthaltend das durch ΩGm inaktivierte *ech*-Gen an Stelle des intakten *ech*-Gens kodierend für Enoyl-CoA-Hydratase-Aldolase (Fig. 1k; Fig. 2k).
12. *Pseudomonas* sp. HR199 $\text{ech}\Delta$, enthaltend das durch Deletion inaktivierte *ech*-Gen an Stelle des intakten *ech*-Gens kodierend für Enoyl-CoA-Hydratase-Aldolase (Fig. 1l; Fig. 2l).
13. *Pseudomonas* sp. HR199 $\text{aat}\Omega\text{Km}$, enthaltend das durch ΩKm inaktivierte *aat*-Gen an Stelle des intakten *aat*-Gens kodierend für beta-Ketothiolase (Fig. 1m; Fig. 2m).
14. *Pseudomonas* sp. HR199 $\text{aat}\Omega\text{Gm}$, enthaltend das durch ΩGm inaktivierte *aat*-Gen an Stelle des intakten *aat*-Gens kodierend für beta-Ketothiolase (Fig. 1n; Fig. 2n).
15. *Pseudomonas* sp. HR199 $\text{aat}\Delta$, enthaltend das durch Deletion inaktivierte *aat*-Gen an Stelle des intakten *aat*-Gens kodierend für beta-Ketothiolase (Fig. 1o; Fig. 2o).
16. *Pseudomonas* sp. HR199 $\text{vdh}\Omega\text{Km}$, enthaltend das durch ΩKm inaktivierte *vdh*-Gen an Stelle des intakten *vdh*-Gens kodierend für Vanillin-Dehydrogenase (Fig. 1p; Fig. 2p).
17. *Pseudomonas* sp. HR199 $\text{vdh}\Omega\text{Gm}$, enthaltend das durch ΩGm inaktivierte *vdh*-Gen an Stelle des intakten *vdh*-Gens kodierend für Vanillin-Dehydrogenase (Fig. 1q; Fig. 2q).
18. *Pseudomonas* sp. HR199 $\text{vdh}\Delta$, enthaltend das durch Deletion inaktivierte *vdh*-Gen an Stelle des intakten *vdh*-Gens kodierend für Vanillin-Dehydrogenase (Fig. 1r; Fig. 2r).
19. *Pseudomonas* sp. HR199 $\text{vdhB}\Omega\text{Km}$, enthaltend das durch ΩKm inaktivierte *vdhB*-Gen an Stelle des intakten *vdhB*-Gens kodierend für Vanillin-Dehydrogenase II.
20. *Pseudomonas* sp. HR199 $\text{vdhB}\Omega\text{Gm}$, enthaltend das durch ΩGm inaktivierte *vdhB*-Gen an Stelle des intakten *vdhB*-Gens kodierend für Vanillin-Dehydrogenase II.

21. *Pseudomonas* sp. HR199 $\text{vdh}\Delta$, enthaltend das durch Deletion inaktivierte vdhB -Gen an Stelle des intakten vdhB -Gens kodierend für Vanillin-Dehydrogenase II.
22. *Pseudomonas* sp. HR199 $\text{adh}\Omega\text{Km}$, enthaltend das durch ΩKm inaktivierte adh -Gen an Stelle des intakten adh -Gens kodierend für Alkohol-Dehydrogenase.
23. *Pseudomonas* sp. HR199 $\text{adh}\Omega\text{Gm}$, enthaltend das durch ΩGm inaktivierte adh -Gen an Stelle des intakten adh -Gens kodierend für Alkohol-Dehydrogenase.
24. *Pseudomonas* sp. HR199 $\text{adh}\Delta$ enthaltend das durch Deletion inaktivierte adh -Gen an Stelle des intakten adh -Gens kodierend für Alkohol-Dehydrogenase.
25. *Pseudomonas* sp. HR199 $\text{vanA}\Omega\text{Km}$, enthaltend das durch ΩKm inaktivierte vanA -Gen an Stelle des intakten vanA -Gens kodierend für die α -Untereinheit der Vanillinsäure-Demethylase.
26. *Pseudomonas* sp. HR199 $\text{vanA}\Omega\text{Gm}$, enthaltend das durch ΩGm inaktivierte vanA -Gen an Stelle des intakten vanA -Gens kodierend für die α -Untereinheit der Vanillinsäure-Demethylase.
27. *Pseudomonas* sp. HR199 $\text{vanA}\Delta$, enthaltend das durch Deletion inaktivierte vanA -Gen an Stelle des intakten vanA -Gens kodierend für die α -Untereinheit der Vanillinsäure-Demethylase.
28. *Pseudomonas* sp. HR199 $\text{vanB}\Omega\text{Km}$, enthaltend das durch ΩKm inaktivierte vanB -Gen an Stelle des intakten vanB -Gens kodierend für die β -Untereinheit der Vanillinsäure-Demethylase.
29. *Pseudomonas* sp. HR199 $\text{vanB}\Omega\text{Gm}$, enthaltend das durch ΩGm inaktivierte vanB -Gen an Stelle des intakten vanB -Gens kodierend für die β -Untereinheit der Vanillinsäure-Demethylase.
30. *Pseudomonas* sp. HR199 $\text{vanB}\Delta$, enthaltend das durch Deletion inaktivierte vanB -Gen an Stelle des intakten vanB -Gens kodierend für die β -Untereinheit der Vanillinsäure-Demethylase.

Gegenstand der Erfindung ist außerdem ein Verfahren zur biotechnischen Herstellung von organischen Verbindungen. Insbesondere können mit diesem Verfahren Alkohole, Aldehyde und organische Säuren hergestellt werden. Vorzugsweise handelt es sich hierbei um Coniferylalkohol, Coniferylaldehyd, Ferulasäure, Vanillin und Vanillinsäure.

In dem erfindungsgemäßen Verfahren werden die oben beschriebenen Organismen eingesetzt. Zu den ganz besonders bevorzugten Organismen gehören Bakterien, insbesondere die *Pseudomonas*-Arten. Im einzelnen können die oben genannten *Pseudomonas*-Arten vorzugsweise für folgende Verfahren eingesetzt werden:

1. *Pseudomonas* sp. HR199calA Ω Km, *Pseudomonas* sp. HR199calA Ω Gm und *Pseudomonas* sp. HR199calA Δ zur Herstellung von Coniferylalkohol aus Eugenol.

2. *Pseudomonas* sp. HR199calB Ω Km, *Pseudomonas* sp. HR199calB Ω Gm und *Pseudomonas* sp. HR199calB Δ zur Herstellung von Coniferylaldehyd aus Eugenol oder Coniferylalkohol.

3. *Pseudomonas* sp. HR199fcs Ω Km, *Pseudomonas* sp. HR199fcs Ω Gm, *Pseudomonas* sp. HR199fcs Δ , *Pseudomonas* sp. HR199ech Ω Km, *Pseudomonas* sp. HR199ech Ω Gm und *Pseudomonas* sp. HR199ech Δ zur Herstellung von Ferulasäure aus Eugenol oder Coniferylalkohol oder Coniferylaldehyd.

4. *Pseudomonas* sp. HR199vdh Ω Km, *Pseudomonas* sp. HR199vdh Ω Gm, *Pseudomonas* sp. HR199vdh Δ , *Pseudomonas* sp. HR199vdh Ω GmvdhB Ω Km, *Pseudomonas* sp. HR199vdh Ω KmvdhB Ω Gm, *Pseudomonas* sp. HR199vdh Δ vdhB Ω Gm und *Pseudomonas* sp. HR199vdh Δ vdhB Ω Km zur Herstellung von Vanillin aus Eugenol oder Coniferylalkohol oder Coniferylaldehyd oder Ferulasäure.

5. *Pseudomonas* sp. HR199vanA Ω Km, *Pseudomonas* sp. HR199vanA Ω Gm, *Pseudomonas* sp. HR199vanA Δ , *Pseudomonas* sp. HR199vanB Ω Km, *Pseudomonas* sp. HR199vanB Ω Gm und *Pseudomonas* sp. HR199vanB Δ zur Herstellung von Vanillinsäure aus Eugenol oder Coniferylalkohol oder Coniferylaldehyd oder Ferulasäure oder Vanillin.

Bevorzugtes Substrat ist Eugenol. Jedoch kann der Zusatz weiterer Substrate oder sogar der Austausch des Eugenol gegen ein anderes Substrat möglich sein.

10

Als Nährmedium für die erfindungsgemäß eingesetzten Organismen kommen synthetische, halbsynthetische oder komplexe Kulturmedien in Betracht. Diese können kohlenstoffhaltige und stickstoffhaltige Verbindungen, anorganische Salze, gegebenenfalls Spurenelemente sowie Vitamine enthalten.

15

Als kohlenstoffhaltige Verbindungen können Kohlenhydrate, Kohlenwasserstoffe oder organische Grundchemikalien in Betracht kommen. Beispiele für vorzugsweise verwendbare Verbindungen sind Zucker, Alkohole bzw. Zuckeralkohole, organische Säuren oder komplexe Gemische.

20

Als Zucker kommt vorzugsweise Glucose in Betracht. Als organische Säuren können vorzugsweise Zitronensäure oder Essigsäure zum Einsatz kommen. Zu den komplexen Gemischen zählen z. B. Malzextrakt, Hefeextrakt, Casein oder Caseinhydrolysat.

25

Als stickstoffhaltige Substrate kommen anorganische Verbindungen in Betracht. Beispiele hierfür sind Nitrate und Ammoniumsalze. Ebenso können organische Stickstoffquellen zum Einsatz kommen. Hierzu zählen Hefextrakt, Sojamehl, Casein, Caseinhydrolysat und Maisquellwasser.

30

Zu den einsetzbaren anorganischen Salzen zählen beispielsweise Sulfate, Nitrate, Chloride, Carbonate und Phosphate. Als Metalle enthalten die genannten Salze vorzugsweise Natrium, Kalium, Magnesium, Mangan, Calcium, Zink und Eisen.

- 5 Die Temperatur für die Kultivierung liegt vorzugsweise im Bereich von 5 bis 100°C. Besonders bevorzugt ist der Bereich von 15 bis 60°C, höchst bevorzugt sind 22 bis 37°C.

- 10 Der pH-Wert des Mediums beträgt bevorzugt 2 bis 12. Besonders bevorzugt ist der Bereich von 4 bis 8.

- Grundsätzlich können für die Durchführung des erfindungsgemäßen Verfahrens alle dem Fachmann bekannten Bioreaktoren eingesetzt werden. Vorzugsweise kommen alle für Submersverfahren geeigneten Vorrichtungen in Betracht. Das heißt, es
15 können erfindungsgemäß Gefäße ohne oder mit mechanischer Mischeinrichtung eingesetzt werden. Zu ersteren zählen z. B. Schüttelapparaturen, Blasensäulen- oder Schlaufenreaktoren. Zu letzteren gehören vorzugsweise alle bekannten Vorrichtungen mit Rührern in beliebiger Gestaltung.

- 20 Das erfindungsgemäße Verfahren kann kontinuierlich oder diskontinuierlich durchgeführt werden. Die Dauer der Fermentation bis zum Erreichen einer maximalen Produktmenge hängt von der speziellen Art des eingesetzten Organismus ab. Grundsätzlich liegen die Zeiten der Fermentation jedoch zwischen 2 und 200 Stunden.

- 25 Im folgenden wird die Erfindung unter Bezugnahme auf Beispiele näher erläutert:

- Von dem Eugenol verwertenden Stamm *Pseudomonas* sp. HR199 (DSM 7063) wurden gezielt Mutanten erzeugt, wobei spezifisch Gene des Eugenol-Katabolismus durch Insertion von Omega-Elementen oder durch Einführen von Deletionen inakti-
30 viert wurden. Als Omega-Elemente dienten DNA-Abschnitte die für Antibiotikaresistenzen gegen Kanamycin (Ω Km) und Gentamycin (Ω Gm) codierten. Diese

- Resistenzgene wurden ausgehend von Tn5 und dem Plasmid pBBR1MCS-5 mit Hilfe von Standardmethoden isoliert. Die Gene *calA*, *calB*, *fcs*, *ech*, *aat*, *vdh*, *adh*, *vdhB*, *vanA* und *vanB*, die für Coniferylalkohol-Dehydrogenase, Coniferylaldehyd-Dehydrogenase, Ferulasäure-CoA Synthetase, Enoyl-CoA Hydratase-Aldolase, beta-Ketothiolase, Vanillin-Dehydrogenase, Alkohol-Dehydrogenase, Vanillin-Dehydrogenase II und Vanillinsäure-Demethylase codieren wurden ausgehend von genomischer DNA des Stammes *Pseudomonas* sp. HR199 mit Hilfe von Standardmethoden isoliert und in pBluescript SK⁻ kloniert. Aus diesen Genen wurden durch Verdauung mit geeigneten Restriktionsendonukleasen DNA-Abschnitte entfernt (Deletion), bzw.
- 10 durch Ω -Elemente substituiert (Insertion), wodurch das jeweilige Gen inaktiviert wurde. Die auf diese Weise mutierten Gene wurden in konjugativ übertragbare Vektoren umklont und anschließend in den Stamm *Pseudomonas* sp. HR199 eingeführt. Durch geeignete Selektion wurden Transkonjuganten erhalten, die das jeweils funktionsfähige wildtyp-Gen gegen das neu eingebrachte inaktivierte Gen aus-
- 15 tauscht hatten. Die so erhaltenen Insertions- und Deletionsmutanten wiesen nur noch das jeweils inaktivierte Gen auf. Auf diese Weise wurden sowohl Mutanten mit nur einem defekten Gen als auch Mehrfachmutanten, in denen mehrere Gene auf diese Weise inaktiviert wurden, erhalten. Diese Mutanten wurden für die Biotransformation von
- 20 a) Eugenol zu Coniferylalkohol, Coniferylaldehyd, Ferulasäure, Vanillin und/oder Vanillinsäure;
- b) Coniferylalkohol zu Coniferylaldehyd, Ferulasäure, Vanillin und/oder Vanillinsäure;
- c) Coniferylaldehyd zu Ferulasäure, Vanillin und/oder Vanillinsäure;
- 25 d) Ferulasäure zu Vanillin und/oder Vanillinsäure und
- e) Vanillin zu Vanillinsäure eingesetzt.

Material und Methoden

Wachstumsbedingungen der Bakterien.

Stämme von *Escherichia coli* wurden bei 37°C in Luria-Bertani (LB) oder M9-Mineralmedium (Sambrook, J., E. F. Fritsch und T. Maniatis. 1989. Molecular cloning: a laboratory manual. 2. Aufl., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.) angezogen. Stämme von *Pseudomonas* sp. wurden bei 30°C in Nutrient Broth (NB, 0,8%, wt/vol) oder in Mineralmedium (MM) (Schlegel, H. G. et al. 1961. Arch. Mikrobiol. 38:209-222) bzw. HR-Mineralmedium (HR-MM) (Rabenhorst, J. 1996. Appl. Microbiol. Biotechnol. 46:470-474.) angezogen. Ferulasäure, Vanillin, Vanillinsäure und Protocatechusäure wurden in Dimethylsulfoxid gelöst, und dem jeweiligen Medium in einer Endkonzentration von 0.1% (wt/vol) zugesetzt. Eugenol wurde dem Medium direkt in einer Endkonzentration von 0.1% (vol/vol) zugesetzt, bzw. in den Deckel von MM-Agarplatten auf Filterpapier (Rundfilter 595, Schleicher & Schuell, Dassel, Deutschland) appliziert. Bei der Anzucht von Transkonjuganten und Mutanten von *Pseudomonas* sp. wurde Tetracyclin, Kanamycin, und Gentamycin in Endkonzentrationen von 25 µg/ml bzw. 100 µg/ml bzw. 7,5 µg/ml eingesetzt.

Qualitativer und quantitativer Nachweis von Stoffwechselintermediaten in Kulturüberständen.

Kulturüberstände wurden direkt, bzw. nach Verdünnung mit H₂O-bidest. mittels Hochdruck-Flüssigkeits-Chromatographie (Knauer-HPLC) analysiert. Die Chromatographie erfolgte an Nucleosil-100 C18 (7 µm, 250 x 4 mm). Als Lösungsmittel diente 0.1% (vol/vol) Ameisensäure und Acetonitril. Der verwendete Gradient zur Elution der Substanzen verlief wie folgt:

00:00 - 06:30 → 26% Acetonitril
06:30 - 08:00 → 100% Acetonitril
08:00 - 12:00 → 100% Acetonitril
12:00 - 13:00 → 26% Acetonitril
13:00 - 18:00 → 26% Acetonitril

Reinigung der Vanillin-Dehydrogenase-II.

Die Aufreinigung erfolgte bei 4°C.

5 Rohextrakt.

Auf Eugenol angezogene Zellen von *Pseudomonas* sp. HR199 wurden in 10 mM Natriumphosphat-Puffer, pH 6.0 gewaschen, im gleichen Puffer resuspendiert und durch zweimalige Passage einer French-Press (Amicon, Silver Spring, Maryland, USA) bei einem Druck von 1000 psi aufgeschlossen. Das Zellhomogenat wurde
10 einer Ultrazentrifugation (1 h, 100 000 x g, 4°C) unterzogen, wodurch die lösliche Fraktion des Rohextraktes als Überstand erhalten wurde.

Anionenaustauschchromatographie an DEAE-Sephacel.

Die lösliche Fraktion des Rohextraktes wurde über Nacht gegen 10 mM Natriumphosphat-Puffer, pH 6.0 dialysiert. Das Dialysat wurde auf eine mit 10 mM
15 Natriumphosphat-Puffer, pH 6.0 äquilibrierte DEAE-Sephacel-Säule (2,6 cm x 35 cm, Bettvolumen[BV]: 186 ml) mit einer Durchflußrate von 0.8 ml/min aufgetragen. Die Säule wurde mit zwei BV 10 mM Natriumphosphat-Puffer, pH 6.0 gespült. Die Elution der Vanillin-Dehydrogenase-II (VDH-II) erfolgte mit einem linearen Salzgradient von 0 bis 400 mM NaCl in 10 mM Natriumphosphat-Puffer, pH 6.0 (750
20 ml). Es wurden 10 ml-Fractionen aufgefangen. Fractionen mit hoher VDH-II-Aktivität wurden zum DEAE-Pool vereinigt.

Bestimmung der Vanillin-Dehydrogenase-Aktivität.

25 Die Bestimmung der VDH-Aktivität erfolgte bei 30°C durch einen optisch enzymatischen Test. Der Reaktionsansatz mit einem Volumen von 1 ml enthielt 0.1 mmol Kalium-Phosphat (pH 7.1), 0.125 µmol Vanillin, 0.5 µmol NAD, 1.2 µmol Pyruvat (Na-Salz), Lactat-Dehydrogenase (1 U; aus Schweineherz) und Enzymlösung. Die Oxidation von Vanillin wurde bei $\lambda = 340$ nm verfolgt ($\epsilon_{\text{Vanillin}} = 11,6 \text{ cm}^2/\mu\text{mol}$).
30 Die Enzymaktivität wurde in Einheiten (U) angegeben, wobei 1 U der Enzymmenge entspricht, die 1 µmol Vanillin pro Minute umsetzt. Die Proteinkonzentrationen in

den Proben wurden nach Lowry et al. (Lowry, O. H., N. J. Rosebrough, A. L. Farr und R. J. Randall. 1951. J. Biol. Chem. 193:265-275) bestimmt.

Bestimmung der Coniferylalkohol-Dehydrogenase-Aktivität.

- 5 Die Bestimmung der CADH-Aktivität erfolgte bei 30°C durch einen optisch enzymatischen Test nach Jaeger et al. (Jaeger, E., L. Eggeling und H. Sahm. 1981. Current Microbiology. 6:333-336). Der Reaktionsansatz mit einem Volumen von 1 ml enthielt 0.2 mmol Tris/HCl (pH 9.0), 0.4 µmol Coniferylalkohol, 2 µmol NAD, 0.1 mmol Semicarbazid und Enzymlösung. Die Reduktion von NAD wurde bei $\lambda =$
10 340 nm verfolgt ($\epsilon = 6.3 \text{ cm}^2/\mu\text{mol}$). Die Enzymaktivität wurde in Einheiten (U) angegeben, wobei 1 U der Enzymmenge entspricht, die 1 µmol Substrat pro Minute umsetzt. Die Proteinkonzentrationen in den Proben wurden nach Lowry et al. (Lowry, O. H., N. J. Rosebrough, A. L. Farr und R. J. Randall. 1951. J. Biol. Chem. 193:265-275) bestimmt.

15

Bestimmung der Coniferylaldehyd-Dehydrogenase-Aktivität.

- Die Bestimmung der CALDH-Aktivität erfolgte bei 30°C durch einen optisch enzymatischen Test. Der Reaktionsansatz mit einem Volumen von 1 ml enthielt 0.1 mmol Tris/HCl (pH 8.8), 0.08 µmol Coniferylaldehyd, 2.7 µmol NAD und Enzymlösung.
20 Die Oxidation von Coniferylaldehyd zu Ferulasäure wurde bei $\lambda = 400 \text{ nm}$ verfolgt ($\epsilon = 34 \text{ cm}^2/\mu\text{mol}$). Die Enzymaktivität wurde in Einheiten (U) angegeben, wobei 1 U der Enzymmenge entspricht, die 1 µmol Substrat pro Minute umsetzt. Die Proteinkonzentrationen in den Proben wurden nach Lowry et al. (Lowry, O. H., N. J. Rosebrough, A. L. Farr und R. J. Randall. 1951. J. Biol. Chem. 193:265-275)
25 bestimmt.

Bestimmung der Ferulasäure-CoA-Synthetase (Ferulasäure-Thiokinase)-Aktivität.

- Die Bestimmung der FCS-Aktivität erfolgte bei 30°C durch einen optisch enzymatischen Test, modifiziert nach Zenk et al. (Zenk et al. 1980. Anal. Biochem. 101:182-187). Der Reaktionsansatz mit einem Volumen von 1 ml enthielt 0.09 mmol Kalium-
- 30

Phosphat (pH 7.0), 2.1 μmol MgCl_2 , 0.7 μmol Ferulasäure, 2 μmol ATP, 0.4 μmol Coenzym A und Enzymlösung. Die Entstehung des CoA-Esters aus Ferulasäure wurde bei $\lambda = 345$ nm verfolgt ($\epsilon = 10 \text{ cm}^2/\mu\text{mol}$). Die Enzymaktivität wurde in Einheiten (U) angegeben, wobei 1 U der Enzymmenge entspricht, die 1 μmol Substrat pro Minute umsetzt. Die Proteinkonzentrationen in den Proben wurden nach Lowry et al. (Lowry, O. H., N. J. Rosebrough, A. L. Farr und R. J. Randall. 1951. J. Biol. Chem. 193:265-275) bestimmt.

Electrophoretische Methoden.

Die Auftrennung von proteinhaltigen Extrakten erfolgte in 7.4% (wt/vol) Polyacrylamidgelen unter nativen Bedingungen nach der Methode von Stegmann et al. (Stegmann et al. 1973. Z. Naturforsch. 28c:722-732) und unter denaturierenden Bedingungen in 11.5% (wt/vol) Polyacrylamidgelen nach der Methode von Laemmli (Laemmli, U. K. 1970. Nature (London) 227:680-685). Zur unspezifischen Proteinfärbung wurde Serva Blue R verwendet. Zur spezifischen Anfärbung der Coniferylalkohol-, Coniferylaldehyd- und Vanillin-Dehydrogenase wurden die Gele für 20 min in 100 mM KP-Puffer (pH 7.0) umgepuffert und anschließend bei 30°C im gleichen Puffer dem 0.08% (wt/vol) NAD, 0.04% (wt/vol) p-Nitroblau-Tetrazoliumchlorid, 0.003% (wt/vol) Phenazine-Methosulfat und 1 mM des jeweiligen Substrates zugesetzt worden war inkubiert, bis entsprechende Farbbanden sichtbar wurden.

Transfer von Proteinen aus Polyacrylamidgelen auf PVDF-Membranen. Proteine wurden aus SDS-Polyacrylamidgelen mit Hilfe eines Semidry-Fastblot Gerätes (B32/33, Biometra, Göttingen, Deutschland) nach Herstellerangaben auf PVDF-Membranen (Waters-Milipore, Bedford, Mass., USA) übertragen.

Bestimmung von N-terminalen Aminosäuresequenzen.

Die Bestimmung von N-terminalen Aminosäuresequenzen erfolgte mit Hilfe eines Protein Peptide Sequenzers (Typ 477 A, Applied Biosystems, Foster City, USA) und eines PTH-Analysers nach Herstellerangaben.

Isolierung und Manipulation von DNA.

Die Isolierung von genomischer DNA erfolgte nach der Methode von Marmur (Marmur, J. 1961. J. Mol. Biol. 3:208-218). Die Isolierung und Analyse von anderer Plasmid-DNA bzw. von DNA-Restriktionsfragmenten erfolgte nach Standardmethoden (Sambrook, J. E. F. Fritsch und T. Maniatis. 1989. Molecular cloning: a laboratory manual. 2. Aufl., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.).

Transfer von DNA.

Die Präparation und Transformation von kompetenten *Escherichia coli*-Zellen erfolgte nach der Methode von Hanahan (Hanahan, D. 1983. J. Mol. Biol. 166:557-580). Konjugativer Plasmidtransfer zwischen Plasmid-tragenden *Escherichia coli* S17-1-Stämmen (Donor) und *Pseudomonas* sp.-Stämmen (Rezipient) erfolgte auf NB-Agarplatten nach der Methode von Friedrich et al. (Friedrich, B. et al. 1981. J. Bacteriol. 147:198-205), oder durch eine "Minikomplementations-Methode" auf MM-Agarplatten mit 0.5% (wt/vol) Gluconat als C-Quelle und 25 µg/ml Tetracyclin oder 100 µg/ml Kanamycin. Dabei wurden Zellen des Rezipienten in einer Richtung als Impfstrich aufgetragen. Nach 5 min wurden dann Zellen der Donor-Stämme als Impfstriche aufgetragen, wobei der Rezipienten-Impfstrich gekreuzt wurde. Nach einer Inkubation für 48 h bei 30°C wuchsen die Transkonjuganten direkt hinter der Kreuzungsstelle, wohingegen weder Donor- noch Rezipienten-Stamm zum Wachstum in der Lage war.

Hybridisierungsexperimente.

DNA-Restriktionsfragmente wurden in einem 0.8% (wt/vol) Agarose-Gel in 50 mM Tris- 50 mM Borsäure- 1.25 mM EDTA-Puffer (pH 8.5) elektrophoretisch aufgetrennt (Sambrook, J. E. F. Fritsch und T. Maniatis. 1989. Molecular cloning: a laboratory manual. 2. Aufl., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.). Die Übertragung der denaturierten DNA aus dem Gel auf eine positiv geladene Nylonmembran (Porengröße: 0.45 µm, Pall Filtrationstechnik, Dreieich, Deutschland), die anschließende Hybridisierung mit biotinylierten, bzw.

Digoxigenin-markierten DNA-Sonden und die Herstellung dieser DNA-Sonden erfolgte nach Standardmethoden (Sambrook, J. E. F. Fritsch und T. Maniatis. 1989. Molecular cloning: a laboratory manual. 2. Aufl., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.).

5

DNA-Sequenzierung.

Die Bestimmung von Nukleotidsequenzen erfolgte nach der Didesoxy-Kettenabbruch-Methode von Sanger et al. (Sanger et al. 1977. Proc. Natl. Acad. Sci. USA 74:5463-5467) "nicht-radioaktiv" mit einem "LI-COR DNA-Sequencer Modell 4000L" (LI-COR Inc., Biotechnology Division, Lincoln, NE, USA) unter Verwendung eines "Thermo Sequenase fluorescent labelled primer cycle sequencing kit with 7-deaza-dGTP" (Amersham Life Science, Amersham International pls, Little Chalfont, Buckinghamshire, England) jeweils nach Vorschrift des Herstellers.

15 Mit Hilfe von synthetischen Oligonukleotiden wurde nach der "Primer-hopping Strategie" von Strauss et al. (Strauss, E. C. et al. 1986. Anal. Biochem. 154:353-360) sequenziert.

Chemikalien, Biochemikalien und Enzyme.

20 Restriktionsenzyme, T4 DNA-Ligase, Lambda-DNA und Enzyme bzw. Substrate für die optisch enzymatischen Tests wurden von C. F. Boehringer & Söhne (Mannheim, Deutschland) oder von GIBCO/BRL (Eggenstein, Deutschland) bezogen. [γ - ^{32}P]ATP kam von Amersham/Buchler (Braunschweig, Deutschland). Oligonukleotide wurden von der Firma MWG-Biotech GmbH (Ebersberg, Deutschland) bezogen.

25 Agarose vom Typ NA wurde von Pharmacia-LKB (Uppsala, Schweden) bezogen. Alle anderen Chemikalien waren von Haarmann & Reimer (Holzminden, Deutschland), E. Merck AG (Darmstadt, Deutschland), Fluka Chemie (Buchs, Schweiz), Serva Feinbiochemica (Heidelberg, Deutschland) oder Sigma Chemie (Deisenhofen, Deutschland).

Beispiele

Beispiel 1

- 5 **Konstruktion von Omega-Elementen, die Resistenzen gegenüber Kanamycin (Ω Km) bzw. Gentamycin(Ω Gm) vermitteln.**

Für die Konstruktion des Ω Km-Elements wurde das 2099 bp *Bgl*I-Fragment des Transposons Tn5 (Auerswald E. A., G. Ludwig und H. Schaller. 1981. Cold Spring Harb. Symp. Quant. Biol. 45:107-113; Beck E., G. Ludwig, E. A. Auerswald, B. Reiss und H. Schaller. 1982. Gene 19:327-336; Mazodier P., P. Cossart, E. Giraud und F. Gasser. 1985. Nucleic Acids Res. 13:195-205.) präparativ isoliert. Das Fragment wurde durch Behandlung mit der Nuklease Bal-31 auf ca. 990 bp verkürzt. Dieses Fragment, das nur noch das Kanamycin-Resistenzgen (codierend für eine Aminoglycosid-3'-O-Phosphotransferase) umfaßte, wurde anschließend mit *Sma*I geschnittener pSKsym-DNA (pBluescript SK⁻-Derivat, welches eine symmetrisch aufgebaute multiple Klonierungsstelle [*Sall*, *Hind*III, *Eco*RI, *Sma*I, *Eco*RI, *Hind*III, *Sall*] enthält) ligiert. Aus dem resultierenden Plasmid konnte das Ω Km-Element als *Sma*I-, *Eco*RI-, *Hind*III- oder *Sall*-Fragment reisoliert werden.

10
15

20 Für die Konstruktion des Ω Gm-Elements wurde das 983 bp *Eae*I-Fragment des Plasmids pBBR1MCS-5 (Kovach M.E., P. H. Elzer, D. S. Hill, G. T. Robertson, M. A. Farris, R. M. Roop und K. M. Peterson. 1995. Gene 166:175-176.) präparativ isoliert und anschließend mit Mung Bean Nuklease (Abdauen von einzelsträngigen DNA-Molekülen) behandelt. Dieses Fragment, das nur noch das Gentamycin-Resistenzgen (codierend für eine Gentamycin-3-Acetyltransferase) umfaßte, wurde anschließend mit *Sma*I geschnittener pSKsym-DNA (s.o.) ligiert. Aus dem resultierenden Plasmid konnte das Ω Gm-Element als *Sma*I-, *Eco*RI-, *Hind*III- oder *Sall*-Fragment reisoliert werden.

25

Beispiel 2

Klonierung der Gene aus *Pseudomonas* sp. HR199 (DSM7063), die durch Insertion von Ω -Elementen oder durch Deletionen inaktiviert werden sollten.

5 Die separaten Klonierungen der Gene *fcs*, *ech*, *vdh* und *aat* erfolgte ausgehend von den *E. coli* S17-1 Stämmen DSM 10439 und DSM 10440 mit den Plasmiden pE207 und pE5-1 (siehe EP-A 0845532). Aus diesen Plasmiden wurden die angegebenen Fragmente präparativ isoliert und wie im weiteren beschrieben behandelt:

10 Für die Klonierung des *fcs*-Gens wurde das 2350 bp große *SalI*/*EcoRI*-Fragment des Plasmids pE207 und das 3700 bp große *EcoRI*/*SalI*-Fragment des Plasmids pE5-1 zusammen in pBluescript SK⁻ in einer Weise kloniert, daß beide Fragmente über die *EcoRI*-Enden miteinander verbunden waren. Ausgehend von dem resultierenden Hybridplasmid wurde das 6050 bp *SalI*-Fragment präparativ isoliert und durch
15 Behandlung mit der Nuklease Bal-31 auf ca. 2480 bp verkürzt. Anschließend wurden an die Fragment-Enden *PstI*-Linker ligiert und das Fragment nach *PstI*-Verdauung in pBluescript SK⁻ kloniert (pSK*fcs*). Nach Transformation von *E. coli* XL1-Blue wurden Klone erhalten, die das *fcs*-Gen exprimierten und eine FCS-Aktivität von 0.2 U/mg Protein aufwiesen.

20

Für die Klonierung des *ech*-Gens wurde das 3800 bp große *HindIII*/*EcoRI*-Fragment des Plasmids pE207 präparativ isoliert und durch Behandlung mit der Nuklease Bal-31 auf ca. 1470 bp verkürzt. Anschließend wurden an die Fragment-Enden *EcoRI*-Linker ligiert und das Fragment nach *EcoRI*-Verdauung in pBluescript SK⁻ kloniert
25 (pSK*ech*).

25

Für die Klonierung des *vdh*-Gens wurde das 2350 bp große *SalI*/*EcoRI*-Fragment des Plasmids pE207 präparativ isoliert. Nach Klonierung in pBluescript SK⁻ wurde das Fragment mit Hilfe eines Exonuklease III/ Mung Bean Nukleasesystems einseitig
30 um ca. 1530 bp verkürzt. Anschließend wurde an das Fragmentende ein *EcoRI*-

30

Linker ligiert und das Fragment nach *EcoRI*-Verdauung in pBluescript SK⁻ kloniert (pSK*vdh*). Nach Transformation von *E. coli* XL1-Blue wurden Klone erhalten, die das *vdh*-Gen exprimierten und eine VDH-Aktivität von 0.01 U/mg Protein aufwiesen.

5

Für die Klonierung des *aat*-Gens wurde das 3700 bp große *EcoRI*/*SaII*-Fragment des Plasmids pE5-1 präparativ isoliert und durch Behandlung mit der Nuklease Bal-31 auf ca. 1590 bp verkürzt. Anschließend wurden an die Fragment-Enden *EcoRI*-Linker ligiert und das Fragment nach *EcoRI*-Verdauung in pBluescript SK⁻ kloniert (pSK*aat*).

10

Beispiel 3

Inaktivierung der oben beschriebenen Gene durch Insertion von Ω -Elementen, bzw durch Deletion von Teilbereichen dieser Gene.

15

Das Plasmid pSK*fcs*, welches das *fcs*-Gen enthielt wurde mit *Bss*HII verdaut, wodurch ein 1290 bp großes Fragment aus dem *fcs*-Gen herausgeschnitten wurde. Nach Religation wurde das Deletions-Derivat des *fcs*-Gens (*fcs* Δ) (siehe Abb. 1i und 2i) kloniert in pBluescript SK⁻ (pSK*fcs* Δ) erhalten. Darüber hinaus wurden nach Herausschneiden des Fragments die Omega-Elemente Ω Km und Ω Gm an dessen Stelle einligiert. Dadurch entstanden die Ω -inaktivierten Derivate des *fcs*-Gens (*fcs* Ω Km, siehe Abb. 1g und 2g) und (*fcs* Ω Gm, siehe Abb. 1h und 2h) kloniert in pBluescript SK⁻ (pSK*fcs* Ω Km und pSK*fcs* Ω Gm). In Rohextrakten der erhaltenen *E. coli* Klone, deren Hybridplasmide ein durch Deletion bzw. Ω -Element-Insertion inaktiviertes *fcs*-Gen aufwiesen, konnte keine FCS-Aktivität nachgewiesen werden.

20

25

Das Plasmid pSK*ech*, welches das *ech*-Gen enthielt, wurde mit *NruI* verdaut, wodurch ein 53 bp und ein 430 bp großes Fragment aus dem *ech*-Gen herausgeschnitten wurde. Nach Religation wurde das Deletions-Derivat des *ech*-Gens (*ech* Δ , siehe Abb. 1l und 2l) kloniert in pBluescript SK⁻ (pSK*ech* Δ) erhalten. Darüber

30

hinaus wurden nach Herausschneiden der Fragmente die Omega-Elemente ΩK_m und ΩG_m an deren Stelle einligiert. Dadurch entstanden die Ω -inaktivierten Derivate des *ech*-Gens (*ech* ΩK_m und *ech* ΩG_m) kloniert in pBluescript SK⁻ (pSK*ech* ΩK_m und pSK*ech* ΩG_m).

5

Das Plasmid pSK*vdh*, welches das *vdh*-Gen enthielt wurde mit *Bss*HIII verdaut, wodurch ein 210 bp großes Fragment aus dem *vdh*-Gen herausgeschnitten wurde. Nach Religation wurde das Deletions-Derivat des *vdh*-Gens (*vdh* Δ , siehe Abb. 1o und 2o) kloniert in pBluescript SK⁻ (pSK*vdh* Δ) erhalten. Darüber hinaus wurden nach Herausschneiden des Fragments die Omega-Elemente ΩK_m und ΩG_m an dessen Stelle einligiert. Dadurch entstanden die Ω -inaktivierten Derivate des *vdh*-Gens (*vdh* ΩK_m und *vdh* ΩG_m) kloniert in pBluescript SK⁻ (pSK*vdh* ΩK_m , siehe Abb. 1m und 2m) und (pSK*vdh* ΩG_m , siehe Abb. 1n und 2n). In Rohextrakten der erhaltenen *E. coli* Klone, deren Hybridplasmide ein durch Deletion bzw. Ω -Element-Insertion inaktiviertes *vdh*-Gen aufwiesen, konnte keine VDH-Aktivität nachgewiesen werden.

Das Plasmid pSK*aat*, welches das *aat*-Gen enthielt wurde mit *Bss*HIII verdaut, wodurch ein 59 bp großes Fragment aus dem *aat*-Gen herausgeschnitten wurde. Nach Religation wurde das Deletions-Derivat des *aat*-Gens (*aat* Δ , siehe Abb. 1r und 2r) kloniert in pBluescript SK⁻ (pSK*aat* Δ) erhalten. Darüber hinaus wurden nach Herausschneiden des Fragments die Omega-Elemente ΩK_m und ΩG_m an dessen Stelle einligiert. Dadurch entstanden die Ω -inaktivierten Derivate des *aat*-Gens (*aat* ΩK_m , siehe Abb. 1p und 2p) und (*aat* ΩG_m , siehe Abb. 1q und 2q) kloniert in pBluescript SK⁻ (pSK*aat* ΩK_m und pSK*aat* ΩG_m).

25

Beispiel 4

Umklonieren der durch Ω -Elemente inaktivierten Gene in das konjugativ übertragbare "Suizid-Plasmid" pSUP202.

- 5 Um die durch Ω -Elemente inaktivierten Gene in *Pseudomonas* sp. HR199 gegen die intakten Gene austauschen zu können, benötigt man einen Vektor, der einerseits in Pseudomonaden übertragen werden kann (konjugativ übertragbare Plasmide), andererseits dort jedoch nicht repliziert werden kann und somit in Pseudomonaden instabil ist ("Suizid-Plasmid"). DNA-Abschnitte, die mit Hilfe eines solchen Plasmid-
- 10 systems in Pseudomonaden übertragen werden, können nur erhalten bleiben, wenn sie durch homologe Rekombination (RecA-abhängige Rekombination) in das Genom der Bakterienzelle integriert werden. Im vorliegenden Fall wurde das "Suizid-Plasmid" pSUP202 (Simon et al. 1983. In: A. Pühler. Molecular genetics of the bacteria-plant interaction. Springer Verlag, Berlin, Heidelberg, New York, S. 98-
- 15 106.) eingesetzt.

Die inaktivierten Gene *fcs* Ω Km und *fcs* Ω Gm wurden nach *Pst*I-Verdauung aus den Plasmiden pSK*fcs* Ω Km und pSK*fcs* Ω Gm isoliert und mit *Pst*I geschnittener pSUP202 DNA ligiert. Die Ligationsansätze wurden nach *E. coli* S17-1 transformiert. Die Selektion erfolgte auf Tetracyclin-haltigem LB-Medium mit Kanamycin

20 bzw. Gentamycin. Es wurden Kanamycin-resistente Transformanden erhalten, deren Hybridplasmid (pSUP*fcs* Ω Km) das inaktivierte Gen *fcs* Ω Km enthielt. Das entsprechende Hybridplasmid (pSUP*fcs* Ω Gm) der Gentamycin-resistenten Transformanden enthielt das inaktivierte Gen *fcs* Ω Gm.

25 Die inaktivierten Gene *ech* Ω Km und *ech* Ω Gm wurden nach *Eco*RI-Verdauung aus den Plasmiden pSKE*ech* Ω Km und pSKE*ech* Ω Gm isoliert und mit *Eco*RI geschnittener pSUP202 DNA ligiert. Die Ligationsansätze wurden nach *E. coli* S17-1 transformiert. Die Selektion erfolgte auf Tetracyclin-haltigem LB-Medium mit Kanamycin

30 bzw. Gentamycin. Es wurden Kanamycin-resistente Transformanden erhalten, deren

Hybridplasmid (pSUP ϵ ch Ω Km) das inaktivierte Gen ϵ ch Ω Km enthielt. Das entsprechende Hybridplasmid (pSUP ϵ ch Ω Gm) der Gentamycin-resistenten Transformanden enthielt das inaktivierte Gen ϵ ch Ω Gm.

- 5 Die inaktivierten Gene ν dh Ω Km und ν dh Ω Gm wurden nach *Eco*RI-Verdauung aus den Plasmiden pSK ν dh Ω Km und pSK ν dh Ω Gm isoliert und mit *Eco*RI geschnittener pSUP202 DNA ligiert. Die Ligationsansätze wurden nach *E. coli* S17-1 transformiert. Die Selektion erfolgte auf Tetracyclin-haltigem LB-Medium mit Kanamycin bzw. Gentamycin. Es wurden Kanamycin-resistente Transformanden erhalten, deren
- 10 Hybridplasmid (pSUP ν dh Ω Km) das inaktivierte Gen ν dh Ω Km enthielt. Das entsprechende Hybridplasmid (pSUP ν dh Ω Gm) der Gentamycin-resistenten Transformanden enthielt das inaktivierte Gen ν dh Ω Gm.

- Die inaktivierten Gene α at Ω Km und α at Ω Gm wurden nach *Eco*RI-Verdauung aus
- 15 den Plasmiden pSK α at Ω Km und pSK α at Ω Gm isoliert und mit *Eco*RI geschnittener pSUP202 DNA ligiert. Die Ligationsansätze wurden nach *E. coli* S17-1 transformiert. Die Selektion erfolgte auf Tetracyclin-haltigem LB-Medium mit Kanamycin bzw. Gentamycin. Es wurden Kanamycin-resistente Transformanden erhalten, deren
- 20 Hybridplasmid (pSUP α at Ω Km) das inaktivierte Gen α at Ω Km enthielt. Das entsprechende Hybridplasmid (pSUP α at Ω Gm) der Gentamycin-resistenten Transformanden enthielt das inaktivierte Gen α at Ω Gm.

Beispiel 5

- 25 **Umklonieren der durch Deletion inaktivierten Gene in das konjugativ übertragbare "Suizid-Plasmid" mit "sacB-Selektionssystem" pHE55.**
- Um die durch Deletion inaktivierten Gene in *Pseudomonas* sp. HR199 gegen die intakten Gene austauschen zu können, benötigt man einen Vektor, der die schon für pSUP202 beschriebenen Eigenschaften aufweist. Da im Gegensatz zu den durch Ω -
- 30 Element inaktivierten Genen bei durch Deletion inaktivierten Genen keine Selektion

tionsmöglichkeit (keine Antibiotika-Resistenz) für den erfolgten Austausch der Gene in *Pseudomonas* sp. HR199 besteht, mußte ein anderes Selektionssystem zur Anwendung kommen. Bei dem "*sacB*-Selektionssystem" wird das auszutauschende, durch Deletion inaktivierte Gen in einem Plasmid kloniert, welches neben einem Antibiotika-Resistenzgen auch über das *sacB*-Gen verfügt. Nach konjugativer Übertragung dieses Hybridplasmids in einen Pseudomonaden wird das Plasmid durch homologe Rekombination an der Stelle in das Genom integriert, an der sich das intakte Gen befindet (erster "Cross over"). Auf diese Weise entsteht ein "heterogener" Stamm, der sowohl über ein intaktes als auch über ein durch Deletion inaktiviertes Gen verfügt, welche durch die pHE55-DNA voneinander getrennt sind. Diese Stämme weisen die durch den Vektor codierte Resistenz auf und besitzen darüber hinaus ein aktives *sacB*-Gen. Durch ein zweites homologes Rekombinationsereignis (zweiter "Cross over"), soll nun die pHE55-DNA zusammen mit dem intakten Gen aus der genomischen DNA ausgegliedert werden. Durch dieses Rekombinationsereignis entsteht ein Stamm, der nur noch über das inaktivierte Gen verfügt. Darüber hinaus kommt es zum Verlust der pHE55-codierten Antibiotika-Resistenz und des *sacB*-Gens. Streicht man Stämme auf Saccharose-haltigen Medien aus, werden Stämme die das *sacB*-Gen exprimieren im Wachstum gehemmt, da das Genprodukt Saccharose zu einem Polymer umsetzt, welches im Periplasma der Zellen akkumuliert wird. Zellen, die durch das zweite Rekombinationsereignis das *sacB*-Gen nicht mehr tragen, werden somit nicht im Wachstum gehemmt. Um eine phänotypische Selektionsmöglichkeit auf die Integration des durch Deletion inaktivierten Gens zu besitzen, tauscht man dieses nicht gegen ein intaktes Gen aus, sondern man bedient sich eines Stammes, in dem das auszutauschende Gen bereits durch Insertion eines Ω -Elements "markiert" vorliegt. Bei erfolgreichem Austausch verliert der resultierende Stamm die durch das Ω -Element codierte Antibiotika-Resistenz.

Das inaktivierte Gen *fcsΔ* wurden nach *Pst*I-Verdauung aus dem Plasmid pSK/*fcsΔ* isoliert und mit *Pst*I geschnittener pHE55 DNA ligiert. Der Ligationsansatz wurde nach *E. coli* S17-1 transformiert. Die Selektion erfolgte auf Tetracyclin-haltigem LB-

Medium. Es wurden Tetracyclin-resistente Transformanden erhalten, deren Hybridplasmid (pHE*fcs*Δ) das inaktivierte Gen *fcs*Δ enthielt.

5 Das inaktivierte Gen *ech*Δ wurden nach *Eco*RI-Verdauung aus dem Plasmid pSK*ech*Δ isoliert und mit Mung Bean Nuklease behandelt (Erzeugung von glatten Enden ["blunt ends"]). Das Fragment wurde mit *Bam*HI geschnittener und Mung Bean Nuklease behandelter pHE55 DNA ligiert. Der Ligationsansatz wurde nach *E. coli* S17-1 transformiert. Die Selektion erfolgte auf Tetracyclin-haltigem LB-Medium. Es wurden Tetracyclin-resistente Transformanden erhalten, deren
10 Hybridplasmid (pHE*ech*Δ) das inaktivierte Gen *ech*Δ enthielt.

Das inaktivierte Gen *vdh*Δ wurden nach *Eco*RI-Verdauung aus dem Plasmid pSK*vdh*Δ isoliert und mit Mung Bean Nuklease behandelt. Das Fragment wurde mit *Bam*HI geschnittener und Mung Bean Nuklease behandelter pHE55 DNA ligiert. Der
15 Ligationsansatz wurde nach *E. coli* S17-1 transformiert. Die Selektion erfolgte auf Tetracyclin-haltigem LB-Medium. Es wurden Tetracyclin-resistente Transformanden erhalten, deren Hybridplasmid (pHE*vdh*Δ) das inaktivierte Gen *vdh*Δ enthielt.

Das inaktivierte Gen *aat*Δ wurden nach *Eco*RI-Verdauung aus dem Plasmid pSK*aat*Δ isoliert und mit Mung Bean Nuklease behandelt. Das Fragment wurde mit *Bam*HI geschnittener und Mung Bean Nuklease behandelter pHE55 DNA ligiert. Der
20 Ligationsansatz wurde nach *E. coli* S17-1 transformiert. Die Selektion erfolgte auf Tetracyclin-haltigem LB-Medium. Es wurden Tetracyclin-resistente Transformanden erhalten, deren Hybridplasmid (pHE*aat*Δ) das inaktivierte Gen *aat*Δ enthielt.

25

Beispiel 6

Erzeugung von Mutanten des Stammes *Pseudomonas* sp. HR199, bei denen spezifisch Gene des Eugenol-Katabolismus durch Insertion eines Ω -Elementes inaktiviert wurden.

Der Stamm *Pseudomonas* sp. HR199 wurde als Rezipient in Konjugationsexperimenten eingesetzt, bei denen Stämme von *E. coli* S17-1 als Donoren eingesetzt wurden, die die unten aufgeführten Hybridplasmide von pSUP202 enthielten. Die Transkonjuganten wurden auf Gluconat-haltigem Mineralmedium selektiert, welches das dem Ω -Element entsprechende Antibiotikum enthielt. "Homogenote" (Austausch des intakten Gens gegen das durch Ω -Element-Insertion inaktivierte Gen durch doppeltes "Cross over") und "heterogenote" (Integration des Hybridplasmids in das Genom durch einfachen "Cross over") Transkonjuganten konnten anhand der durch pSUP202 codierten Tetracyclin-Resistenz unterschieden werden.

Die Mutanten *Pseudomonas* sp. HR199 *fcs* Ω Km und *Pseudomonas* sp. HR199 *fcs* Ω Gm wurden nach Konjugation von *Pseudomonas* sp. HR199 mit *E. coli* S17-1 (pSUP*fcs* Ω Km) bzw. *E. coli* S17-1 (pSUP*fcs* Ω Gm) erhalten. Der Austausch des intakten *fcs*-Gens gegen das durch Ω Km bzw. Ω Gm inaktivierte Gen (*fcs* Ω Km bzw. *fcs* Ω Gm) wurde mittels DNA-Sequenzierung verifiziert.

Die Mutanten *Pseudomonas* sp. HR199 *ech* Ω Km und *Pseudomonas* sp. HR199 *ech* Ω Gm wurden nach Konjugation von *Pseudomonas* sp. HR199 mit *E. coli* S17-1 (pSUP*ech* Ω Km) bzw. *E. coli* S17-1 (pSUP*ech* Ω Gm) erhalten. Der Austausch des intakten *ech*-Gens gegen das durch Ω Km bzw. Ω Gm inaktivierte Gen (*ech* Ω Km bzw. *ech* Ω Gm) wurde mittels DNA-Sequenzierung verifiziert.

Die Mutanten *Pseudomonas* sp. HR199 *vdh* Ω Km und *Pseudomonas* sp. HR199 *vdh* Ω Gm wurden nach Konjugation von *Pseudomonas* sp. HR199 mit *E. coli* S17-1 (pSUP*vdh* Ω Km) bzw. *E. coli* S17-1 (pSUP*vdh* Ω Gm) erhalten. Der Austausch des

intakten *vdh*-Gens gegen das durch Ω Km bzw. Ω Gm inaktivierte Gen (*vdh* Ω Km bzw. *vdh* Ω Gm) wurde mittels DNA-Sequenzierung verifiziert.

- 5 Die Mutanten *Pseudomonas* sp. HR199 *aat* Ω Km und *Pseudomonas* sp. HR199 *aat* Ω Gm wurden nach Konjugation von *Pseudomonas* sp. HR199 mit *E. coli* S17-1 (pSUP*aat* Ω Km) bzw. *E. coli* S17-1 (pSUP*aat* Ω Gm) erhalten. Der Austausch des intakten *aat*-Gens gegen das durch Ω Km bzw. Ω Gm inaktivierte Gen (*aat* Ω Km bzw. *aat* Ω Gm) wurde mittels DNA-Sequenzierung verifiziert.
- 10 Die Mutante *Pseudomonas* sp. HR199 *fcs* Ω Km*vdh* Ω Gm wurden nach Konjugation von *Pseudomonas* sp. HR199 *fcs* Ω Km mit *E. coli* S17-1 (pSUP*vdh* Ω Gm) erhalten. Der Austausch des intakten *vdh*-Gens gegen das durch Ω Gm inaktivierte Gen (*vdh* Ω Gm) wurde mittels DNA-Sequenzierung verifiziert.
- 15 Die Mutante *Pseudomonas* sp. HR199 *vdh* Ω Km*aat* Ω Gm wurden nach Konjugation von *Pseudomonas* sp. HR199 *vdh* Ω Km mit *E. coli* S17-1 (pSUP*aat* Ω Gm) erhalten. Der Austausch des intakten *aat*-Gens gegen das durch Ω Gm inaktivierte Gen (*aat* Ω Gm) wurde mittels DNA-Sequenzierung verifiziert.
- 20 Die Mutante *Pseudomonas* sp. HR199 *vdh* Ω Km*ech* Ω Gm wurden nach Konjugation von *Pseudomonas* sp. HR199 *vdh* Ω Km mit *E. coli* S17-1 (pSUP*ech* Ω Gm) erhalten. Der Austausch des intakten *ech*-Gens gegen das durch Ω Gm inaktivierte Gen (*ech* Ω Gm) wurde mittels DNA-Sequenzierung verifiziert.

Beispiel 7

Erzeugung von Mutanten des Stammes *Pseudomonas* sp. HR199, bei denen spezifisch Gene des Eugenol-Katabolismus durch Deletion eines Teilbereiches inaktiviert wurden.

Die Stämme *Pseudomonas* sp. HR199 *fcs*ΩKm, *Pseudomonas* sp. HR199 *ech*ΩKm, *Pseudomonas* sp. HR199 *vdh*ΩKm und *Pseudomonas* sp. HR199 *aat*ΩKm wurden als Rezipient in Konjugationsexperimenten eingesetzt, bei denen Stämme von *E. coli* S17-1 als Donoren eingesetzt wurden, die die unten aufgeführten Hybridplasmide von pHE55 enthielten. Die "heterogenoten" Transkonjuganten wurden auf Gluconathaltigem Mineralmedium selektiert, welches neben Tetracyclin (pHE55 codierte Resistenz) das dem Ω-Element entsprechende Antibiotikum enthielt. Nach Ausstreichen auf Saccharose-haltigem Mineralmedium wurden Transkonjuganten erhalten, die durch ein zweites Rekombinationsereignis (zweiter "Cross over") die Vektor-DNA eliminiert hatten. Durch Ausstreichen auf Mineralmedium ohne Antibiotika bzw. mit dem Ω-Element entsprechenden Antibiotikum konnten die Mutanten erkannt werden, bei denen das durch Ω-Element inaktivierte Gen gegen das durch Deletion inaktivierte Gen ausgetauscht worden war (keine Antibiotika-Resistenz).

Die Mutante *Pseudomonas* sp. HR199 *fcs*Δ wurde nach Konjugation von *Pseudomonas* sp. HR199 *fcs*ΩKm mit *E. coli* S17-1 (pHE*fcs*Δ) erhalten. Der Austausch des durch ΩKm inaktivierten Gens (*fcs*ΩKm) gegen das durch Deletion inaktivierte Gen (*fcs*Δ) wurde mittels DNA-Sequenzierung verifiziert.

Die Mutanten *Pseudomonas* sp. HR199 *ech*Δ wurde nach Konjugation von *Pseudomonas* sp. HR199 *ech*ΩKm mit *E. coli* S17-1 (pHE*ech*Δ) erhalten. Der Austausch des durch ΩKm inaktivierten Gens (*ech*ΩKm) gegen das durch Deletion inaktivierte Gen (*ech*Δ) wurde mittels DNA-Sequenzierung verifiziert.

Die Mutanten *Pseudomonas* sp. HR199 *vdh* Δ wurde nach Konjugation von *Pseudomonas* sp. HR199 *vdh* Ω Km mit *E. coli* S17-1 (pHE*vdh* Δ) erhalten. Der Austausch des durch Ω Km inaktivierten Gens (*vdh* Ω Km) gegen das durch Deletion inaktivierte Gen (*vdh* Δ) wurde mittels DNA-Sequenzierung verifiziert.

5

Die Mutanten *Pseudomonas* sp. HR199 *aat* Δ wurde nach Konjugation von *Pseudomonas* sp. HR199 *aat* Ω Km mit *E. coli* S17-1 (pHE*aat* Δ) erhalten. Der Austausch des durch Ω Km inaktivierten Gens (*aat* Ω Km) gegen das durch Deletion inaktivierte Gen (*aat* Δ) wurde mittels DNA-Sequenzierung verifiziert.

10

Beispiel 8

Biotransformation von Eugenol zu Vanillin mit der Mutante *Pseudomonas* sp. HR199 *vdh* Ω Km.

15

Der Stamm *Pseudomonas* sp. HR199 *vdh* Ω Km wurde in 50 ml HR-MM mit 6 mM Eugenol bis zu einer optischen Dichte von ca. OD_{600nm} = 0.6 angezogen. Nach 17 h waren 2.9 mM Vanillin, 1.4 mM Ferulasäure und 0.4 mM Vanillinsäure im Kulturüberstand nachweisbar.

20

Beispiel 9

Biotransformation von Eugenol zu Ferulasäure mit der Mutante *Pseudomonas* sp. HR199 *vdh* Ω G*maat* Ω Km.

25

Der Stamm *Pseudomonas* sp. HR199 *vdh* Ω G*maat* Ω Km wurde in 50 ml HR-MM mit 6 mM Eugenol bis zu einer optischen Dichte von ca. OD_{600nm} = 0.6 angezogen. Nach 18 h waren 1.9 mM Vanillin, 2.4 mM Ferulasäure und 0.6 mM Vanillinsäure im Kulturüberstand nachweisbar.

Beispiel 10

Biotransformation von Eugenol zu Coniferylalkohol mit der Mutante *Pseudomonas* sp. HR199 *vdh*Ω*Gmaat*Ω*Km*.

- 5 Der Stamm *Pseudomonas* sp. HR199 *vdh*Ω*Gmaat*Ω*Km* wurde in 50 ml HR-MM mit 6 mM Eugenol bis zu einer optischen Dichte von ca. OD_{600nm} = 0.4 angezogen. Nach 15 h waren 1.7 mM Coniferylalkohol, 1.4 mM Vanillin, 1.4 mM Ferulasäure und 0.2 mM Vanillinsäure im Kulturüberstand nachweisbar.

10 **Beispiel 11**

Fermentative Produktion von natürlichem Vanillin aus Eugenol im 10 l Fermenter mit der Mutante *Pseudomonas* sp. HR 199 *vdh*Ω*Km*.

- Mit 100 ml einer 24 Stunden alten Vorkultur, die auf einer Schüttelmaschine (120 Upm) bei 32°C in einem auf pH 7,0 eingestellten Medium aus 12,5 g/l Glyzerin, 10 g/l Hefeextrakt und 0,37 g/l Essigsäure angezogen wurde, wurde der Produktionsfermenter beimpft. Der Fermenter enthielt 9,9 l Medium mit folgender Zusammensetzung: 1,5 g/l Hefeextrakt, 1,6 g/l KH₂PO₄, 0,2 g/l NaCl, 0,2 g/l MgSO₄. Der pH-Wert wurde mit Natronlauge auf pH 7,0 eingestellt. Nach der Sterilisation wurde dem Medium 4 g Eugenol zugefügt. Die Temperatur betrug 32°C, die Belüftung 3 NI/min und die Rührerdrehzahl 600 Upm. Der pH-Wert wurde mit Natronlauge bei pH 6,5 gehalten.
- 15
20

- Vier Stunden nach dem Animpfen wurde mit der kontinuierlichen Zugabe von Eugenol begonnen, so daß am Ende der Fermentation nach 65 Stunden 255 g Eugenol zur Kultur gegeben worden waren. Außerdem wurden während der Fermentation 40 g Hefeextrakt zugefüttert. Die Konzentration an Eugenol lag am Ende der Fermentation bei 0,2 g/l. Der Gehalt an Vanillin betrug 2,6 g/l. Zusätzlich lagen noch 3,4 g/l Ferulasäure vor.
- 25

Das so erhaltene Vanillin kann durch bekannte physikalische Verfahren wie Chromatographie, Destillation und/oder Extraktion isoliert und zur Herstellung natürlicher Aromen verwendet werden.

5 Erläuterungen zu den Figuren:

FIG. 1a bis 1r:

Gen-Strukturen zur Gewinnung von Organismen und Mutanten

10

*calA**: Teil des inaktivierten Gens der Coniferylalkohol-Dehydrogenase

*calB**: Teil des inaktivierten Gens der Coniferylaldehyd-Dehydrogenase

*fcs**: Teil des inaktivierten Gens der Ferulasäure-CoA Synthetase

*ech**: Teil des inaktivierten Gens der Enoyl-CoA Hydratase-Aldolase

15 *vdh**: Teil des inaktivierten Gens der Vanillin-Dehydrogenase

*aat**: Teil des inaktivierten Gens der beta-Ketothiolase

Die mit "*" versehenen Restriktionsenzym-Schnittstellen kamen für die Konstruktion zum Einsatz, sind jedoch in dem resultierenden Konstrukt nicht mehr funktionsfähig.

20

FIG. 2a: Nukleotidsequenz der Gen-Struktur *calA*ΩKm

FIG. 2b: Nukleotidsequenz der Gen-Struktur *calA*ΩGm:

FIG. 2c: Nukleotidsequenz der Gen-Struktur *calA*Δ

FIG. 2d: Nukleotidsequenz der Gen-Struktur *calB*ΩKm

5 FIG. 2e: Nukleotidsequenz der Gen-Struktur *calB*ΩGm

FIG. 2f: Nukleotidsequenz der Gen-Struktur *calB*Δ

FIG. 2g: Nukleotidsequenz der Gen-Struktur *fcs*ΩKm

FIG. 2h: Nukleotidsequenz der Gen-Struktur *fcs*ΩGm

FIG. 2i: Nukleotidsequenz der Gen-Struktur *fcs*Δ

10 FIG. 2j: Nukleotidsequenz der Gen-Struktur *ech*ΩKm

FIG. 2k: Nukleotidsequenz der Gen-Struktur *ech*ΩGm

FIG. 2l: Nukleotidsequenz der Gen-Struktur *ech*Δ

FIG. 2m: Nukleotidsequenz der Gen-Struktur *vdh*ΩKm

FIG. 2n: Nukleotidsequenz der Gen-Struktur *vdh*ΩGm

15 FIG. 2o: Nukleotidsequenz der Gen-Struktur *vdh*Δ

FIG. 2p: Nukleotidsequenz der Gen-Struktur *aat*ΩKm

FIG. 2q: Nukleotidsequenz der Gen-Struktur *aat*ΩGm

FIG. 2r: Nukleotidsequenz der Gen-Struktur *aat*Δ

Patentansprüche

1. Transformierter und/oder mutagenisierter ein- oder mehrzelliger Organismus, der dadurch gekennzeichnet ist, daß Enzyme des Eugenol- und/oder Ferulasäure-Katabolismus derart inaktiviert sind, daß eine Akkumulation der Intermediate Coniferylalkohol, Coniferylaldehyd, Ferulasäure, Vanillin und/oder Vanillinsäure erfolgt.
5
2. Organismus nach Anspruch 1, dadurch gekennzeichnet, daß der Eugenol- und/oder Ferulasäure-Katabolismus durch Ω -Element-Insertion oder Einführen von Deletionen in entsprechende Gene verändert ist.
10
3. Organismus nach einem der Ansprüche 1 oder 2, dadurch gekennzeichnet, daß ein oder mehrere Gene, die für die Enzyme Coniferylalkohol-Dehydrogenasen, Coniferylaldehyd-Dehydrogenasen, Ferulasäure-CoA Synthetasen, Enoyl-CoA Hydratasen-Aldolasen, beta-Ketothiolasen, Vanillin-Dehydrogenasen oder Vanillinsäure-Demethylasen Enzyme kodieren, verändert und/oder inaktiviert sind.
15
4. Organismus nach einem der Ansprüche 1 bis 3, dadurch gekennzeichnet, daß er einzellig, vorzugsweise ein Mikroorganismus oder eine pflanzliche oder eine tierische Zelle ist.
20
5. Organismus nach einem der Ansprüche 1 bis 4, dadurch gekennzeichnet, daß er ein Bakterium, vorzugsweise eine *Pseudomonas*-Art ist.
25
6. Gen-Strukturen, bei denen die für die Enzyme Coniferylalkohol-Dehydrogenasen, Coniferylaldehyd-Dehydrogenasen, Ferulasäure-CoA Synthetasen, Enoyl-CoA Hydratasen-Aldolasen, beta-Ketothiolasen, Vanillin-Dehydrogenasen oder Vanillinsäure-Demethylasen oder zweier oder mehrerer dieser Enzyme kodierenden Nukleotidsequenzen verändert und/oder inaktiviert sind.
30

7. Gen-Strukturen mit den in Figur 1a bis 1r angegebenen Strukturen.
8. Gen-Strukturen mit den in Figur 2a bis 2r angegebenen Sequenzen.
- 5 9. Vektoren enthaltend wenigstens eine Gen-Struktur nach einem der Ansprüche 6 bis 8.
- 10 10. Transformierter Organismus nach einem der Ansprüche 1 bis 5, dadurch gekennzeichnet, daß er wenigstens einen Vektor gemäß Anspruch 9 enthält.
11. Organismus nach einem der Ansprüche 1 bis 5, dadurch gekennzeichnet, daß er wenigstens eine Gen-Struktur nach einem der Ansprüche 6 bis 8 an Stelle des jeweiligen intakten Gens im Genom integriert enthält.
- 15 12. Verfahren zur biotechnischen Herstellung von organischen Verbindungen, insbesondere von Alkoholen, Aldehyden und organischen Säuren, dadurch gekennzeichnet, daß ein Organismus nach einem der Ansprüche 1 bis 5 oder 10 bis 11 eingesetzt wird.
- 20 13. Verfahren zur Herstellung der Organismen nach einem der Ansprüche 1 bis 5, dadurch gekennzeichnet, daß die Veränderung des Eugenol- und/oder Ferulasäure-Katabolismus mittels an sich bekannter mikrobiologischer Züchtungsmethoden erzielt wird.
- 25 14. Verfahren zur Herstellung eines Organismus nach einem der Ansprüche 1 bis 5 oder 10 bis 11, dadurch gekennzeichnet, daß die Veränderung des Eugenol- und/oder Ferulasäure-Katabolismus und/oder die Inaktivierung der entsprechenden Gene mittels gentechnischer Methoden erzielt wird.
- 30

15. Verwendung der Organismen nach einem der Ansprüche 1 bis 5 oder 10 bis 11 zur Herstellung von Coniferylalkohol, Coniferylaldehyd, Ferulasäure, Vanillin und/oder Vanillinsäure.
- 5 16. Verwendung von Gen-Strukturen nach einem der Ansprüche 6 bis 8 oder eines Vektors nach Anspruch 9 zur Herstellung transformierter und/oder mutagenisierter Organismen.

Sequenzen

CTGCAGCCAG GGCTGAAAAG GAGGGATTCA GTGAGGTCAT GAAGGGAGGG GACGGCGCCT	60
GGCTCCAATT GCTCGATGGC GCCGCGATTG AGTGTCTTGG GCGCGGTCTT GGAGAGTTCTG	120
GCTAGGGAGA TAAATTTGCT GGCCATGGTG GCGGCCCTTG ATGGGTTGGA TGATTTTCTG	180
CATTCTGCAT CATGAAATTC ATGAAATCAT CACTTTTCGG GGGGTGGGTG CACGGGATTG	240
AAGGTTGCTA GGAGAGTGCA TTGCTCGTAA GCCCAGGAAG CACGCGGGTT TCAGGATGGT	300
GCATGGAAAT GGCATGAGCT TTGCTGGATA TGATTAGAGA CATTAACTAT TTTGGCGGAA	360
TGGAAGCACG ATTCCTCGCC CGGTAGAGCG GTAACCGCGA CATTCAAGAC CGTAAAAAGG	420
AAAGAGCATG CAA CTG ACC AAC AAG AAA ATC GTC GTC ACC GGA GTG TCC TCC	472
Met Gln Leu Thr Asn Lys Lys Ile Val Val Thr Gly Val Ser Ser	
1 5 10 15	
GGT ATC GGT GCC GAA ACT GCC CGC GTT CTG CGC TCT CAC GGC GCC ACA	520
Gly Ile Gly Ala Glu Thr Ala Arg Val Leu Arg Ser His Gly Ala Thr	
20 25 30	
GTG ATT GGC GTA GAT CGC AAC ATG CCG AGC CTG ACT CTG GAT GCT TTC	568
Val Ile Gly Val Asp Arg Asn Met Pro Ser Leu Thr Leu Asp Ala Phe	
35 40 45	
GTT CAG GCT GAC CTG AGC CAT CCT GAA GGC ATC GAT AAG GCC ATC GGG	616
Val Gln Ala Asp Leu Ser His Pro Glu Gly Ile Asp Lys Ala Ile	
50 55 60 62	
ACAGCAAGCG AACCGGAATT GCCAGCTGGG GCGCCCTCTG GTAAGGTTGG GAAGCCCTGC	676
AAAGTAAACT GGATGGCTTT CTTGCCGCCA AGGATCTGAT GGCGCAGGGG ATCAAGATCT	736
GATCAAGAGA CAGGATGAGG ATCGTTTCGC ATG ATT GAA CAA GAT GGA TTG CAC	790
Met Ile Glu Gln Asp Gly Leu His	
1 5	
GCA GGT TCT CCG GCC GCT TGG GTG GAG AGG CTA TTC GGC TAT GAC TGG	838
Ala Gly Ser Pro Ala Ala Trp Val Glu Arg Leu Phe Gly Tyr Asp Trp	
10 15 20	
GCA CAA CAG ACA ATC GGC TGC TCT GAT GCC GCC GTG TTC CGG CTG TCA	886
Ala Gln Gln Thr Ile Gly Cys Ser Asp Ala Ala Val Phe Arg Leu Ser	
25 30 35 40	
GCG CAG GGG CGC CCG GTT CTT TTT GTC AAG ACC GAC CTG TCC GGT GCC	934
Ala Gln Gly Arg Pro Val Leu Phe Val Lys Thr Asp Leu Ser Gly Ala	
45 50 55	
CTG AAT GAA CTG CAG GAC GAG GCA GCG CGG CTA TCG TGG CTG GCC ACG	982
Leu Asn Glu Leu Gln Asp Glu Ala Ala Arg Leu Ser Trp Leu Ala Thr	
60 65 70	

ACG GGC GTT CCT TGC GCA GCT GTG CTC GAC GTT GTC ACT GAA GCG GGA	1030
Thr Gly Val Pro Cys Ala Ala Val Leu Asp Val Val Thr Glu Ala Gly	
75 80 85	
AGG GAC TGG CTG CTA TTG GGC GAA GTG CCG GGG CAG GAT CTC CTG TCA	1078
Arg Asp Trp Leu Leu Leu Gly Glu Val Pro Gly Gln Asp Leu Leu Ser	
90 95 100	
TCT CAC CTT GCT CCT GCC GAG AAA GTA TCC ATC ATG GCT GAT GCA ATG	1126
Ser His Leu Ala Pro Ala Glu Lys Val Ser Ile Met Ala Asp Ala Met	
105 110 115 120	
CGG CGG CTG CAT ACG CTT GAT CCG GCT ACC TGC CCA TTC GAC CAC CAA	1174
Arg Arg Leu His Thr Leu Asp Pro Ala Thr Cys Pro Phe Asp His Gln	
125 130 135	
GCG AAA CAT CGC ATC GAG CGA GCA CGT ACT CGG ATG GAA GCC GGT CTT	1222
Ala Lys His Arg Ile Glu Arg Ala Arg Thr Arg Met Glu Ala Gly Leu	
140 145 150	
GTC GAT CAG GAT GAT CTG GAC GAA GAG CAT CAG GGG CTC GCG CCA GCC	1270
Val Asp Gln Asp Asp Leu Asp Glu Glu His Gln Gly Leu Ala Pro Ala	
155 160 165	
GAA CTG TTC GCC AGG CTC AAG GCG CGC ATG CCC GAC GGC GAG GAT CTC	1318
Glu Leu Phe Ala Arg Leu Lys Ala Arg Met Pro Asp Gly Glu Asp Leu	
170 175 180	
GTC GTG ACC CAT GGC GAT GCC TGC TTG CCG AAT ATC ATG GTG GAA AAT	1366
Val Val Thr His Gly Asp Ala Cys Leu Pro Asn Ile Met Val Glu Asn	
185 190 195 200	
GGC CGC TTT TCT GGA TTC ATC GAC TGT GGC CGG CTG GGT GTG GCG GAC	1414
Gly Arg Phe Ser Gly Phe Ile Asp Cys Gly Arg Leu Gly Val Ala Asp	
205 210 215	
CGC TAT CAG GAC ATA GCG TTG GCT ACC CGT GAT ATT GCT GAA GAG CTT	1462
Arg Tyr Gln Asp Ile Ala Leu Ala Thr Arg Asp Ile Ala Glu Glu Leu	
220 225 230	
GGC GGC GAA TGG GCT GAC CGC TTC CTC GTG CTT TAC GGT ATC GCC GCT	1510
Gly Gly Glu Trp Ala Asp Arg Phe Leu Val Leu Tyr Gly Ile Ala Ala	
235 240 245	
CCC GAT TCG CAG CGC ATC GCC TTC TAT CGC CTT CTT GAC GAG TTC TTC	1558
Pro Asp Ser Gln Arg Ile Ala Phe Tyr Arg Leu Leu Asp Glu Phe Phe	
250 255 260 264	
TGAGCGGGAC TCTGGGGTTC GAAATGACCG ACCAAGCGAC GCCCTG GCC GCG GTG	1613
Ala Ala Val	
225	
ATT GCA TTC ATG TGT GCT GAG GAG TCA CGT TGG ATC AAC GGC ATA AAT	1661
Ile Ala Phe Met Cys Ala Glu Glu Ser Arg Trp Ile Asn Gly Ile Asn	
230 235 240	

ATT CCA GTG GAC GGA GGT TTG GCA TCG ACC TAC GTG TAA GTTCGTGGAC 1710
Ile Pro Val Asp Gly Gly Leu Ala Ser Thr Tyr Val
245 250 255

GCCCTTTGCA CGCGCACTAT ATCTCTATGC AGCAGCTGAA AGCAGCTTTG GTTTTGATCG 1770

GAGGTAGCGG GCGGAAAGGT GCAGAATGTC TAAATAATAA AGGATTCTTG TGAAGCTTTA 1830

GTTGTCCGTA AACGAAAATA AAAATAAAGA GGAATGATAT GAAAGCAAGT AGATCAGTCT 1890

GCACTTTCAA AATAGCTACC CTGGCAGGCG CCATTTATGC AGCGCTGCCA ATGTCAGCTG 1950

CAAACGAT GCAGCTGGAT GTAGGTAGCT CGGATTGGAC GGTGCGTTGG GGACAACACC 2010

CTCAAGTATA GCCTTGCCCTC TCGCCTGAAT GAGCAAGACT CAAGTCTGAC AAATGCGCCG 2070

ACTGTCAATG GTTATATCCG GATATTCAAA GTCAGGGTGA TCGTAACTTT GACCGGGGGC 2130

TTGGTATCCA ATCGTCTCGA TATTCTGGCT GCAG 2164

FIG. 2a:

CTGCAGCCAG GGCTGAAAAG GAGGGATTCA GTGAGGTCAT GAAGGGAGGG GACGGCGCCT	60
GGCTCCAATT GCTCGATGGC GCCGCGATTG AGTGTCTTGG GCGCGGTCTT GGAGAGTTCG	120
GCTAGGGAGA TAAATTTGCT GGCCATGGTG GCGGCCCTG ATGGGTGGA TGATTTTCTG	180
CATTCTGCAT CATGAAATTC ATGAAATCAT CACTTTTCGG GGGGTGGGTG CACGGGATTG	240
AAGGTTGCTA GGAGAGTGCA TTGCTCGTAA GCCCAGGAAG CACGCGGGTT TCAGGATGGT	300
GCATGGAAAT GGCATGAGCT TTGCTGGATA TGATTAGAGA CATTAAGTAT TTTGGCGGAA	360
TGGAAGCACG ATTCCTCGCC CGGTAGAGCG GTAACCGCGA CATTCAGGAC CGTAAAAAGG	420
AAAGAGCATG CAA CTG ACC AAC AAG AAA ATC GTC GTC ACC GGA GTG TCC TCC	472
Met Gln Leu Thr Asn Lys Lys Ile Val Val Thr Gly Val Ser Ser	
1 5 10 15	
GGT ATC GGT GCC GAA ACT GCC CGC GTT CTG CGC TCT CAC GGC GCC ACA	520
Gly Ile Gly Ala Glu Thr Ala Arg Val Leu Arg Ser His Gly Ala Thr	
20 25 30	
GTG ATT GGC GTA GAT CGC AAC ATG CCG AGC CTG ACT CTG GAT GCT TTC	568
Val Ile Gly Val Asp Arg Asn Met Pro Ser Leu Thr Leu Asp Ala Phe	
35 40 45	
GTT CAG GCT GAC CTG AGC CAT CCT GAGGGGAGAG GCGGTTTGCG TATTGGGCGC	622
Val Gln Ala Asp Leu Ser His Pro	
50 55	
ATGCATAAAA ACTGTTGTAA TTCATTAAGC ATTCTGCCGA CATGGAAGCC ATCACAAACG	682
GCATGATGAA CCTGAATCGC CAGCGGCATC AGCACCTTGT CGCCTTGCGT ATAATATTTG	742
CCCATGGACG CACACCGTGG AAACGGATGA AGGCACGAAC CCAGTTGACA TAAGCCTGTT	802
CGGTTCTGTA ACTGTAATGC AAGTAGCGTA TGCGCTCACG CAACTGGTCC AGAACCTTGA	862
CCGAACGCAG CGGTGGTAAC GGCGCAGTGG CGGTTTTTCAT GGCTTGTTAT GACTGTTTTT	922
TTGTACAGTC TATGCCTCGG GCATCCAAGC AGCAAGCGCG TTACGCCGTG GGTGATGTT	982
TGATGTTATG GAGCAGCAAC G ATG TTA CGC AGC AGC AAC GAT GTT ACG CAG	1033
Met Leu Arg Ser Ser Asn Asp Val Thr Gln	
1 5 10	
CAG GGC AGT CGC CCT AAA ACA AAG TTA GGT GGC TCA AGT ATG GGC ATC	1081
Gln Gly Ser Arg Pro Lys Thr Lys Leu Gly Gly Ser Ser Met Gly Ile	
15 20 25	
ATT CGC ACA TGT AGG CTC GGC CCT GAC CAA GTC AAA TCC ATG CGG GCT	1129
Ile Arg Thr Cys Arg Leu Gly Pro Asp Gln Val Lys Ser Met Arg Ala	
30 35 40	
GCT CTT GAT CTT TTC GGT CGT GAG TTC GGA GAC GTA GCC ACC TAC TCC	1177
Ala Leu Asp Leu Phe Gly Arg Glu Phe Gly Asp Val Ala Thr Tyr Ser	
45 50 55	

CAA CAT CAG CCG GAC TCC GAT TAC CTC GGG AAC TTG CTC CGT AGT AAG Gln His Gln Pro Asp Ser Asp Tyr Leu Gly Asn Leu Leu Arg Ser Lys 60 65 70	1225
ACA TTC ATC GCG CTT GCT GCC TTC GAC CAA GAA GCG GTT GTT GGC GCT Thr Phe Ile Ala Leu Ala Ala Phe Asp Gln Glu Ala Val Val Gly Ala 75 80 85 90	1273
CTC GCG GCT TAC GTT CTG CCC AGG TTT GAG CAG CCG CGT AGT GAG ATC Leu Ala Ala Tyr Val Leu Pro Arg Phe Glu Gln Pro Arg Ser Glu Ile 95 100 105	1321
TAT ATC TAT GAT CTC GCA GTC TCC GGC GAG CAC CGG AGG CAG GGC ATT Tyr Ile Tyr Asp Leu Ala Val Ser Gly Glu His Arg Arg Gln Gly Ile 110 115 120	1369
GCC ACC GCG CTC ATC AAT CTC CTC AAG CAT GAG GCC AAC GCG CTT GGT Ala Thr Ala Leu Ile Asn Leu Leu Lys His Glu Ala Asn Ala Leu Gly 125 130 135	1417
GCT TAT GTG ATC TAC GTG CAA GCA GAT TAC GGT GAC GAT CCC GCA GTG Ala Tyr Val Ile Tyr Val Gln Ala Asp Tyr Gly Asp Asp Pro Ala Val 140 145 150	1465
GCT CTC TAT ACA AAG TTG GGC ATA CGG GAA GAA GTG ATG CAC TTT GAT Ala Leu Tyr Thr Lys Leu Gly Ile Arg Glu Glu Val Met His Phe Asp 155 160 165 170	1513
ATC GAC CCA AGT ACC GCC ACC TAA CAATTCGTTC AAGCCGAGAT CGGCTTCCCT Ile Asp Pro Ser Thr Ala Thr 175 177	1567
G ATT GCA TTC ATG TGT GCT GAG GAG TCA CGT TGG ATC AAC GGC ATA AAT Ile Ala Phe Met Cys Ala Glu Glu Ser Arg Trp Ile Asn Gly Ile Asn 228 230 235 240	1616
ATT CCA GTG GAC GGA GGT TTG GCA TCG ACC TAC GTG TAA GTTCGTGGAC Ile Pro Val Asp Gly Gly Leu Ala Ser Thr Tyr Val 245 250 255	1665
GCCCTTTGCA CGCGCACTAT ATCTCTATGC AGCAGCTGAA AGCAGCTTTG GTTTTGATCG	1725
GAGGTAGCGG GCGGAAAGGT GCAGAATGTC TAAATAATAA AGGATTCTTG TGAAGCTTTA	1785
GTTGTCCGTA AACGAAAATA AAAATAAAGA GGAATGATAT GAAAGCAAGT AGATCAGTCT	1845
GCACTTTCAA AATAGCTACC CTGGCAGGCG CCATTTATGC AGCGCTGCCA ATGTCAGCTG	1905
CAAACCTCGAT GCAGCTGGAT GTAGGTAGCT CGGATTGGAC GGTGCGTTGG GGACAACACC	1965
CTCAAGTATA GCCTTGCCCTC TCGCCTGAAT GAGCAAGACT CAAGTCTGAC AAATGCGCCG	2025
ACTGTCAATG GTTATATCCG GATATTCAA GTCAGGGTGA TCGTAACTTT GACCGGGGGC	2085
TTGGTATCCA ATCGTCTCGA TATTCTGGCT GCAG	2119

FIG. 2b:

CTGCAGCCAG GGCTGAAAAG GAGGGATTCA GTGAGGTCAT GAAGGGAGGG GACGGCGCCT	60
GGCTCCAATT GCTCGATGGC GCCGCGATTG AGTGTCTTGG GCGCGGTCTT GGAGAGTTCG	120
GCTAGGGAGA TAAATTTGCT GGCCATGGTG GCGGCCCCCTG ATGGGTTGGA TGATTTTCTG	180
CATTCTGCAT CATGAAATTC ATGAAATCAT CACTTTTTCGG GGGGTGGGTG CACGGGATTG	240
AAGGTTGCTA GGAGAGTGCA TTGCTCGTAA GCCCAGGAAG CACGCGGGTT TCAGGATGGT	300
GCATGGAAAT GGCATGAGCT TTGCTGGATA TGATTAGAGA CATTA ACTAT TTTGGCGGAA	360
TGGAAGCACG ATTCCTCGCC CGGTAGAGCG GTAACCGCGA CATT CAGGAC CGTAAAAAGG	420
AAAGAGCATG CAA CTG ACC AAC AAG AAA ATC GTC GTC ACC GGA GTG TCC TCC	472
Met Gln Leu Thr Asn Lys Lys Ile Val Val Thr Gly Val Ser Ser	
1 5 10 15	
GGT ATC GGT GCC GAA ACT GCC CGC GTT CTG CGC TCT CAC GGC GCC ACA	520
Gly Ile Gly Ala Glu Thr Ala Arg Val Leu Arg Ser His Gly Ala Thr	
20 25 30	
GTG ATT GGC GTA GAT CGC AAC ATG CCG AGC CTG ACT CTG GAT GCT TTC	568
Val Ile Gly Val Asp Arg Asn Met Pro Ser Leu Thr Leu Asp Ala Phe	
35 40 45	
GTT CAG GCT GAC CTG AGC CAT CCT GAA GGC ATC GATC AAC GGC ATA AAT	617
Val Gln Ala Asp Leu Ser His Pro Glu Gly Ile Asn Gly Ile Asn	
50 55 58 240	
ATT CCA GTG GAC GGA GGT TTG GCA TCG ACC TAC GTG TAA GTTCGTGGAC	666
Ile Pro Val Asp Gly Gly Leu Ala Ser Thr Tyr Val	
245 250 255	
GCCCTTTGCA CGCGCACTAT ATCTCTATGC AGCAGCTGAA AGCAGCTTTG GTTTTGATCG	726
GAGGTAGCGG GCGGAAAGGT GCAGAATGTC TAAATAATAA AGGATTCTTG TGAAGCTTTA	786
GTTGTCCGTA AACGAAAATA AAAATAAAGA GGAATGATAT GAAAGCAAGT AGATCAGTCT	846
GCACTTTCAA AATAGCTACC CTGGCAGGCG CCATTTATGC AGCGCTGCCA ATGTCAGCTG	906
CAAACTCGAT GCAGCTGGAT GTAGGTAGCT CGGATTGGAC GGTGCGTTGG GGACAACACC	966
CTCAAGTATA GCCTTGCCCTC TCGCCTGAAT GAGCAAGACT CAAGTCTGAC AAATGCGCCG	1026
ACTGTCAATG GTTATATCCG GATATTCAAA GTCAGGGTGA TCGTAACTTT GACCGGGGGC	1086
TTGGTATCCA ATCGTCTCGA TATTCTGGCT GCAG	1120

FIG. 2c:

GAATTCGCG TATCGCCCGG TTCTATCAGC GGGCCGCTTT CGAAAGTCAT GGTGTTAGCC	60
GGTAGGGTCT TTTTCTTGGC CATGCTTGTT GCCTGAACCT TCGTTGACAT AGGGCAGAGG	120
TGCGTTTGCC GCTTCGCTTC GCGATGAACC GCATCGAGAT GCTGAGGTCA GGATTTTTCC	180
TTAACTCGCG TAAGCATTCT GTCATTTTTT TGGTGGCTTT GAACAGCCTG ATGAAAGGTG	240
GTCTCGCCCT TTGAGGCCGA TTCTTGGGCG CTTGGCGGCG TCGAAGCGAT GCTCCACTAC	300
CGATTAAGAT AATTAAAATA AGGAAACCGC ATGGTTTCTT ATGTGAATTT GTCTGGCATA	360
CTCCAGCTCA AGGGCAATTT TTGGGCTATT GGCTGAGCAG TTGCCTCTAT ATGGTTATTC	420
AGAATAACAA TTGACTCCTC AGGAGGTCAG CG ATG AGC ATT CTT GGT TTG AAT	473
Met Ser Ile Leu Gly Leu Asn	
1 5	
GGT GCC CCG GTC GGA GCT GAG CAG CTG GGC TCG GCT CTT GAT CGC ATG	521
Gly Ala Pro Val Gly Ala Glu Gln Leu Gly Ser Ala Leu Asp Arg Met	
10 15 20	
AAG AAG GCG CAC CTG GAG CAG GGG CCT GCA AAC TTG GAG CTG CGT CTG	569
Lys Lys Ala His Leu Glu Gln Gly Pro Ala Asn Leu Glu Leu Arg Leu	
25 30 35	
AGT AGG CTG GAT CGT GCG ATT GCA ATG CTT CTG GAA AAT CGT GAA GCA	617
Ser Arg Leu Asp Arg Ala Ile Ala Met Leu Leu Glu Asn Arg Glu Ala	
40 45 50 55	
ATT GCC GAC GCG GTT TCT GCT GAC TTT GGC AAT CGC AGC CGT GAG CAA	665
Ile Ala Asp Ala Val Ser Ala Asp Phe Gly Asn Arg Ser Arg Glu Gln	
60 65 70	
ACA CTG CTT TGC GAC ATT GCT GGC TCG GTG GCA AGC CTG AAG GAT AGC	713
Thr Leu Leu Cys Asp Ile Ala Gly Ser Val Ala Ser Leu Lys Asp Ser	
75 80 85	
CGC GAG CAC GTG GCC AAA TGG ATG GAG CCC GAA CAT CAC AAG GCG ATG	761
Arg Glu His Val Ala Lys Trp Met Glu Pro Glu His His Lys Ala Met	
90 95 100	
TTT CCA GGG GCG GAG GCA CGC GTT GAG TTT CAG CCG CTG GGT GTC GTT	809
Phe Pro Gly Ala Glu Ala Arg Val Glu Phe Gln Pro Leu Gly Val Val	
105 110 115	
GGG GTC ATT AGT CCC TGG AAC TTC CCT ATC GTA CTG GCC TTT GGG CCG	857
Gly Val Ile Ser Pro Trp Asn Phe Pro Ile Val Leu Ala Phe Gly Pro	
120 125 130 135	
CTG GCC GGC ATA TTC GCA GCA GGT AAT CGC GCC ATG CTC AAG CCG TCC	905
Leu Ala Gly Ile Phe Ala Ala Gly Asn Arg Ala Met Leu Lys Pro Ser	
140 145 150	
GAG CTT ACC CCG CGG ACT TCT GCC CTG CTT GCG GAG CTA ATT GCT CGT	953
Glu Leu Thr Pro Arg Thr Ser Ala Leu Leu Ala Glu Leu Ile Ala Arg	
155 160 165	

TAC TTC GAT GAA ACT GAG CTG ACT ACA GTG CTG GGC GAC GCT GAA GTC	1001
Tyr Phe Asp Glu Thr Glu Leu Thr Thr Val Leu Gly Asp Ala Glu Val	
170 175 180	
GGT GCG CTG TTC AGT GCT CAG CCT TTC GAT CAT CTG ATC TTC ACC GGC	1049
Gly Ala Leu Phe Ser Ala Gln Pro Phe Asp His Leu Ile Phe Thr Gly	
185 190 195	
GGC ACT GCC GTG GCC AAG CAC ATC ATG CGT GCC GCG GCG GAT AAC CTA	1097
Gly Thr Ala Val Ala Lys His Ile Met Arg Ala Ala Ala Asp Asn Leu	
200 205 210 215	
GTG CCC GTT ACC CTG GAA TTG GGT GGC AAA TCG CCG GTG ATC GTT TCC	1145
Val Pro Val Thr Leu Glu Leu Gly Gly Lys Ser Pro Val Ile Val Ser	
220 225 230	
CGC AGT GCA GAT ATG GCG GAC GTT GCA CAA CGG GTG TTG ACG GTG AAA	1193
Arg Ser Ala Asp Met Ala Asp Val Ala Gln Arg Val Leu Thr Val Lys	
235 240 245	
ACC TTC AAT GCC GGG CAA ATC TGT CTG GCA CCG GAC TAT GTG CTG CTG	1241
Thr Phe Asn Ala Gly Gln Ile Cys Leu Ala Pro Asp Tyr Val Leu Leu	
250 255 260	
CCG GAA GGGACAGCAA GCGAACC GGA ATTGCCAGCT GGGGCGCCCT CTGGTAAGGT	1297
Pro Glu	
265	
TGGGAAGCCC TGCAAAGTAA ACTGGATGGC TTTCTTGCCG CCAAGGATCT GATGGCGCAG	1357
GGGATCAAGA TCTGATCAAG AGACAGGATG AGGATCGTTT CGC ATG ATT GAA CAA	1412
Met Ile Glu Gln	
1	
GAT GGA TTG CAC GCA GGT TCT CCG GCC GCT TGG GTG GAG AGG CTA TTC	1460
Asp Gly Leu His Ala Gly Ser Pro Ala Ala Trp Val Glu Arg Leu Phe	
5 10 15 20	
GGC TAT GAC TGG GCA CAA CAG ACA ATC GGC TGC TCT GAT GCC GCC GTG	1508
Gly Tyr Asp Trp Ala Gln Gln Thr Ile Gly Cys Ser Asp Ala Ala Val	
25 30 35	
TTC CGG CTG TCA GCG CAG GGG CGC CCG GTT CTT TTT GTC AAG ACC GAC	1556
Phe Arg Leu Ser Ala Gln Gly Arg Pro Val Leu Phe Val Lys Thr Asp	
40 45 50	
CTG TCC GGT GCC CTG AAT GAA CTG CAG GAC GAG GCA GCG CGG CTA TCG	1604
Leu Ser Gly Ala Leu Asn Glu Leu Gln Asp Glu Ala Ala Arg Leu Ser	
55 60 65	
TGG CTG GCC ACG ACG GGC GTT CCT TGC GCA GCT GTG CTC GAC GTT GTC	1652
Trp Leu Ala Thr Thr Gly Val Pro Cys Ala Ala Val Leu Asp Val Val	
70 75 80	

ACT	GAA	GCG	GGA	AGG	GAC	TGG	CTG	CTA	TTG	GGC	GAA	GTG	CCG	GGG	CAG	1700
Thr	Glu	Ala	Gly	Arg	Asp	Trp	Leu	Leu	Leu	Gly	Glu	Val	Pro	Gly	Gln	
85					90					95					100	
GAT	CTC	CTG	TCA	TCT	CAC	CTT	GCT	CCT	GCC	GAG	AAA	GTA	TCC	ATC	ATG	1748
Asp	Leu	Leu	Ser	Ser	His	Leu	Ala	Pro	Ala	Glu	Lys	Val	Ser	Ile	Met	
			105						110					115		
GCT	GAT	GCA	ATG	CGG	CGG	CTG	CAT	ACG	CTT	GAT	CCG	GCT	ACC	TGC	CCA	1796
Ala	Asp	Ala	Met	Arg	Arg	Leu	His	Thr	Leu	Asp	Pro	Ala	Thr	Cys	Pro	
			120					125					130			
TTC	GAC	CAC	CAA	GCG	AAA	CAT	CGC	ATC	GAG	CGA	GCA	CGT	ACT	CGG	ATG	1844
Phe	Asp	His	Gln	Ala	Lys	His	Arg	Ile	Glu	Arg	Ala	Arg	Thr	Arg	Met	
		135					140					145				
GAA	GCC	GGT	CTT	GTC	GAT	CAG	GAT	GAT	CTG	GAC	GAA	GAG	CAT	CAG	GGG	1892
Glu	Ala	Gly	Leu	Val	Asp	Gln	Asp	Asp	Leu	Asp	Glu	Glu	His	Gln	Gly	
150						155					160					
CTC	GCG	CCA	GCC	GAA	CTG	TTC	GCC	AGG	CTC	AAG	GCG	CGC	ATG	CCC	GAC	1940
Leu	Ala	Pro	Ala	Glu	Leu	Phe	Ala	Arg	Leu	Lys	Ala	Arg	Met	Pro	Asp	
165					170					175					180	
GGC	GAG	GAT	CTC	GTC	GTG	ACC	CAT	GGC	GAT	GCC	TGC	TTG	CCG	AAT	ATC	1988
Gly	Glu	Asp	Leu	Val	Val	Thr	His	Gly	Asp	Ala	Cys	Leu	Pro	Asn	Ile	
			185					190						195		
ATG	GTG	GAA	AAT	GGC	CGC	TTT	TCT	GGA	TTC	ATC	GAC	TGT	GGC	CGG	CTG	2036
Met	Val	Glu	Asn	Gly	Arg	Phe	Ser	Gly	Phe	Ile	Asp	Cys	Gly	Arg	Leu	
			200					205					210			
GGT	GTG	GCG	GAC	CGC	TAT	CAG	GAC	ATA	GCG	TTG	GCT	ACC	CGT	GAT	ATT	2084
Gly	Val	Ala	Asp	Arg	Tyr	Gln	Asp	Ile	Ala	Leu	Ala	Thr	Arg	Asp	Ile	
		215					220					225				
GCT	GAA	GAG	CTT	GGC	GGC	GAA	TGG	GCT	GAC	CGC	TTC	CTC	GTG	CTT	TAC	2132
Ala	Glu	Glu	Leu	Gly	Gly	Glu	Trp	Ala	Asp	Arg	Phe	Leu	Val	Leu	Tyr	
230						235					240					
GGT	ATC	GCC	GCT	CCC	GAT	TCG	CAG	CGC	ATC	GCC	TTC	TAT	CGC	CTT	CTT	2180
Gly	Ile	Ala	Ala	Pro	Asp	Ser	Gln	Arg	Ile	Ala	Phe	Tyr	Arg	Leu	Leu	
245					250				255					260		
GAC	GAG	TTC	TTC	TGA	GCGGGACTCT	GGGGTTCGAA	ATGACCGACC	AAGCGACGCC								2235
Asp	Glu	Phe	Phe													
			264													
CGC	CAT	GCC	AAG	CCT	GTT	CTC	GTG	CAA	AGT	CCT	GTG	GGT	GAG	TCG	AAC	2283
	His	Ala	Lys	Pro	Val	Leu	Val	Gln	Ser	Pro	Val	Gly	Glu	Ser	Asn	
	444	445					450					455				
TTG	GCG	ATG	CGC	GCA	CCC	TAC	GGA	GAA	GCG	ATC	CAC	GGA	CTG	CTC	TCT	2331
Leu	Ala	Met	Arg	Ala	Pro	Tyr	Gly	Glu	Ala	Ile	His	Gly	Leu	Leu	Ser	
460						465					470					

GTC CTC CTT TCA ACG GAG TGT TAG AACCGTTGGT AGTGGTTTTG GACGGGCCCCA 2385
Val Leu Leu Ser Thr Glu Cys
475 480 481

GGAGCATGCG CTTCTGGGCC CGTTTCTTGA GTATTCATTG GATAGTCACG CGTGGTAGCT 2445

TCGAGCCTGC ACAGCTGATG AGCACCCCTGG AAGGCGCGCT GTACGCGGAC GACTGGGTTC 2505

ATCTTCGCCA TTCATGACGG AACTCCGTTC CCCAGTACCG CGATGACTAT TTTGCCTCTT 2565

CCGATGTCCG ATTCCACGCC GCCTGACGCT AAGCGGGGGC GGGGGCGCCC GCATCCCAGC 2625

CCAGACAGCA ACAAATGAGT AGGCTCTTGG ATGCCGCGGC GGCTGAGATT GGTAACGGCA 2685

ATTTGCTCAA TGTGACGATG GATTGATTG CCCGTGCTGC CGGCGTCTCA AAAAAACGC 2745

TGTACGTCTT GGTGGCGAGC AAGGAAGAAC TCATTCCCG GTTAGTGGCT CGAGACATGT 2805

CCAACCTTGA GGAATTC 2822

FIG. 2d:

GAATTCGCG TATCGCCCGG TTCTATCAGC GGGCCGCTTT CGAAAGTCAT GGTGTTAGCC	60
GGTAGGGTCT TTTTCTTGGC CATGCTTGTT GCCTGAACCT TCGTTGACAT AGGGCAGAGG	120
TGCGTTTGCC GCTTCGCTTC GCGATGAACC GCATCGAGAT GCTGAGGTCA GGATTTTTCC	180
TTAACTCGCG TAAGCATTCT GTCATTTTTT TGGTGGCTTT GAACAGCCTG ATGAAAGGTG	240
GTCTCGCCCT TTGAGGCCGA TTCTTGGGCG CTTGGCGGCG TCGAAGCGAT GCTCCACTAC	300
CGATTAAGAT AATTAAAATA AGGAAACCGC ATGGTTTCTT ATGTGAATTT GTCTGGCATA	360
CTCCAGCTCA AGGGCAATTT TTGGGCTATT GGCTGAGCAG TTGCCTCTAT ATGGTTATTC	420
AGAATAACAA TTGACTCCTC AGGAGGTCAG CG ATG AGC ATT CTT GGT TTG AAT	473
Met Ser Ile Leu Gly Leu Asn	
1 5	
GGT GCC CCG GTC GGA GCT GAG CAG CTG GGC TCG GCT CTT GAT CGC ATG	521
Gly Ala Pro Val Gly Ala Glu Gln Leu Gly Ser Ala Leu Asp Arg Met	
10 15 20	
AAG AAG GCG CAC CTG GAG CAG GGG CCT GCA AAC TTG GAG CTG CGT CTG	569
Lys Lys Ala His Leu Glu Gln Gly Pro Ala Asn Leu Glu Leu Arg Leu	
25 30 35	
AGT AGG CTG GAT CGT GCG ATT GCA ATG CTT CTG GAA AAT CGT GAA GCA	617
Ser Arg Leu Asp Arg Ala Ile Ala Met Leu Leu Glu Asn Arg Glu Ala	
40 45 50 55	
ATT GCC GAC GCG GTT TCT GCT GAC TTT GGC AAT CGC AGC CGT GAG CAA	665
Ile Ala Asp Ala Val Ser Ala Asp Phe Gly Asn Arg Ser Arg Glu Gln	
60 65 70	
ACA CTG CTT TGC GAC ATT GCT GGC TCG GTG GCA AGC CTG AAG GAT AGC	713
Thr Leu Leu Cys Asp Ile Ala Gly Ser Val Ala Ser Leu Lys Asp Ser	
75 80 85	
CGC GAG CAC GTG GCC AAA TGG ATG GAG CCC GAA CAT CAC AAG GCG ATG	761
Arg Glu His Val Ala Lys Trp Met Glu Pro Glu His His Lys Ala Met	
90 95 100	
TTT CCA GGG GCG GAG GCA CGC GTT GAG TTT CAG CCG CTG GGT GTC GTT	809
Phe Pro Gly Ala Glu Ala Arg Val Glu Phe Gln Pro Leu Gly Val Val	
105 110 115	
GGG GTC ATT AGT CCC TGG AAC TTC CCT ATC GTA CTG GCC TTT GGG CCG	857
Gly Val Ile Ser Pro Trp Asn Phe Pro Ile Val Leu Ala Phe Gly Pro	
120 125 130 135	
CTG GCC GGC ATA TTC GCA GCA GGT AAT CGC GCC ATG CTC AAG CCG TCC	905
Leu Ala Gly Ile Phe Ala Ala Gly Asn Arg Ala Met Leu Lys Pro Ser	
140 145 150	
GAG CTT ACC CCG CGG ACT TCT GCC CTG CTT GCG GAG CTA ATT GCT CGT	953
Glu Leu Thr Pro Arg Thr Ser Ala Leu Leu Ala Glu Leu Ile Ala Arg	
155 160 165	

TAC TTC GAT GAA ACT GAG CTG ACT ACA GTG CTG GGC GAC GCT GAA GTC Tyr Phe Asp Glu Thr Glu Leu Thr Thr Val Leu Gly Asp Ala Glu Val 170 175 180	1001
GGT GCG CTG TTC AGT GCT CAG CCT TTC GAT CAT CTG ATC TTC ACC GGC Gly Ala Leu Phe Ser Ala Gln Pro Phe Asp His Leu Ile Phe Thr Gly 185 190 195	1049
GGC ACT GCC GTG GCC AAG CAC ATC ATG CGT GCC GCG GCG GAT AAC CTA Gly Thr Ala Val Ala Lys His Ile Met Arg Ala Ala Ala Asp Asn Leu 200 205 210 215	1097
GTG CCC GTT ACC CTG GAA TTG GGT GGC AAA TCG CCG GTG ATC GTT TCC Val Pro Val Thr Leu Glu Leu Gly Gly Lys Ser Pro Val Ile Val Ser 220 225 230	1145
CGC AGT GCA GAT ATG GCG GAC GTT GCA CAA CGG GTG TTG ACG GTG AAA Arg Ser Ala Asp Met Ala Asp Val Ala Gln Arg Val Leu Thr Val Lys 235 240 245	1193
ACC TTC AAT GCC GGG CAA ATC TGT CTG GCA CCG GAC TAT GTG CTG GGG Thr Phe Asn Ala Gly Gln Ile Cys Leu Ala Pro Asp Tyr Val Leu 250 255 260 262	1241
GAGAGGCGGT TTGCGTATTG GGCGCATGCA TAAAACTGT TGTAATTCAT TAAGCATTCT	1301
GGCGACATGG AAGCCATCAC AAACGGCATG ATGAACCTGA ATCGCCAGCG GCATCAGCAC	1361
CTTGTCGCCT TCGGTATAAT ATTTGCCCAT GGACGCACAC CGTGGAACG GATGAAGGCA	1421
CGAACCCAGT TGACATAAGC CTGTTGCGTT CGTAACTGT AATGCAAGTA GCGTATGCGC	1481
TCACGCAACT GGTCCAGAAC CTTGACCGAA CGCAGCGGTG GTAACGGCGC AGTGGCGGTT	1541
TTCATGGCTT GTTATGACTG TTTTTTTGTA CAGTCTATGC CTCGGGCATC CAAGCAGCAA	1601
GCGCGTTACG CCGTGGGTCG ATGTTTGATG TTATGGAGCA GCAACG ATG TTA CGC Met Leu Arg 1	1656
AGC AGC AAC GAT GTT ACG CAG CAG GGC AGT CGC CCT AAA ACA AAG TTA Ser Ser Asn Asp Val Thr Gln Gln Gly Ser Arg Pro Lys Thr Lys Leu 5 10 15	1704
GGT GGC TCA AGT ATG GGC ATC ATT CGC ACA TGT AGG CTC GGC CCT GAC Gly Gly Ser Ser Met Gly Ile Ile Arg Thr Cys Arg Leu Gly Pro Asp 20 25 30 35	1752
CAA GTC AAA TCC ATG CGG GCT GCT CTT GAT CTT TTC GGT CGT GAG TTC Gln Val Lys Ser Met Arg Ala Ala Leu Asp Leu Phe Gly Arg Glu Phe 40 45 50	1800
GGA GAC GTA GCC ACC TAC TCC CMA CAT CAG CCG GAC TCC GAT TAC CTC Gly Asp Val Ala Thr Tyr Ser Gln His Gln Pro Asp Ser Asp Tyr Leu 55 60 65	1848
GGG AAC TTG CTC CGT AGT AAG ACA TTC ATC GCG CTT GCT GCC TTC GAC Gly Asn Leu Leu Arg Ser Lys Thr Phe Ile Ala Leu Ala Ala Phe Asp 70 75 80	1896

CAA GAA GCG GTT GTT GGC GCT CTC GCG GCT TAC GTT CTG CCC AGG TTT Gln Glu Ala Val Val Gly Ala Leu Ala Ala Tyr Val Leu Pro Arg Phe 85 90 95	1944
GAG CAG CCG CGT AGT GAG ATC TAT ATC TAT GAT CTC GCA GTC TCC GGC Glu Gln Pro Arg Ser Glu Ile Tyr Ile Tyr Asp Leu Ala Val Ser Gly 100 105 110 115	1992
GAG CAC CGG AGG CAG GGC ATT GCC ACC GCG CTC ATC AAT CTC CTC AAG Glu His Arg Arg Gln Gly Ile Ala Thr Ala Leu Ile Asn Leu Leu Lys 120 125 130	2040
CAT GAG GCC AAC GCG CTT GGT GCT TAT GTG ATC TAC GTG CAA GCA GAT His Glu Ala Asn Ala Leu Gly Ala Tyr Val Ile Tyr Val Gln Ala Asp 135 140 145	2088
TAC GGT GAC GAT CCC GCA GTG GCT CTC TAT ACA AAG TTG GGC ATA CGG Tyr Gly Asp Asp Pro Ala Val Ala Leu Tyr Thr Lys Leu Gly Ile Arg 150 155 160	2136
GAA GAA GTG ATG CAC TTT GAT ATC GAC CCA AGT ACC GCC ACC TAA CAA Glu Glu Val Met His Phe Asp Ile Asp Pro Ser Thr Ala Thr 165 170 175 177	2184
TTCGTTCAAG CCGAGATCGG CTTCCCTG CAA AGT CCT GTG GGT GAG TCG AAC Gln Ser Pro Val Gly Glu Ser Asn 451 455	2236
TTG GCG ATG CGC GCA CCC TAC GGA GAA GCG ATC CAC GGA CTG CTC TCT Leu Ala Met Arg Ala Pro Tyr Gly Glu Ala Ile His Gly Leu Leu Ser 460 465 470	2284
GTC CTC CTT TCA ACG GAG TGT TAG AACCGTTGGT AGTGGTTTTG GACGGGCCCA Val Leu Leu Ser Thr Glu Cys 475 480 481	2338
GGAGCATGCG CTTCTGGGCC CGTTTCTTGA GTATTCATTG GATAGTCACG CGTGGTAGCT	2398
TCGAGCCTGC ACAGCTGATG AGCACCCCTGG AAGGCGCGCT GTACGCGGAC GACTGGGTTC	2458
ATCTTCGCCA TTCATGACGG AACTCCGTTT CCCAGTACCG CGATGACTAT TTTGCCTCTT	2518
CCGATGTCCG ATTCCACGCC GCCTGACGCT AAGCGGGGGC GGGGGCGCCC GCATCCCAGC	2578
CCAGACAGCA ACAAATGAGT AGGCTCTTGG ATGCCGCGGC GGCTGAGATT GGTAACGGCA	2638
ATTCGTCAA TGTGACGATG GATTGATTG CCCGTGCTGC CGGCGTCTCA AAAAAACGC	2698
TGTACGTCTT GGTGGCGAGC AAGGAAGAAC TCATTTCCCG GTTAGTGGCT CGAGACATGT	2758
CCAACCTTGA GGAATTC	2775

FIG. 2e:

GAATTCCGCG TATCGCCCGG TTCTATCAGC GGGCCGCTTT CGAAAGTCAT GGTGTTAGCC	60
GGTAGGGTCT TTTTCTTGGC CATGCTTGTT GCCTGAACCT TCGTTGACAT AGGGCAGAGG	120
TGCGTTTGCC GCTTCGCTTC GCGATGAACC GCATCGAGAT GCTGAGGTCA GGATTTTTC	180
TTAACTCGCG TAAGCATTCT GTCATTTTTT TGGTGGCTTT GAACAGCCTG ATGAAAGGTG	240
GTCTCGCCCT TTGAGGCCGA TTCTTGGGCG CTTGGCGGCG TCGAAGCGAT GCTCCACTAC	300
CGATTAAGAT AATTAAATA AGGAAACCGC ATGGTTTCTT ATGTGAATTT GTCTGGCATA	360
CTCCAGCTCA AGGGCAATTT TTGGGCTATT GGCTGAGCAG TTGCCTCTAT ATGGTTATTC	420
AGAATAACAA TTGACTCCTC AGGAGGTCAG CG ATG AGC ATT CTT GGT TTG AAT	473
Met Ser Ile Leu Gly Leu Asn	
1 5	
GGT GCC CCG GTC GGA GCT GAG CAG CTG GGC TCG GCT CTT GAT CGC ATG	521
Gly Ala Pro Val Gly Ala Glu Gln Leu Gly Ser Ala Leu Asp Arg Met	
10 15 20	
AAG AAG GCG CAC CTG GAG CAG GGG CCT GCA AAC TTG GAG CTG CGT CTG	569
Lys Lys Ala His Leu Glu Gln Gly Pro Ala Asn Leu Glu Leu Arg Leu	
25 30 35	
AGT AGG CTG GAT CGT GCG ATT GCA ATG CTT CTG GAA AAT CGT GAA GCA	617
Ser Arg Leu Asp Arg Ala Ile Ala Met Leu Leu Glu Asn Arg Glu Ala	
40 45 50 55	
ATT GCC GAC GCG GTT TCT GCT GAC TTT GGC AAT CGC AGC CGT GAG CAA	665
Ile Ala Asp Ala Val Ser Ala Asp Phe Gly Asn Arg Ser Arg Glu Gln	
60 65 70	
ACA CTG CTT TGC GAC ATT GCT GGC TCG GTG GCA AGC CTG AAG GAT AGC	713
Thr Leu Leu Cys Asp Ile Ala Gly Ser Val Ala Ser Leu Lys Asp Ser	
75 80 85	
CGC GAG CAC GTG GCC AAA TGG ATG GAG CCC GAA CAT CAC AAG GCG ATG	761
Arg Glu His Val Ala Lys Trp Met Glu Pro Glu His His Lys Ala Met	
90 95 100	
TTT CCA GGG GCG GAG GCA CGC GTT GAG TTT CAG CCG CTG GGT GTC GTT	809
Phe Pro Gly Ala Glu Ala Arg Val Glu Phe Gln Pro Leu Gly Val Val	
105 110 115	
GGG GTC ATT AGT CCC TGG AAC TTC CCT ATC GTA CTG GCC TTT GGG CCG	857
Gly Val Ile Ser Pro Trp Asn Phe Pro Ile Val Leu Ala Phe Gly Pro	
120 125 130 135	
CTG GCC GGC ATA TTC GCA GCA GGT AAT CGC GCC ATG CTC AAG CCG TCC	905
Leu Ala Gly Ile Phe Ala Ala Gly Asn Arg Ala Met Leu Lys Pro Ser	
140 145 150	
GAG CTT ACC CCG CGG ACT TCT GCC CTG CTT GCG GAG CTA ATT GCT CGT	953
Glu Leu Thr Pro Arg Thr Ser Ala Leu Leu Ala Glu Leu Ile Ala Arg	
155 160 165	

TAC TTC GAT GAA ACT GAG CTG ACT ACA GTG CTG GGC GAC GCT GAA GTC	1001
Tyr Phe Asp Glu Thr Glu Leu Thr Thr Val Leu Gly Asp Ala Glu Val	
170 175 180	
GGT GCG CTG TTC AGT GCT CAG CCT TTC GAT CAT CTG ATC TTC ACC GGC	1049
Gly Ala Leu Phe Ser Ala Gln Pro Phe Asp His Leu Ile Phe Thr Gly	
185 190 195	
GGC ACT GCC GTG GCC AAG CAC ATC ATG CGT GCC GCG GCG GAT AAC CTA	1097
Gly Thr Ala Val Ala Lys His Ile Met Arg Ala Ala Ala Asp Asn Leu	
200 205 210 215	
GTG CCC GTT ACC CTG GAA TTG GGT GGC AAA TCG CCG GTG ATC GTT TCC	1145
Val Pro Val Thr Leu Glu Leu Gly Gly Lys Ser Pro Val Ile Val Ser	
220 225 230	
CGC AGT GCA GAT ATG GCG GAC GTT GCA CAA CGG GTG TTG ACG GTG AAA	1193
Arg Ser Ala Asp Met Ala Asp Val Ala Gln Arg Val Leu Thr Val Lys	
235 240 245	
ACC TTC AAT GCC GGG CAA ATC TGT CTG GCA CC GTG GGT GAG TCG AAC	1240
Thr Phe Asn Ala Gly Gln Ile Cys Leu Ala Val Gly Glu Ser Asn	
250 255 257 454 455	
TTG GCG ATG CGC GCA CCC TAC GGA GAA GCG ATC CAC GGA CTG CTC TCT	1288
Leu Ala Met Arg Ala Pro Tyr Gly Glu Ala Ile His Gly Leu Leu Ser	
460 465 470	
GTC CTC CTT TCA ACG GAG TGT TAG AACC GTTGGT AGTGGT TTTG GACGGG CCGCA	1342
Val Leu Leu Ser Thr Glu Cys	
475 480 481	
GGAGCATGCG CTTCTGGGCC CGTTTCTTGA GTATTCATTG GATAGTCACG CGTGGTAGCT	1402
TCGAGCCTGC ACAGCTGATG AGCACCCTGG AAGGCGCGCT GTACGCGGAC GACTGGGTTC	1462
ATCTTCGCCA TTCATGACGG AACTCCGTTT CCCAGTACCG CGATGACTAT TTTGCCTCTT	1522
CCGATGTCCG ATTCCACGCC GCCTGACGCT AAGCGGGGGC GGGGGCGCCC GCATCCCAGC	1582
CCAGACAGCA ACAAATGAGT AGGCTCTTGG ATGCCGCGGC GGCTGAGATT GGTAACGGCA	1642
ATTCGTCAA TGTGACGATG GATTCGATTG CCCGTGCTGC CGGCGTCTCA AAAAAACGC	1702
TGTACGTCTT GGTGGCGAGC AAGGAAGAAC TCATTTCCCG GTTAGTGGCT CGAGACATGT	1762
CCAACCTTGA GGAATTC	1779

FIG. 2f:

CTGCAGCCGA GCATCGATTG AGCACTTTAC CCAGCTGCGC TGGCTGACCA TTCAGAATGG	60
CCCCGCGGCAC TATCCAATCT AAATCGATCT TCGGGCGCCG CGGGCATCAT GCCCCGCGCG	120
CTCGCCTCAT TTCAATCTCT AACTTGATAA AAACAGAGCT GTTCTCCGGT CTTGGTGGAT	180
CAAGGCCAGT CGCGGAGAGT CTCGAAGAGG AGAGTACAGT GAACGCCGAG TCCACATTGC	240
AACCGCAGGC ATCATCATGC TCTGCTCAGC CACGCTACCG CAGTGTGTCG ATTGGTCATC	300
CTCCGGTTGA GGTTACGCAA GACGCTGGAG GTATTGTCCG G ATG CGT TCT CTC GAG	356
Met Arg Ser Leu Glu	
1 5	
GCG CTT CTT CCC TTC CCG GGT CGA ATT CTT GAG CGT CTC GAG CAT TGG	404
Ala Leu Leu Pro Phe Pro Gly Arg Ile Leu Glu Arg Leu Glu His Trp	
10 15 20	
GCT AAG ACC CGT CCA GAA CAA ACC TGC GTT GCT GCC AGG GCG GCA AAT	452
Ala Lys Thr Arg Pro Glu Gln Thr Cys Val Ala Ala Arg Ala Ala Asn	
25 30 35	
GGG GAA TGG CGT CGT ATC AGC TAC GCG GAA ATG TTC CAC AAC GTC CGC	500
Gly Glu Trp Arg Arg Ile Ser Tyr Ala Glu Met Phe His Asn Val Arg	
40 45 50	
GCC ATC GCA CAG AGC TTG CTT CCT TAC GGA CTA TCG GCA GAG CGT CCG	548
Ala Ile Ala Gln Ser Leu Leu Pro Tyr Gly Leu Ser Ala Glu Arg Pro	
55 60 65	
CTG CTT ATC GTC TCT GGA AAT GAC CTG GAA CAT CTT CAG CTG GCA TTT	596
Leu Leu Ile Val Ser Gly Asn Asp Leu Glu His Leu Gln Leu Ala Phe	
70 75 80 85	
GGG GCT ATG TAT GCG GGC ATT CCC TAT TGC CCG GTG TCT CCT GCT TAT	644
Gly Ala Met Tyr Ala Gly Ile Pro Tyr Cys Pro Val Ser Pro Ala Tyr	
90 95 100	
TCA CTG CTG TCG CAA GAT TTG GCG AAG CTG CGT CAC ATC GTA GGT CTT	692
Ser Leu Leu Ser Gln Asp Leu Ala Lys Leu Arg His Ile Val Gly Leu	
105 110 115	
CTG CAA CCG GGA CTG GTC TTT GCT GCC GAT GCA GCA CCT TTC CAG GGG	740
Leu Gln Pro Gly Leu Val Phe Ala Ala Asp Ala Ala Pro Phe Gln	
120 125 130 132	
ACAGCAAGCG AACCGGAATT GCCAGCTGGG GCGCCCTCTG GTAAGGTTGG GAAGCCCTGC	800
AAAGTAACT GGATGGCTTT CTTGCCGCCA AGGATCTGAT GGCGCAGGGG ATCAAGATCT	860
GATCAAGAGA CAGGATGAGG ATCGTTTCGC ATG ATT GAA CAA GAT GGA TTG CAC	914
Met Ile Glu Gln Asp Gly Leu His	
1 5	
GCA GGT TCT CCG GCC GCT TGG GTG GAG AGG CTA TTC GGC TAT GAC TGG	962
Ala Gly Ser Pro Ala Ala Trp Val Glu Arg Leu Phe Gly Tyr Asp Trp	
10 15 20	

GCA CAA CAG ACA ATC GGC TGC TCT GAT GCC GCC GTG TTC CGG CTG TCA	1010
Ala Gln Gln Thr Ile Gly Cys Ser Asp Ala Ala Val Phe Arg Leu Ser	
25 30 35 40	
GCG CAG GGG CGC CCG GTT CTT TTT GTC AAG ACC GAC CTG TCC GGT GCC	1058
Ala Gln Gly Arg Pro Val Leu Phe Val Lys Thr Asp Leu Ser Gly Ala	
45 50 55	
CTG AAT GAA CTG CAG GAC GAG GCA GCG CGG CTA TCG TGG CTG GCC ACG	1106
Leu Asn Glu Leu Gln Asp Glu Ala Ala Arg Leu Ser Trp Leu Ala Thr	
60 65 70	
ACG GGC GTT CCT TGC GCA GCT GTG CTC GAG GTT GTC ACT GAA GCG GGA	1154
Thr Gly Val Pro Cys Ala Ala Val Leu Asp Val Val Thr Glu Ala Gly	
75 80 85	
AGG GAC TGG CTG CTA TTG GGC GAA GTG CCG GGG CAG GAT CTC CTG TCA	1202
Arg Asp Trp Leu Leu Leu Gly Glu Val Pro Gly Gln Asp Leu Leu Ser	
90 95 100	
TCT CAC CTT GCT CCT GCC GAG AAA GTA TCC ATC ATG GCT GAT GCA ATG	1250
Ser His Leu Ala Pro Ala Glu Lys Val Ser Ile Met Ala Asp Ala Met	
105 110 115 120	
CGG CGG CTG CAT ACG CTT GAT CCG GCT ACC TGC CCA TTC GAC CAC CAA	1298
Arg Arg Leu His Thr Leu Asp Pro Ala Thr Cys Pro Phe Asp His Gln	
125 130 135	
GCG AAA CAT CGC ATC GAG CGA GCA CGT ACT CGG ATG GAA GCC GGT CTT	1346
Ala Lys His Arg Ile Glu Arg Ala Arg Thr Arg Met Glu Ala Gly Leu	
140 145 150	
GTC GAT CAG GAT GAT CTG GAC GAA GAG CAT CAG GGG CTC GCG CCA GCC	1394
Val Asp Gln Asp Asp Leu Asp Glu Glu His Gln Gly Leu Ala Pro Ala	
155 160 165	
GAA CTG TTC GCC AGG CTC AAG GCG CGC ATG CCC GAC GGC GAG GAT CTC	1442
Glu Leu Phe Ala Arg Leu Lys Ala Arg Met Pro Asp Gly Glu Asp Leu	
170 175 180	
GTC GTG ACC CAT GGC GAT GCC TGC TTG CCG AAT ATC ATG GTG GAA AAT	1490
Val Val Thr His Gly Asp Ala Cys Leu Pro Asn Ile Met Val Glu Asn	
185 190 195 200	
GGC CGC TTT TCT GGA TTC ATC GAC TGT GGC CGG CTG GGT GTG GCG GAC	1538
Gly Arg Phe Ser Gly Phe Ile Asp Cys Gly Arg Leu Gly Val Ala Asp	
205 210 215	
CGC TAT CAG GAC ATA GCG TTG GCT ACC CGT GAT ATT GCT GAA GAG CTT	1586
Arg Tyr Gln Asp Ile Ala Leu Ala Thr Arg Asp Ile Ala Glu Glu Leu	
220 225 230	
GGC GGC GAA TGG GCT GAC CGC TTC CTC GTG CTT TAC GGT ATC GCC GCT	1634
Gly Gly Glu Trp Ala Asp Arg Phe Leu Val Leu Tyr Gly Ile Ala Ala	
235 240 245	

CCC GAT TCG CAG CGC ATC GCC TTC TAT CGC CTT CTT GAC GAG TTC TTC	1682
Pro Asp Ser Gln Arg Ile Ala Phe Tyr Arg Leu Leu Asp Glu Phe Phe	
250 255 260 264	
TGAGCGGGAC TCTGGGGTTC GAAATGACCG ACCAAGCGAC GCCCCT GTT TTG CAA	1737
Val Leu Gln	
563 565	
TGG CGG TCG GCG AAA GTT GAT GCG CTG TAT CGT GGT GAA GAT CAA TCC	1785
Trp Arg Ser Ala Lys Val Asp Ala Leu Tyr Arg Gly Glu Asp Gln Ser	
570 575 580	
ATG CTG CGT GAC GAG GCC ACA CTG TGA GTTGGTCAGG GGGGGCTTAC	1832
Met Leu Arg Asp Glu Ala Thr Leu	
585 589	
TCGGCGTTTT CCGACACTGC GTTGGTTGCG GCAGTGC GCA CCCCCTGGAT TGATTGCGGG	1892
GGTGCCCTGT CGCTGGTGTC GCCTATCGAC TTAGGGGTAA AGGTCGCTCG CGAAGTTCTG	1952
ATGCGTGCGT CGCTTGAACC ACAAATGGTC GATAGCGTAC TCGCAGGCTC TATGGCTCAA	2012
GCAAGCTTTG ATGCTTACCT GCTCCCGCGG CACATTGGCT TGTACAGCGG TGTTCCCAAG	2072
TCGGTTCCGG CCTTGGGGGT GCAGCGCATT TGCGGCACAG GCTTCGAACT GCTTCGGCAG	2132
GCCGGCGAGC AGATTTCCCA AGGCGCTGAT CACGTGCTGT GTGTCGCGGG CTGCAG	2188

FIG. 2g:

CTGCAGCCGA GCATCGATTG AGCACTTTAC CCAGCTGCGC TGGCTGACCA TTCAGAATGG	60
CCCGCGGCAC TATCCAATCT AAATCGATCT TCGGGCGCCG CGGGCATCAT GCCCGCGGCC	120
CTCGCCTCAT TTCAATCTCT AACTTGATAA AAACAGAGCT GTTCTCCGGT CTTGGTGGAT	180
CAAGGCCAGT CGCGGAGAGT CTCGAAGAGG AGAGTACAGT GAACGCCGAG TCCACATTGC	240
AACCGCAGGC ATCATCATGC TCTGCTCAGC CACGCTACCG CAGTGTGTCG ATTGGTCATC	300
CTCCGGTTGA GGTTACGCAA GACGCTGGAG GTATTGTCCG G ATG CGT TCT CTC GAG	356
Met Arg Ser Leu Glu	5
1	
GCG CTT CTT CCC TTC CCG GGT CGA ATT CTT GAG CGT CTC GAG CAT TGG	404
Ala Leu Leu Pro Phe Pro Gly Arg Ile Leu Glu Arg Leu Glu His Trp	
10 15 20	
GCT AAG ACC CGT CCA GAA CAA ACC TGC GTT GCT GCC AGG GCG GCA AAT	452
Ala Lys Thr Arg Pro Glu Gln Thr Cys Val Ala Ala Arg Ala Ala Asn	
25 30 35	
GGG GAA TGG CGT CGT ATC AGC TAC GCG GAA ATG TTC CAC AAC GTC CGC	500
Gly Glu Trp Arg Arg Ile Ser Tyr Ala Glu Met Phe His Asn Val Arg	
40 45 50	
GCC ATC GCA CAG AGC TTG CTT CCT TAC GGA CTA TCG GCA GAG CGT CCG	548
Ala Ile Ala Gln Ser Leu Leu Pro Tyr Gly Leu Ser Ala Glu Arg Pro	
55 60 65	
CTG CTT ATC GTC TCT GGA AAT GAC CTG GAA CAT CTT CAG CTG GCA TTT	596
Leu Leu Ile Val Ser Gly Asn Asp Leu Glu His Leu Gln Leu Ala Phe	
70 75 80 85	
GGG GCT ATG TAT GCG GGC ATT CCC TAT TGC CCG GTG TCT CCT GCT TAT	644
Gly Ala Met Tyr Ala Gly Ile Pro Tyr Cys Pro Val Ser Pro Ala Tyr	
90 95 100	
TCA CTG CTG TCG CAA GAT TTG GCG AAG CTG CGT CAC ATC GTA GGT CTT	692
Ser Leu Leu Ser Gln Asp Leu Ala Lys Leu Arg His Ile Val Gly Leu	
105 110 115	
CTG CAA CCG GGA CTG GTC TTT GCT GCC GAT GCA GCA CCT TTC CAG GGG	740
Leu Gln Pro Gly Leu Val Phe Ala Ala Asp Ala Ala Pro Phe Gln	
120 125 130 132	
GAGAGGCGGT TTGCGTATTG GGCGCATGCA TAAAACTGT TGTAATTCAT TAAGCATTCT	800
GCCGACATGG AAGCCATCAC AAACGGCATG ATGAACCTGA ATCGCCAGCG GCATCAGCAC	860
CTTGTCGCCT TGC GTATAAT ATTTGCCCAT GGACGCACAC CGTGGAACG GATGAAGGCA	920
CGAACCCAGT TGACATAAGC CTGTTCCGGT CGTAACTGT AATGCAAGTA GCGTATGCCG	980
TCACGCAACT GGTCCAGAAC CTTGACCGAA CGCAGCGGTG GTAACGGCCG AGTGGCGGTT	1040
TTCATGGCTT GTTATGACTG TTTTTTGTG CAGTCTATGC CTCGGGCATC CAAGCAGCAA	1100

GCGCGTTACG CCGTGGGTCG ATGTTTGATG TTATGGAGCA GCAACG ATG TTA CGC	1155
Met Leu Arg 1	
AGC AGC AAC GAT GTT ACG CAG CAG GGC AGT CGC CCT AAA ACA AAG TTA	1203
Ser Ser Asn Asp Val Thr Gln Gln Gly Ser Arg Pro Lys Thr Lys Leu	
5 10 15	
GGT GGC TCA AGT ATG GGC ATC ATT CGC ACA TGT AGG CTC GGC CCT GAC	1251
Gly Gly Ser Ser Met Gly Ile Ile Arg Thr Cys Arg Leu Gly Pro Asp	
20 25 30 35	
CAA GTC AAA TCC ATG CGG GCT GCT CTT GAT CTT TTC GGT CGT GAG TTC	1299
Gln Val Lys Ser Met Arg Ala Ala Leu Asp Leu Phe Gly Arg Glu Phe	
40 45 50	
GGA GAC GTA GCC ACC TAC TCC CAA CAT CAG CCG GAC TCC GAT TAC CTC	1347
Gly Asp Val Ala Thr Tyr Ser Gln His Gln Pro Asp Ser Asp Tyr Leu	
55 60 65	
GGG AAC TTG CTC CGT AGT AAG ACA TTC ATC GCG CTT GCT GCC TTC GAC	1395
Gly Asn Leu Leu Arg Ser Lys Thr Phe Ile Ala Leu Ala Ala Phe Asp	
70 75 80	
CAA GAA GCG GTT GTT GGC GCT CTC GCG GCT TAC GTT CTG CCC AGG TTT	1443
Gln Glu Ala Val Val Gly Ala Leu Ala Ala Tyr Val Leu Pro Arg Phe	
85 90 95	
GAG CAG CCG CGT AGT GAG ATC TAT ATC TAT GAT CTC GCA GTC TCC GGC	1491
Glu Gln Pro Arg Ser Glu Ile Tyr Ile Tyr Asp Leu Ala Val Ser Gly	
100 105 110 115	
GAG CAC CGG AGG CAG GGC ATT GCC ACC GCG CTC ATC AAT CTC CTC AAG	1539
Glu His Arg Arg Gln Gly Ile Ala Thr Ala Leu Ile Asn Leu Leu Lys	
120 125 130	
CAT GAG GCC AAC GCG CTT GGT GCT TAT GTG ATC TAC GTG CAA GCA GAT	1587
His Glu Ala Asn Ala Leu Gly Ala Tyr Val Ile Tyr Val Gln Ala Asp	
135 140 145	
TAC GGT GAC GAT CCC GCA GTG GCT CTC TAT ACA AAG TTG GGC ATA CGG	1635
Tyr Gly Asp Asp Pro Ala Val Ala Leu Tyr Thr Lys Leu Gly Ile Arg	
150 155 160	
GAA GAA GTG ATG CAC TTT GAT ATC GAC CCA AGT ACC GCC ACC TAA CAA	1683
Glu Glu Val Met His Phe Asp Ile Asp Pro Ser Thr Ala Thr	
165 170 175 177	
TTTCGTTCAAG CCGAGATCGG CTTCCCTT GTT TTG CAA TGG CGG TCG GCG AAA	1735
Val Leu Gln Trp Arg Ser Ala Lys	
563 565 570	
GTT GAT GCG CTG TAT CGT GGT GAA GAT CAA TCC ATG CTG CGT GAC GAG	1783
Val Asp Ala Leu Tyr Arg Gly Glu Asp Gln Ser Met Leu Arg Asp Glu	
575 580 585	

GCC ACA CTG TGA GTTGGTCAGG GGGGGCTTAC TCGGCGTTTT CCGACACTGC 1835
Ala Thr Leu
589

GTTGGTTGCG GCAGTGCGCA CCCCTGGAT TGATTGCGGG GGTGCCCTGT CGCTGGTGTC 1895

GCCTATCGAC TTAGGGGTAA AGGTCGCTCG CGAAGTTCTG ATGCGTGCGT CGCTTGAACC 1955

ACAAATGGTC GATAGCGTAC TCGCAGGCTC TATGGCTCAA GCAAGCTTTG ATGCTTACCT 2015

GCTCCCGCGG CACATTGGCT TGTACAGCGG TGTTCCCAAG TCGGTTCCGG CCTTGGGGGT 2075

GCAGCGCATT TGCGGCACAG GCTTCGAACT GCTTCGGCAG GCCGGCGAGC AGATTTCCTA 2135

AGGCGCTGAT CACGTGCTGT GTGTCGCGGG CTGCAG 2171

FIG. 2h:

CTGCAGCCGA GCATCGATTG AGCACTTTAC CCAGCTGCGC TGGCTGACCA TTCAGAATGG	60
CCCGCGGCAC TATCCAATCT AAATCGATCT TCGGGCGCCG CGGGCATCAT GCCCGCGGCG	120
CTCGCCTCAT TTCAATCTCT AACTTGATAA AAACAGAGCT GTTCTCCGGT CTTGGTGGAT	180
CAAGGCCAGT CGCGGAGAGT CTCGAAGAGG AGAGTACAGT GAACGCCGAG TCCACATTGC	240
AACCGCAGGC ATCATCATGC TCTGCTCAGC CACGCTACCG CAGTGTGTCG ATTGGTCATC	300
CTCCGGTTGA GGTACGCAA GACGCTGGAG GTATTGTCCG G ATG CGT TCT CTC GAG	356
Met Arg Ser Leu Glu	
1 5	
GCG CTT CTT CCC TTC CCG GGT CGA ATT CTT GAG CGT CTC GAG CAT TGG	404
Ala Leu Leu Pro Phe Pro Gly Arg Ile Leu Glu Arg Leu Glu His Trp	
10 15 20	
GCT AAG ACC CGT CCA GAA CAA ACC TGC GTT GCT GCC AGG GCG GCA AAT	452
Ala Lys Thr Arg Pro Glu Gln Thr Cys Val Ala Ala Arg Ala Ala Asn	
25 30 35	
GGG GAA TGG CGT CGT ATC AGC TAC GCG GAA ATG TTC CAC AAC GTC CGC	500
Gly Glu Trp Arg Arg Ile Ser Tyr Ala Glu Met Phe His Asn Val Arg	
40 45 50	
GCC ATC GCA CAG AGC TTG CTT CCT TAC GGA CTA TCG GCA GAG CGT CCG	548
Ala Ile Ala Gln Ser Leu Leu Pro Tyr Gly Leu Ser Ala Glu Arg Pro	
55 60 65	
CTG CTT ATC GTC TCT GGA AAT GAC CTG GAA CAT CTT CAG CTG GCA TTT	596
Leu Leu Ile Val Ser Gly Asn Asp Leu Glu His Leu Gln Leu Ala Phe	
70 75 80 85	
GGG GCT ATG TAT GCG GGC ATT CCC TAT TGC CCG GTG TCT CCT GCT TAT	644
Gly Ala Met Tyr Ala Gly Ile Pro Tyr Cys Pro Val Ser Pro Ala Tyr	
90 95 100	
TCA CTG CTG TCG CAA GAT TTG GCG AAG CTG CGT CAC ATC GTA GGT CTT	692
Ser Leu Leu Ser Gln Asp Leu Ala Lys Leu Arg His Ile Val Gly Leu	
105 110 115	
CTG CAA CCG GGA CTG GTC TTT GCT GCC GAT GCA GCA CCT TTC CAG CGC	740
Leu Gln Pro Gly Leu Val Phe Ala Ala Asp Ala Ala Pro Phe Gln Arg	
120 125 130 133	
GCT GTT TTG CAA TGG CGG TCG GCG AAA GTT GAT GCG CTG TAT CGT GGT	788
Ala Val Leu Gln Trp Arg Ser Ala Lys Val Asp Ala Leu Tyr Arg Gly	
562 565 570 575	
GAA GAT CAA TCC ATG CTG CGT GAC GAG GCC ACA CTG TGA GTTGGTCAGG	837
Glu Asp Gln Ser Met Leu Arg Asp Glu Ala Thr Leu	
580 585 589	
GGGGGCTTAC TCGGCGTTTT CCGACACTGC GTTGGTTGCG GCAGTGC GCA CCCCCTGGAT	897
TGATTGCGGG GGTGCCCTGT CGCTGGTGTC GCCTATCGAC TTAGGGGTAA AGGTCGCTCG	957

CGAAGTTCTG ATGCGTGCGT CGCTTGAACC ACAAATGGTC GATAGCGTAC TCGCAGGCTC 1017
TATGGCTCAA GCAAGCTTTG ATGCTTACCT GCTCCCGCGG CACATTGGCT TGTACAGCGG 1077
TGTTCCCAAG TCGGTTCCGG CCTTGGGGGT GCAGCGCATT TCGGGCACAG GCTTCGAACT 1137
GCTTCGGCAG GCCGGCGAGC AGATTTCCTA AGGCGCTGAT CACGTGCTGT GTGTCGCGGG 1197
CTGCAG 1203

FIG. 2i:

GAATTCCCCT GGCGACGAAA GGGCGGCAGG CCGCATGGCC ACGGCTGGGC GGTAAC TGAT 60

GCTTGCGTTA ATCGTTAACC GTTTGAAATT CCTTGCCAAA TTTCGGCGAG AGAATCATGC 120

GGGTACGCCT TTCCGTGCGC TTTGATCTGC GCTTCCGTGC CTTGAATCAG AAAAATAGTT 180

AATTGACAGA ACTATAGGTT CGCAGTAGCT TTTGCTCACC CACCAAATCC ACAGCACTGG 240

GGTGCACG ATG AAT AGC TAC GAT GGC CGT TGG TCT ACC GTT GAT GTG AAG 290
Met Asn Ser Tyr Asp Gly Arg Trp Ser Thr Val Asp Val Lys
1 5 10

GTT GAA GAA GGT ATC GCT TGG GTC ACG CTG AAC CGC CCG GAG AAG CGC 338
Val Glu Glu Gly Ile Ala Trp Val Thr Leu Asn Arg Pro Glu Lys Arg
15 20 25 30

AAC GCA ATG AGC CCA ACT CTC AAT CGA GAG ATG GTC GAG GTT CTG GAG 386
Asn Ala Met Ser Pro Thr Leu Asn Arg Glu Met Val Glu Val Leu Glu
35 40 45

GTG CTG GAG CAG GAC GCA GAT GCT CGC GTG CTT GTT CTG ACT GGT GCA 434
Val Leu Glu Gln Asp Ala Asp Ala Arg Val Leu Val Leu Thr Gly Ala
50 55 60

GGC GAA TCC TGG ACC GCG GGC ATG GAC CTG AAG GAG TAT TTC CGC GAG 482
Gly Glu Ser Trp Thr Ala Gly Met Asp Leu Lys Glu Tyr Phe Arg Glu
65 70 75

ACC GAT GCT GGC CCC GAA ATT CTG CAA GAG AAG ATT CGT CGGGGACAGC 531
Thr Asp Ala Gly Pro Glu Ile Leu Gln Glu Lys Ile Arg
80 85 90 91

AAGCGAACCG GAATTGCCAG CTGGGGCGCC CTCTGGTAAG GTTGGGAAGC CCTGCAAAGT 591

AAACTGGATG GCTTTCCTGC CGCCAAGGAT CTGATGGCGC AGGGGATCAA GATCTGATCA 651

AGAGACAGGA TGAGGATCGT TTCGC ATG ATT GAA CAA GAT GGA TTG CAC GCA 703
Met Ile Glu Gln Asp Gly Leu His Ala
1 5

GGT TCT CCG GCC GCT TGG GTG GAG AGG CTA TTC GGC TAT GAC TGG GCA 751
Gly Ser Pro Ala Ala Trp Val Glu Arg Leu Phe Gly Tyr Asp Trp Ala
10 15 20 25

CAA CAG ACA ATC GGC TGC TCT GAT GCC GCC GTG TTC CGG CTG TCA GCG 799
Gln Gln Thr Ile Gly Cys Ser Asp Ala Ala Val Phe Arg Leu Ser Ala
30 35 40

CAG GGG CGC CCG GTT CTT TTT GTC AAG ACC GAC CTG TCC GGT GCC CTG 847
Gln Gly Arg Pro Val Leu Phe Val Lys Thr Asp Leu Ser Gly Ala Leu
45 50 55

AAT GAA CTG CAG GAC GAG GCA GCG CGG CTA TCG TGG CTG GCC ACG ACG 895
Asn Glu Leu Gln Asp Glu Ala Ala Arg Leu Ser Trp Leu Ala Thr Thr
60 65 70

GGC GTT CCT TGC GCA GCT GTG CTC GAC GTT GTC ACT GAA GCG GGA AGG Gly Val Pro Cys Ala Ala Val Leu Asp Val Val Thr Glu Ala Gly Arg 75 80 85	943
GAC TGG CTG CTA TTG GGC GAA GTG CCG GGG CAG GAT CTC CTG TCA TCT Asp Trp Leu Leu Leu Gly Glu Val Pro Gly Gln Asp Leu Leu Ser Ser 90 95 100 105	991
CAC CTT GCT CCT GCC GAG AAA GTA TCC ATC ATG GCT GAT GCA ATG CGG His Leu Ala Pro Ala Glu Lys Val Ser Ile Met Ala Asp Ala Met Arg 110 115 120	1039
CGG CTG CAT ACG CTT GAT CCG GCT ACC TGC CCA TTC GAC CAC CAA GCG Arg Leu His Thr Leu Asp Pro Ala Thr Cys Pro Phe Asp His Gln Ala 125 130 135	1087
AAA CAT CGC ATC GAG CGA GCA CGT ACT CGG ATG GAA GCC GGT CTT GTC Lys His Arg Ile Glu Arg Ala Arg Thr Arg Met Glu Ala Gly Leu Val 140 145 150	1135
GAT CAG GAT GAT CTG GAC GAA GAG CAT CAG GGG CTC GCG CCA GCC GAA Asp Gln Asp Asp Leu Asp Glu Glu His Gln Gly Leu Ala Pro Ala Glu 155 160 165	1183
CTG TTC GCC AGG CTC AAG GCG CGC ATG CCC GAC GGC GAG GAT CTC GTC Leu Phe Ala Arg Leu Lys Ala Arg Met Pro Asp Gly Glu Asp Leu Val 170 175 180 185	1231
GTG ACC CAT GGC GAT GCC TGC TTG CCG AAT ATC ATG GTG GAA AAT GGC Val Thr His Gly Asp Ala Cys Leu Pro Asn Ile Met Val Glu Asn Gly 190 195 200	1279
CGC TTT TCT GGA TTC ATC GAC TGT GGC CGG CTG GGT GTG GCG GAC CGC Arg Phe Ser Gly Phe Ile Asp Cys Gly Arg Leu Gly Val Ala Asp Arg 205 210 215	1327
TAT CAG GAC ATA GCG TTG GCT ACC CGT GAT ATT GCT GAA GAG CTT GGC Tyr Gln Asp Ile Ala Leu Ala Thr Arg Asp Ile Ala Glu Glu Leu Gly 220 225 230	1375
GGC GAA TGG GCT GAC CGC TTC CTC GTG CTT TAC GGT ATC GCC GCT CCC Gly Glu Trp Ala Asp Arg Phe Leu Val Leu Tyr Gly Ile Ala Ala Pro 235 240 245	1423
GAT TCG CAG CGC ATC GCC TTC TAT CGC CTT CTT GAC GAG TTC TTC TGA Asp Ser Gln Arg Ile Ala Phe Tyr Arg Leu Leu Asp Glu Phe Phe 250 255 260 264	1471
CGGGGACTCT GGGGTTTCGAA ATGACCGACC AAGCGACGCC CC GAG CAG GGC ATG Glu Gln Gly Met 255	1525
AAG CAG TTC CTT GAC GAG AAA AGC ATC AAG CCG GGC TTG CAG ACC TAC Lys Gln Phe Leu Asp Glu Lys Ser Ile Lys Pro Gly Leu Gln Thr Tyr 260 265 270	1573

AAG CGC TGA TAAATGCGCC GGGGCCCTCG CTGCGCCCCC GGCCTTCCAA TAATGACAAT 1632
Lys Arg
275 276

AATGAGGAGT GCCCAATGTT TCACGTGCCC CTGCTTATTG GTGGTAAGCC TTGTTGAGCA 1692

TCTGATGAGC GCACCTTCGA GCGTCGTAGC CCGCTGACCG GAGAAGTGGT ATCGCGCGTC 1752

GCTGCTGCCA GTTTGGAAGA TGCGGACGCC GCAGTGGCCG CTGCACAGGC TCGTTTCCT 1812

GAATGGGCGG CGCTTGCTCC GAGCGAACGC CGTGCCCGAC TGCTGCGAGC GCGGATCTT 1872

CTAGAGGACC GTTCTTCCGA GTTCACCGCC GCAGCGAGTG AACTGGCGC AGCGGGAAAC 1932

TGGTATGGGT TTAACGTTTA CCTGGCGGCG GGCATGTTGC GGGGAATTC 1981

FIG. 2j:

GAATTCCTT	GGCGACGAAA	GGGCGGCAGG	CCGCATGGCC	ACGGCTGGGC	GGTAACTGAT	60
GCTTGCGTTA	ATCGTTAACC	GTTTGAAATT	CCTTGCCAAA	TTTCGGCGAG	AGAATCATGC	120
GGGTACGCCT	TTCCGTGCGC	TTTGATCTGC	GCTTCCGTGC	CTTGAATCAG	AAAAATAGTT	180
AATTGACAGA	ACTATAGGTT	CGCAGTAGCT	TTTGCTCACC	CACCAAATCC	ACAGCACTGG	240
GGTGACAG	ATG AAT AGC TAC GAT GGC CGT TGG TCT ACC GTT GAT GTG AAG	290				
	Met Asn Ser Tyr Asp Gly Arg Trp Ser Thr Val Asp Val Lys					
	1 5 10					
GTT GAA GAA GGT ATC GCT TGG GTC ACG CTG AAC CGC CCG GAG AAG CGC	338					
Val Glu Glu Gly Ile Ala Trp Val Thr Leu Asn Arg Pro Glu Lys Arg						
15 20 25 30						
AAC GCA ATG AGC CCA ACT CTC AAT CGA GAG ATG GTC GAG GTT CTG GAG	386					
Asn Ala Met Ser Pro Thr Leu Asn Arg Glu Met Val Glu Val Leu Glu						
35 40 45						
GTG CTG GAG CAG GAC GCA GAT GCT CGC GTG CTT GTT CTG ACT GGT GCA	434					
Val Leu Glu Gln Asp Ala Asp Ala Arg Val Leu Val Leu Thr Gly Ala						
50 55 60						
GGC GAA TCC TGG ACC GCG GGC ATG GAC CTG AAG GAG TAT TTC CGC GAG	482					
Gly Glu Ser Trp Thr Ala Gly Met Asp Leu Lys Glu Tyr Phe Arg Glu						
65 70 75						
ACC GAT GCT GGC CCC GAA ATT CTG CAA GAG AAG ATT CGT CGGGGGAGAG	531					
Thr Asp Ala Gly Pro Glu Ile Leu Gln Glu Lys Ile Arg						
80 85 90 91						
GCGGTTTGCG TATTGGGCGC ATGCATAAAA ACTGTTGTAA TTCATTAAGC ATTCTGCCGA	591					
CATGGAAGCC ATCACAAACG GCATGATGAA CCTGAATCGC CAGCGGCATC AGCACCTTGT	651					
CGCCTTGCGT ATAATATTTG CCCATGGACG CACACCGTGG AAACGGATGA AGGCACGAAC	711					
CCAGTTGACA TAAGCCTGTT CGGTTTCGTAA ACTGTAATGC AAGTAGCGTA TCGCTCAGC	771					
CAACTGGTCC AGAACCTTGA CCGAACGCAG CGGTGGTAAC GGCGCAGTGG CGGTTTTTCAT	831					
GGCTTGTTAT GACTGTTTTT TTGTACAGTC TATGCCTCGG GCATCCAAGC AGCAAGCGCG	891					
TTACGCCGTG GGTGCATGTT TGATGTTATG GAGCAGCAAC G ATG TTA CGC AGC AGC	947					
	Met Leu Arg Ser Ser					
	1 5					
AAC GAT GTT ACG CAG CAG GGC AGT CGC CCT AAA ACA AAG TTA GGT GGC	995					
Asn Asp Val Thr Gln Gln Gly Ser Arg Pro Lys Thr Lys Leu Gly Gly						
10 15 20						
TCA AGT ATG GGC ATC ATT CGC ACA TGT AGG CTC GGC CCT GAC CAA GTC	1043					
Ser Ser Met Gly Ile Ile Arg Thr Cys Arg Leu Gly Pro Asp Gln Val						
25 30 35						

AAA TCC ATG CGG GCT GCT CTT GAT CTT TTC GGT CGT GAG TTC GGA GAC	1091
Lys Ser Met Arg Ala Ala Leu Asp Leu Phe Gly Arg Glu Phe Gly Asp	
40 45 50	
GTA GCC ACC TAC TCC CAA CAT CAG CCG GAC TCC GAT TAC CTC GGG AAC	1139
Val Ala Thr Tyr Ser Gln His Gln Pro Asp Ser Asp Tyr Leu Gly Asn	
55 60 65	
TTG CTC CGT AGT AAG ACA TTC ATC GCG CTT GCT GCC TTC GAC CAA GAA	1187
Leu Leu Arg Ser Lys Thr Phe Ile Ala Leu Ala Ala Phe Asp Gln Glu	
70 75 80 85	
GCG GTT GTT GGC GCT CTC GCG GCT TAC GTT CTG CCC AGG TTT GAG CAG	1235
Ala Val Val Gly Ala Leu Ala Ala Tyr Val Leu Pro Arg Phe Glu Gln	
90 95 100	
CCG CGT AGT GAG ATC TAT ATC TAT GAT CTC GCA GTC TCC GGC GAG CAC	1283
Pro Arg Ser Glu Ile Tyr Ile Tyr Asp Leu Ala Val Ser Gly Glu His	
105 110 115	
CGG AGG CAG GGC ATT GCC ACC GCG CTC ATC AAT CTC CTC AAG CAT GAG	1331
Arg Arg Gln Gly Ile Ala Thr Ala Leu Ile Asn Leu Leu Lys His Glu	
120 125 130	
GCC AAC GCG CTT GGT GCT TAT GTG ATC TAC GTG CAA GCA GAT TAC GGT	1379
Ala Asn Ala Leu Gly Ala Tyr Val Ile Tyr Val Gln Ala Asp Tyr Gly	
135 140 145	
GAC GAT CCC GCA GTG GCT CTC TAT ACA AAG TTG GGC ATA CGG GAA GAA	1427
Asp Asp Pro Ala Val Ala Leu Tyr Thr Lys Leu Gly Ile Arg Glu Glu	
150 155 160 165	
GTG ATG CAC TTT GAT ATC GAC CCA AGT ACC GCC ACC TAA CAATTCGTTT	1476
Val Met His Phe Asp Ile Asp Pro Ser Thr Ala Thr	
170 175 177	
AAGCCGAGAT CGGCTTCCCC GAG CAG GGC ATG AAG CAG TTC CTT GAC GAG	1526
Glu Gln Gly Met Lys Gln Phe Leu Asp Glu	
255 260	
AAA AGC ATC AAG CCG GGC TTG CAG ACC TAC AAG CGC TGA TAAATGCGCC	1575
Lys Ser Ile Lys Pro Gly Leu Gln Thr Tyr Lys Arg	
265 270 275 276	
GGGGCCCTCG CTGCGCCCCC GGCCTTCCAA TAATGACAAT AATGAGGAGT GCCCAATGTT	1635
TCACGTGCCC CTGCTTATTG GTGGTAAGCC TTGTTTACGCA TCTGATGAGC GCACCTTCGA	1695
GCGTCGTAGC CCGCTGACCG GAGAAGTGGT ATCGCGCGTC GCTGCTGCCA GTTTGGAAGA	1755
TGCGGACGCC GCAGTGGCCG CTGCACAGGC TGCGTTTCCT GAATGGGCGG CGCTTGCTCC	1815
GAGCGAACGC CGTGCCCGAC TGCTGCGAGC GGCGGATCTT CTAGAGGACC GTTCTTCCGA	1875

G TTCACCGCC GCAGCGAGTG AACTGGCGC AGCGGGAAAC TGGTATGGGT TTAACGTTTA 1935

CCTGGCGGCG GGCATGTTGC GGGGAATTC 1964

FIG. 2k:

GAATTCCTCCCT GGCACGAAA GGGCGGCAGG CCGCATGGCC ACGGCTGGGC GGTAACCTGAT	60
GCTTGCGTTA ATCGTTAACC GTTTGAAATT CCTTGCCAAA TTTCGGCGAG AGAATCATGC	120
GGGTACGCCT TTCCGTGCGC TTTGATCTGC GCTTCCGTGC CTTGAATCAG AAAAAATAGTT	180
AATTGACAGA ACTATAGGTT CGCAGTAGCT TTTGCTCACC CACCAAATCC ACAGCACTGG	240
GGTGCACG ATG AAT AGC TAC GAT GGC CGT TGG TCT ACC GTT GAT GTG AAG	290
Met Asn Ser Tyr Asp Gly Arg Trp Ser Thr Val Asp Val Lys	
1 5 10	
GTT GAA GAA GGT ATC GCT TGG GTC ACG CTG AAC CGC CCG GAG AAG CGC	338
Val Glu Glu Gly Ile Ala Trp Val Thr Leu Asn Arg Pro Glu Lys Arg	
15 20 25 30	
AAC GCA ATG AGC CCA ACT CTC AAT CGA GAG ATG GTC GAG GTT CTG GAG	386
Asn Ala Met Ser Pro Thr Leu Asn Arg Glu Met Val Glu Val Leu Glu	
35 40 45	
GTG CTG GAG CAG GAC GCA GAT GCT CGC GTG CTT GTT CTG ACT GGT GCA	434
Val Leu Glu Gln Asp Ala Asp Ala Arg Val Leu Val Leu Thr Gly Ala	
50 55 60	
GGC GAA TCC TGG ACC GCG GGC ATG GAC CTG AAG GAG TAT TTC CGC GAG	482
Gly Glu Ser Trp Thr Ala Gly Met Asp Leu Lys Glu Tyr Phe Arg Glu	
65 70 75	
ACC GAT GCT GGC CCC GAA ATT CTG CAA GAG AAG ATT CGT CGC GAG CAG	530
Thr Asp Ala Gly Pro Glu Ile Leu Gln Glu Lys Ile Arg Arg Glu Gln	
80 85 90 92 255	
GGC ATG AAG CAG TTC CTT GAC GAG AAA AGC ATC AAG CCG GGC TTG CAG	578
Gly Met Lys Gln Phe Leu Asp Glu Lys Ser Ile Lys Pro Gly Leu Gln	
260 265 270	
ACC TAC AAG CGC TGA TAAATGCGCC GGGGCCCTCG CTGCGCCCCC GGCCTTCCAA	633
Thr Tyr Lys Arg	
275 276	
TAATGACAAT AATGAGGAGT GCCCAATGTT TCACGTGCCC CTGCTTATTG GTGGTAAGCC	693
TTGTTTACGCA TCTGATGAGC GCACCTTCGA GCGTCGTAGC CCGCTGACCG GAGAAGTGGT	753
ATCGCGCGTC GCTGCTGCCA GTTTGGAAGA TGCGGACGCC GCAGTGGCCG CTGCACAGGC	813
TGCGTTTCCT GAATGGGCGG CGCTTGCTCC GAGCGAACGC CGTGCCCGAC TGCTGCGAGC	873
GGCGGATCTT CTAGAGGACC GTTCTTCCGA GTTCACCGCC GCAGCGAGTG AAACCTGGCGC	933
AGCGGGAAAC TGGTATGGGT TTAACGTTTA CCTGGCGGCG GGCATGTTGC GGGGAATTC	992

FIG. 21:

GAATTCCAAT AATGACAATA ATGAGGAGTG CCCA ATG TTT CAC GTG CCC CTG CTT	55
Met Phe His Val Pro Leu Leu	
1 5	
ATT GGT GGT AAG CCT TGT TCA GCA TCT GAT GAG CGC ACC TTC GAG CGT	103
Ile Gly Gly Lys Pro Cys Ser Ala Ser Asp Glu Arg Thr Phe Glu Arg	
10 15 20	
CGT AGC CCG CTG ACC GGA GAA GTG GTA TCG CGC GTC GCT GCT GCC AGT	151
Arg Ser Pro Leu Thr Gly Glu Val Val Ser Arg Val Ala Ala Ala Ser	
25 30 35	
TTG GAA GAT GCG GAC GCC GCA GTG GCC GCT GCA CAG GCT GCG TTT CCT	199
Leu Glu Asp Ala Asp Ala Ala Val Ala Ala Ala Gln Ala Ala Phe Pro	
40 45 50 55	
GAA TGG GCG GCG CTT GCT CCG AGC GAA CGC CGT GCC CGA CTG CTG CGA	247
Glu Trp Ala Ala Leu Ala Pro Ser Glu Arg Arg Ala Arg Leu Leu Arg	
60 65 70	
GCG GCG GAT CTT CTA GAG GAC CGT TCT TCC GAG TTC ACC GCC GCA GCG	295
Ala Ala Asp Leu Leu Glu Asp Arg Ser Ser Glu Phe Thr Ala Ala Ala	
75 80 85	
AGT GAA ACT GGC GCA GCG GGA AAC TGG TAT GGG TTT AAC GTT TAC CTG	343
Ser Glu Thr Gly Ala Ala Gly Asn Trp Tyr Gly Phe Asn Val Tyr Leu	
90 95 100	
GCG GCG GGC ATG TTG CGG GAA GCC GCG GCC ATG ACC ACA CAG ATT CAG	391
Ala Ala Gly Met Leu Arg Glu Ala Ala Ala Met Thr Thr Gln Ile Gln	
105 110 115	
GGC GAT GTC ATT CCG TCC AAT GTG CCC GGT AGC TTT GCC ATG GCG GTT	439
Gly Asp Val Ile Pro Ser Asn Val Pro Gly Ser Phe Ala Met Ala Val	
120 125 130 135	
CGA CAG CCA TGT GGC GTG GTG CTC GGT ATT GCG CCT TGG AAT GCT CCG	487
Arg Gln Pro Cys Gly Val Val Leu Gly Ile Ala Pro Trp Asn Ala Pro	
140 145 150	
GTA ATC CTT GGC GTA CGG GCT GTT GCG ATG CCG TTG GCA TGC GGC AAT	535
Val Ile Leu Gly Val Arg Ala Val Ala Met Pro Leu Ala Cys Gly Asn	
155 160 165	
ACC GTG GTG TTG AAA AGC TCT GAG CTG AGT CCC TTT ACC CAT CGC CTG	583
Thr Val Val Leu Lys Ser Ser Glu Leu Ser Pro Phe Thr His Arg Leu	
170 175 180	
ATT GGT CAG GTG TTG CAT GAT GCT GGT CTG GGG GAT GGC GTG GTG AAT	631
Ile Gly Gln Val Leu His Asp Ala Gly Leu Gly Asp Gly Val Val Asn	
185 190 195	
GTC ATC AGC AAT GCC CCG CAA GAC GCT CCT GCG GTG GTG GAG CGA CTG	679
Val Ile Ser Asn Ala Pro Gln Asp Ala Pro Ala Val Val Glu Arg Leu	
200 205 210 215	

ATT GCA AAT CCT GCG GTA CGT CGA GTG AAC TTC ACC GGT TCG ACC CAC Ile Ala Asn Pro Ala Val Arg Arg Val Asn Phe Thr Gly Ser Thr His 220 225 230	727
GTT GGA CGG ATC ATT GGT GAG CTG TCT GCG CGT CAT CTG AAG CCT GCT Val Gly Arg Ile Ile Gly Glu Leu Ser Ala Arg His Leu Lys Pro Ala 235 240 245	775
GTG CTG GAA TTA GGT GGT AAG GCT CCG TTC TTG GTC TTG GAC GAT GCC Val Leu Glu Leu Gly Gly Lys Ala Pro Phe Leu Val Leu Asp Asp Ala 250 255 260	823
GAC CTC GAT GCG GCG GTC GAA GCG GCG GCC TTT GGT GCC TAC TTC AAT Asp Leu Asp Ala Ala Val Glu Ala Ala Ala Phe Gly Ala Tyr Phe Asn 265 270 275	871
CAG GGT CAA ATC TGC ATG TCC ACT GAG CGT CTG ATT GTG ACA GCA GTC Gln Gly Gln Ile Cys Met Ser Thr Glu Arg Leu Ile Val Thr Ala Val 280 285 290 295	919
GCA GAC GCC TTT GTT GAA AAG CTG GCG AGG AAG GTC GCC ACA CTG CGT Ala Asp Ala Phe Val Glu Lys Leu Ala Arg Lys Val Ala Thr Leu Arg 300 305 310	967
GCT GGC GAT CCT AAT GAT CCG CAA TCG GTC TTG GGT TCG TTG ATT GAT Ala Gly Asp Pro Asn Asp Pro Gln Ser Val Leu Gly Ser Leu Ile Asp 315 320 325	1015
GCC AAT GCA GGT CAA CGC ATC CAG GTT CTG GTC GAT GAT GCG CTC GGG Ala Asn Ala Gly Gln Arg Ile Gln Val Leu Val Asp Asp Ala Leu 330 335 340 342	1063
GACAGCAAGC GAACCGGAAT TGCCAGCTGG GGCGCCCTCT GGTAAGGTTG GGAAGCCCTG	1123
CAAAGTAAAC TGGATGGCTT TCTTGCCGCC AAGGATCTGA TGGCGCAGGG GATCAAGATC	1183
TGATCAAGAG ACAGGATGAG GATCGTTTCG C ATG ATT GAA CAA GAT GGA TTG Met Ile Glu Gln Asp Gly Leu 1 5	1235
CAC GCA GGT TCT CCG GCC GCT TGG GTG GAG AGG CTA TTC GGC TAT GAC His Ala Gly Ser Pro Ala Ala Trp Val Glu Arg Leu Phe Gly Tyr Asp 10 15 20	1283
TGG GCA CAA CAG ACA ATC GGC TGC TCT GAT GCC GCC GTG TTC CGG CTG Trp Ala Gln Gln Thr Ile Gly Cys Ser Asp Ala Ala Val Phe Arg Leu 25 30 35	1331
TCA GCG CAG GGG CGC CCG GTT CTT TTT GTC AAG ACC GAC CTG TCC GGT Ser Ala Gln Gly Arg Pro Val Leu Phe Val Lys Thr Asp Leu Ser Gly 40 45 50 55	1379
GCC CTG AAT GAA CTG CAG GAC GAG GCA GCG CGG CTA TCG TGG CTG GCC Ala Leu Asn Glu Leu Gln Asp Glu Ala Ala Arg Leu Ser Trp Leu Ala 60 65 70	1427

ACG	ACG	GGC	GTT	CCT	TGC	GCA	GCT	GTG	CTC	GAC	GTT	GTC	ACT	GAA	GCG	1475
Thr	Thr	Gly	Val	Pro	Cys	Ala	Ala	Val	Leu	Asp	Val	Val	Thr	Glu	Ala	
		75						80					85			
GGA	AGG	GAC	TGG	CTG	CTA	TTG	GGC	GAA	GTG	CCG	GGG	CAG	GAT	CTC	CTG	1523
Gly	Arg	Asp	Trp	Leu	Leu	Leu	Gly	Glu	Val	Pro	Gly	Gln	Asp	Leu	Leu	
	90						95				100					
TCA	TCT	CAC	CTT	GCT	CCT	GCC	GAG	AAA	GTA	TCC	ATC	ATG	GCT	GAT	GCA	1571
Ser	Ser	His	Leu	Ala	Pro	Ala	Glu	Lys	Val	Ser	Ile	Met	Ala	Asp	Ala	
	105					110					115					
ATG	CGG	CGG	CTG	CAT	ACG	CTT	GAT	CCG	GCT	ACC	TGC	CCA	TTC	GAC	CAC	1619
Met	Arg	Arg	Leu	His	Thr	Leu	Asp	Pro	Ala	Thr	Cys	Pro	Phe	Asp	His	
120					125					130					135	
CAA	GCG	AAA	CAT	CGC	ATC	GAG	CGA	GCA	CGT	ACT	CGG	ATG	GAA	GCC	GGT	1667
Gln	Ala	Lys	His	Arg	Ile	Glu	Arg	Ala	Arg	Thr	Arg	Met	Glu	Ala	Gly	
			140					145					150			
CTT	GTC	GAT	CAG	GAT	GAT	CTG	GAC	GAA	GAG	CAT	CAG	GGG	CTC	GCG	CCA	1715
Leu	Val	Asp	Gln	Asp	Asp	Leu	Asp	Glu	Glu	His	Gln	Gly	Leu	Ala	Pro	
		155					160					165				
GCC	GAA	CTG	TTC	GCC	AGG	CTC	AAG	GCG	CGC	ATG	CCC	GAC	GGC	GAG	GAT	1763
Ala	Glu	Leu	Phe	Ala	Arg	Leu	Lys	Ala	Arg	Met	Pro	Asp	Gly	Glu	Asp	
	170						175					180				
CTC	GTC	GTG	ACC	CAT	GGC	GAT	GCC	TGC	TTG	CCG	AAT	ATC	ATG	GTG	GAA	1811
Leu	Val	Val	Thr	His	Gly	Asp	Ala	Cys	Leu	Pro	Asn	Ile	Met	Val	Glu	
	185					190				195						
AAT	GGC	CGC	TTT	TCT	GGA	TTC	ATC	GAC	TGT	GGC	CGG	CTG	GGT	GTG	GCG	1859
Asn	Gly	Arg	Phe	Ser	Gly	Phe	Ile	Asp	Cys	Gly	Arg	Leu	Gly	Val	Ala	
200					205					210					215	
GAC	CGC	TAT	CAG	GAC	ATA	GCG	TTG	GCT	ACC	CGT	GAT	ATT	GCT	GAA	GAG	1907
Asp	Arg	Tyr	Gln	Asp	Ile	Ala	Leu	Ala	Thr	Arg	Asp	Ile	Ala	Glu	Glu	
			220					225					230			
CTT	GGC	GGC	GAA	TGG	GCT	GAC	CGC	TTC	CTC	GTG	CTT	TAC	GGT	ATC	GCC	1955
Leu	Gly	Gly	Glu	Trp	Ala	Asp	Arg	Phe	Leu	Val	Leu	Tyr	Gly	Ile	Ala	
		235					240					245				
GCT	CCC	GAT	TCG	CAG	CGC	ATC	GCC	TTC	TAT	CGC	CTT	CTT	GAC	GAG	TTC	2003
Ala	Pro	Asp	Ser	Gln	Arg	Ile	Ala	Phe	Tyr	Arg	Leu	Leu	Asp	Glu	Phe	
	250						255					260				
TTC	TGA	GCGGGACTCT	GGGGTTCGAA	ATGACCGACC	AAGCGACGCC	CG	GCC	CAG								2057
Phe							Ala	Gln								
264							421									
CGC	GTC	GAT	TCG	GCC	ATT	TGC	CAT	ATC	AAT	GGA	CCG	ACT	GTG	CAT	GAC	2105
Arg	Val	Asp	Ser	Gly	Ile	Cys	His	Ile	Asn	Gly	Pro	Thr	Val	His	Asp	
	425						430					435				

GAG GCT CAG ATG CCA TTC GGT GGG GTG AAG TCC AGC GGC TAC GGC AGC 2153
Glu Ala Gln Met Pro Phe Gly Gly Val Lys Ser Ser Gly Tyr Gly Ser
440 445 450

TTC GGC AGT CGA GCA TCG ATT GAG CAC TTT ACC CAG CTG CGC TGG CTG 2201
Phe Gly Ser Arg Ala Ser Ile Glu His Phe Thr Gln Leu Arg Trp Leu
455 460 465 470

ACC ATT CAG AAT GGC CCG CGG CAC TAT CCA ATC TAA ATCGATCTTC 2247
Thr Ile Gln Asn Gly Pro Arg His Tyr Pro Ile
475 480 481

GGGCGCCGCG GGCATCATGC CCGCGGCGCT CGCCTCATTT CAATCTCTAA CTTGATAAAA 2307

ACAGAGCTGT TCTCCGGTCT TGGTGGATCA AGGCCAGTCG CGGAGAGTCT CGAAGAGGAG 2367

AGTACAGTGA ACGCCGAGTC CACATTGCAA CCGCAGGCAT CATCATGCTC TGCTCAGCCA 2427

CGCTACCGCA GTGTGTCGAT TGGTCATCCT CCGGTTGAGG TTACGCAAGA CGCTGGAGGT 2487

ATTGTCCGGA TGC GTTCTCT CGAGGCGCTT CTTCCCTTCC CGGGTGAAT TC 2539

FIG. 2m:

ATT GCA AAT CCT GCG GTA CGT CGA GTG AAC TTC ACC GGT TCG ACC CAC Ile Ala Asn Pro Ala Val Arg Arg Val Asn Phe Thr Gly Ser Thr His 220 225 230	727
GTT GGA CGG ATC ATT GGT GAG CTG TCT GCG CGT CAT CTG AAG CCT GCT Val Gly Arg Ile Ile Gly Glu Leu Ser Ala Arg His Leu Lys Pro Ala 235 240 245	775
GTG CTG GAA TTA GGT GGT AAG GCT CCG TTC TTG GTC TTG GAC GAT GCC Val Leu Glu Leu Gly Gly Lys Ala Pro Phe Leu Val Leu Asp Asp Ala 250 255 260	823
GAC CTC GAT GCG GCG GTC GAA GCG GCG GCC TTT GGT GCC TAC TTC AAT Asp Leu Asp Ala Ala Val Glu Ala Ala Ala Phe Gly Ala Tyr Phe Asn 265 270 275	871
CAG GGT CAA ATC TGC ATG TCC ACT GAG CGT CTG ATT GTG ACA GCA GTC Gln Gly Gln Ile Cys Met Ser Thr Glu Arg Leu Ile Val Thr Ala Val 280 285 290 295	919
GCA GAC GCC TTT GTT GAA AAG CTG GCG AGG AAG GTC GCC ACA CTG CGT Ala Asp Ala Phe Val Glu Lys Leu Ala Arg Lys Val Ala Thr Leu Arg 300 305 310	967
GCT GGC GAT CCT AAT GAT CCG CAA TCG GTC TTG GGT TCG TTG ATT GAT Ala Gly Asp Pro Asn Asp Pro Gln Ser Val Leu Gly Ser Leu Ile Asp 315 320 325	1015
GCC AAT GCA GGT CAA CGC ATC CAG GTGGGGAGAG GCGGTTTGCG TATTGGGCGC Ala Asn Ala Gly Gln Arg Ile Gln 330 335	1069
ATGCATAAAA ACTGTTGTAA TTCATTAAGC ATTCTGCCGA CATGGAAGCC ATCACAAACG	1129
GCATGATGAA CCTGAATCGC CAGCGGCATC AGCACCTTGT CGCCTTGCGT ATAATATTTG	1189
CCCATGGACG CACACCGTGG AAACGGATGA AGGCACGAAC CCAGTTGACA TAAGCCTGTT	1249
CGGTTCGTAA ACTGTAATGC AAGTAGCGTA TCGGCTCAGC CAACTGGTCC AGAACCTTGA	1309
CCGAACGCAG CGGTGGTAAC GGCGCAGTGG CGGTTTTTCAT GGCTTGTTAT GACTGTTTTT	1369
TTGTACAGTC TATGCCTCGG GCATCCAAGC AGCAAGCGCG TTACGCCGTG GGTGATGTT	1429
TGATGTTATG GAGCAGCAAC G ATG TTA CGC AGC AGC AAC GAT GTT ACG CAG Met Leu Arg Ser Ser Asn Asp Val Thr Gln 1 5 10	1480
CAG GGC AGT CGC CCT AAA ACA AAG TTA GGT GGC TCA AGT ATG GGC ATC Gln Gly Ser Arg Pro Lys Thr Lys Leu Gly Gly Ser Ser Met Gly Ile 15 20 25	1528
ATT CGC ACA TGT AGG CTC GGC CCT GAC CAA GTC AAA TCC ATG CGG GCT Ile Arg Thr Cys Arg Leu Gly Pro Asp Gln Val Lys Ser Met Arg Ala 30 35 40	1576

GCT CTT GAT CTT TTC GGT CGT GAG TTC GGA GAC GTA GCC ACC TAC TCC Ala Leu Asp Leu Phe Gly Arg Glu Phe Gly Asp Val Ala Thr Tyr Ser 45 50 55	1624
CAA CAT CAG CCG GAC TCC GAT TAC CTC GGG AAC TTG CTC CGT AGT AAG Gln His Gln Pro Asp Ser Asp Tyr Leu Gly Asn Leu Leu Arg Ser Lys 60 65 70	1672
ACA TTC ATC GCG CTT GCT GCC TTC GAC CAA GAA GCG GTT GTT GGC GCT Thr Phe Ile Ala Leu Ala Ala Phe Asp Gln Glu Ala Val Val Gly Ala 75 80 85 90	1720
CTC GCG GCT TAC GTT CTG CCC AGG TTT GAG CAG CCG CGT AGT GAG ATC Leu Ala Ala Tyr Val Leu Pro Arg Phe Glu Gln Pro Arg Ser Glu Ile 95 100 105	1768
TAT ATC TAT GAT CTC GCA GTC TCC GGC GAG CAC CGG AGG CAG GGC ATT Tyr Ile Tyr Asp Leu Ala Val Ser Gly Glu His Arg Arg Gln Gly Ile 110 115 120	1816
GCC ACC GCG CTC ATC AAT CTC CTC AAG CAT GAG GCC AAC GCG CTT GGT Ala Thr Ala Leu Ile Asn Leu Leu Lys His Glu Ala Asn Ala Leu Gly 125 130 135	1864
GCT TAT GTG ATC TAC GTG CAA GCA GAT TAC GGT GAC GAT CCC GCA GTG Ala Tyr Val Ile Tyr Val Gln Ala Asp Tyr Gly Asp Asp Pro Ala Val 140 145 150	1912
GCT CTC TAT ACA AAG TTG GGC ATA CGG GAA GAA GTG ATG CAC TTT GAT Ala Leu Tyr Thr Lys Leu Gly Ile Arg Glu Glu Val Met His Phe Asp 155 160 165 170	1960
ATC GAC CCA AGT ACC GCC ACC TAA CAATTCGTTC AAGCCGAGAT CGGCTTCCCA Ile Asp Pro Ser Thr Ala Thr 175 177	2014
A TTG GCC CAG CGC GTC GAT TCG GGC ATT TGC CAT ATC AAT GGA CCG ACT Leu Ala Gln Arg Val Asp Ser Gly Ile Cys His Ile Asn Gly Pro Thr 420 425 430 435	2063
GTG CAT GAC GAG GCT CAG ATG CCA TTC GGT GGG GTG AAG TCC AGC GGC Val His Asp Glu Ala Gln Met Pro Phe Gly Gly Val Lys Ser Ser Gly 440 445 450	2111
TAC GGC AGC TTC GGC AGT CGA GCA TCG ATT GAG CAC TTT ACC CAG CTG Tyr Gly Ser Phe Gly Ser Arg Ala Ser Ile Glu His Phe Thr Gln Leu 455 460 465	2159
CGC TGG CTG ACC ATT CAG AAT GGC CCG CGG CAC TAT CCA ATC TAA Arg Trp Leu Thr Ile Gln Asn Gly Pro Arg His Tyr Pro Ile 470 475 480 481	2204
ATCGATCTTC GGGCGCCGCG GGCATCATGC CCGCGGCGCT CGCCTCATTT CAATCTCTAA	2264
CTTGATAAAA ACAGAGCTGT TCTCCGCTCT TGGTGGATCA AGGCCAGTCG CGGAGAGTCT	2324

CGAAGAGGAG AGTACAGTGA ACGCCGAGTC CACATTGCAA CCGCAGGCAT CATCATGCTC 2384
TGCTCAGCCA CGCTACCGCA GTGTGTCGAT TGGTCATCCT CCGGTTGAGG TTACGCAAGA 2444
CGCTGGAGGT ATTGTCCGGA TGC GTTCTCT CGAGGCGCTT CTTCCCTTCC CGGGTGGAAT 2504
TC 2506

FIG. 2n:

GAATTCCAAT AATGACAATA ATGAGGAGTG CCCA ATG TTT CAC GTG CCC CTG CTT	55
Met Phe His Val Pro Leu Leu	
1 5	
ATT GGT GGT AAG CCT TGT TCA GCA TCT GAT GAG CGC ACC TTC GAG CGT	103
Ile Gly Gly Lys Pro Cys Ser Ala Ser Asp Glu Arg Thr Phe Glu Arg	
10 15 20	
CGT AGC CCG CTG ACC GGA GAA GTG GTA TCG CGC GTC GCT GCT GCC AGT	151
Arg Ser Pro Leu Thr Gly Glu Val Val Ser Arg Val Ala Ala Ala Ser	
25 30 35	
TTG GAA GAT GCG GAC GCC GCA GTG GCC GCT GCA CAG GCT GCG TTT CCT	199
Leu Glu Asp Ala Asp Ala Ala Val Ala Ala Ala Gln Ala Ala Phe Pro	
40 45 50 55	
GAA TGG GCG GCG CTT GCT CCG AGC GAA CGC CGT GCC CGA CTG CTG CGA	247
Glu Trp Ala Ala Leu Ala Pro Ser Glu Arg Arg Ala Arg Leu Leu Arg	
60 65 70	
GCG GCG GAT CTT CTA GAG GAC CGT TCT TCC GAG TTC ACC GCC GCA GCG	295
Ala Ala Asp Leu Leu Glu Asp Arg Ser Ser Glu Phe Thr Ala Ala Ala	
75 80 85	
AGT GAA ACT GGC GCA GCG GGA AAC TGG TAT GGG TTT AAC GTT TAC CTG	343
Ser Glu Thr Gly Ala Ala Gly Asn Trp Tyr Gly Phe Asn Val Tyr Leu	
90 95 100	
GCG GCG GGC ATG TTG CGG GAA GCC GCG GCC ATG ACC ACA CAG ATT CAG	391
Ala Ala Gly Met Leu Arg Glu Ala Ala Ala Met Thr Thr Gln Ile Gln	
105 110 115	
GGC GAT GTC ATT CCG TCC AAT GTG CCC GGT AGC TTT GCC ATG GCG GTT	439
Gly Asp Val Ile Pro Ser Asn Val Pro Gly Ser Phe Ala Met Ala Val	
120 125 130 135	
CGA CAG CCA TGT GGC GTG GTG CTC GGT ATT GCG CCT TGG AAT GCT CCG	487
Arg Gln Pro Cys Gly Val Val Leu Gly Ile Ala Pro Trp Asn Ala Pro	
140 145 150	
GTA ATC CTT GGC GTA CGG GCT GTT GCG ATG CCG TTG GCA TGC GGC AAT	535
Val Ile Leu Gly Val Arg Ala Val Ala Met Pro Leu Ala Cys Gly Asn	
155 160 165	
ACC GTG GTG TTG AAA AGC TCT GAG CTG AGT CCC TTT ACC CAT CGC CTG	583
Thr Val Val Leu Lys Ser Ser Glu Leu Ser Pro Phe Thr His Arg Leu	
170 175 180	
ATT GGT CAG GTG TTG CAT GAT GCT GGT CTG GGG GAT GGC GTG GTG AAT	631
Ile Gly Gln Val Leu His Asp Ala Gly Leu Gly Asp Gly Val Val Asn	
185 190 195	
GTC ATC AGC AAT GCC CCG CAA GAC GCT CCT GCG GTG GTG GAG CGA CTG	679
Val Ile Ser Asn Ala Pro Gln Asp Ala Pro Ala Val Val Glu Arg Leu	
200 205 210 215	

ATT GCA AAT CCT GCG GTA CGT CGA GTG AAC TTC ACC GGT TCG ACC CAC Ile Ala Asn Pro Ala Val Arg Arg Val Asn Phe Thr Gly Ser Thr His 220 225 230	727
GTT GGA CGG ATC ATT GGT GAG CTG TCT GCG CGT CAT CTG AAG CCT GCT Val Gly Arg Ile Ile Gly Glu Leu Ser Ala Arg His Leu Lys Pro Ala 235 240 245	775
GTG CTG GAA TTA GGT GGT AAG GCT CCG TTC TTG GTC TTG GAC GAT GCC Val Leu Glu Leu Gly Gly Lys Ala Pro Phe Leu Val Leu Asp Asp Ala 250 255 260	823
GAC CTC GAT GCG GCG GTC GAA GCG GCG GCC TTT GGT GCC TAC TTC AAT Asp Leu Asp Ala Ala Val Glu Ala Ala Ala Phe Gly Ala Tyr Phe Asn 265 270 275	871
CAG GGT CAA ATC TGC ATG TCC ACT GAG CGT CTG ATT GTG ACA GCA GTC Gln Gly Gln Ile Cys Met Ser Thr Glu Arg Leu Ile Val Thr Ala Val 280 285 290 295	919
GCA GAC GCC TTT GTT GAA AAG CTG GCG AGG AAG GTC GCC ACA CTG CGT Ala Asp Ala Phe Val Glu Lys Leu Ala Arg Lys Val Ala Thr Leu Arg 300 305 310	967
GCT GGC GAT CCT AAT GAT CCG CAA TCG GTC TTG GGT TCG TTG ATT GAT Ala Gly Asp Pro Asn Asp Pro Gln Ser Val Leu Gly Ser Leu Ile Asp 315 320 325	1015
GCC AAT GCA GGT CAA CGC ATC CAG GTT CTG GTC GAT GAT GCG CTC GCA Ala Asn Ala Gly Gln Arg Ile Gln Val Leu Val Asp Asp Ala Leu Ala 330 335 340	1063
AAA GGC GCG CAATGGAA TTG GCC CAG CGC GTC GAT TCG GGC ATT TGC CAT Lys Gly Ala Leu Ala Gln Arg Val Asp Ser Gly Ile Cys His 345 346 420 425 430	1113
ATC AAT GGA CCG ACT GTG CAT GAC GAG GCT CAG ATG CCA TTC GGT GGG Ile Asn Gly Pro Thr Val His Asp Glu Ala Gln Met Pro Phe Gly Gly 435 440 445	1161
GTG AAG TCC AGC GGC TAC GGC AGC TTC GGC AGT CGA GCA TCG ATT GAG Val Lys Ser Ser Gly Tyr Gly Ser Phe Gly Ser Arg Ala Ser Ile Glu 450 455 460	1209
CAC TTT ACC CAG CTG CGC TGG CTG ACC ATT CAG AAT GGC CCG CGG CAC His Phe Thr Gln Leu Arg Trp Leu Thr Ile Gln Asn Gly Pro Arg His 465 470 475	1257
TAT CCA ATC TAA ATCGATCTTC GGGCGCCGCG GGCATCATGC CCGCGGCGCT Tyr Pro Ile 480 481	1309
CGCCTCATTTC CAATCTCTAA CTTGATAAAA ACAGAGCTGT TCTCCGGTCT TGGTGGATCA	1369
AGGCCAGTCG CGGAGAGTCT CGAAGAGGAG AGTACAGTGA ACGCCGAGTC CACATTGCAA	1429

CCGCAGGCAT CATCATGCTC TGCTCAGCCA CGCTACCGCA GTGTGTCGAT TGGTCATCCT	1489
CCGGTTGAGG TTACGCAAGA CGCTGGAGGT ATTGTCCGGA TCGGTTCTCT CGAGGCGCTT	1549
CTTCCCTTCC CGGGTGGAAT TC	1571

FIG. 2o:

GAATTCGCG GTCGGCGAAA GTTGATGCGC TGTATCGTGG TGAAGATCAA TCCATGCTGC	60
GTGACGAGGC CACACT GTG AGT TGG TCA GGG GGG GCT TAC TCG GCG TTT TCC Met Ser Trp Ser Gly Gly Ala Tyr Ser Ala Phe Ser 1 5 10	112
GAC ACT GCG TTG GTT GCG GCA GTG CGC ACC CCC TGG ATT GAT TGC GGG Asp Thr Ala Leu Val Ala Ala Val Arg Thr Pro Trp Ile Asp Cys Gly 15 20 25	160
GGT GCC CTG TCG CTG GTG TCG CCT ATC GAC TTA GGG GTA AAG GTC GCT Gly Ala Leu Ser Leu Val Ser Pro Ile Asp Leu Gly Val Lys Val Ala 30 35 40	208
CGC GAA GTT CTG ATG CGT GCG TCG CTT GAA CCA CAA ATG GTC GAT AGC Arg Glu Val Leu Met Arg Ala Ser Leu Glu Pro Gln Met Val Asp Ser 45 50 55 60	256
GTA CTC GCA GGC TCT ATG GCT CAA GCA AGC TTT GAT GCT TAC CTG CTC Val Leu Ala Gly Ser Met Ala Gln Ala Ser Phe Asp Ala Tyr Leu Leu 65 70 75	304
CCG CGG CAC ATT GGC TTG TAC AGC GGT GTT CCC AAG TCG GTT CCG GCC Pro Arg His Ile Gly Leu Tyr Ser Gly Val Pro Lys Ser Val Pro Ala 80 85 90	352
TTG GGG GTG CAG CGC ATT TGC GGC ACA GGC TTC GAA CTG CTT CGG CAG Leu Gly Val Gln Arg Ile Cys Gly Thr Gly Phe Glu Leu Leu Arg Gln 95 100 105	400
GCC GGC GAG CAG ATT TCC CAA GGC GCT GAT CAC GTG CTG TGT GTC GCG Ala Gly Glu Gln Ile Ser Gln Gly Ala Asp His Val Leu Cys Val Ala 110 115 120	448
GCA GAG TCC ATG TCG CGT AAC CCC ATC GCG TCG TAT ACA CAC CGG GGC Ala Glu Ser Met Ser Arg Asn Pro Ile Ala Ser Tyr Thr His Arg Gly 125 130 135 140	496
GGG TTC CGC CTC GGT GCG CCC GTT GAG TTC AAG GAT TTT TTG TGG GAG Gly Phe Arg Leu Gly Ala Pro Val Glu Phe Lys Asp Phe Leu Trp Glu 145 150 155	544
GCA TTG TTT GAT CCT GCT CCA GGA CTC GAC ATG ATC GCT ACC GCA GAA Ala Leu Phe Asp Pro Ala Pro Gly Leu Asp Met Ile Ala Thr Ala Glu 160 165 170	592
AAC CTG GGGACAGCAA GCGAACCGGA ATTGCCAGCT GGGGCGCCCT CTGGTAAGGT Asn Leu 174	648
TGGGAAGCCC TGCAAAGTAA ACTGGATGGC TTTCTTGCCG CCAAGGATCT GATGGCGCAG	708
GGGATCAAGA TCTGATCAAG AGACAGGATG AGGATCGTTT CGC ATG ATT GAA CAA Met Ile Glu Gln 1	763

GAT Asp 5	GGA Gly	TTG Leu	CAC His	GCA Ala	GGT Gly	TCT Ser	CCG Pro	GCC Ala	GCT Ala	TGG Trp	GTG Val	GAG Glu	AGG Arg	CTA Leu	TTC Phe	811
GGC Gly	TAT Tyr	GAC Asp	TGG Trp	GCA Ala	CAA Gln	CAG Gln	ACA Thr	ATC Ile	GGC Gly	TGC Cys	TCT Ser	GAT Asp	GCC Ala	GCC Ala	GTG Val	859
TTC Phe	CGG Arg	CTG Leu	TCA Ser	GCG Ala	CAG Gln	GGG Gly	CGC Arg	CCG Pro	GTT Val	CTT Leu	TTT Phe	GTC Val	AAG Lys	ACC Thr	GAC Asp	907
CTG Leu	TCC Ser	GGT Gly	GCC Ala	CTG Leu	AAT Asn	GAA Glu	CTG Leu	CAG Gln	GAC Asp	GAG Glu	GCA Ala	GCG Ala	CGG Arg	CTA Leu	TCG Ser	955
TGG Trp	CTG Leu	GCC Ala	ACG Thr	ACG Thr	GGC Gly	GTT Val	CCT Pro	TGC Cys	GCA Ala	GCT Ala	GTG Val	CTC Leu	GAC Asp	GTT Val	GTC Val	1003
ACT Thr 85	GAA Glu	GCG Ala	GGA Gly	AGG Arg	GAC Asp	TGG Trp	CTG Leu	CTA Leu	TTG Leu	GGC Gly	GAA Glu	GTG Val	CCG Pro	GGG Gly	CAG Gln 100	1051
GAT Asp	CTC Leu	CTG Leu	TCA Ser	TCT Ser	CAC His	CTT Leu	GCT Ala	CCT Pro	GCC Ala	GAG Glu	AAA Lys	GTA Val	TCC Ser	ATC Ile	ATG Met	1099
GCT Ala	GAT Asp	GCA Ala	ATG Met	CGG Arg	CGG Arg	CTG Leu	CAT His	ACG Thr	CTT Leu	GAT Asp	CCG Pro	GCT Ala	ACC Thr	TGC Cys	CCA Pro	1147
TTC Phe	GAC Asp	CAC His	CAA Gln	GCG Ala	AAA Lys	CAT His	CGC Arg	ATC Ile	GAG Glu	CGA Arg	GCA Ala	CGT Arg	ACT Thr	CGG Arg	ATG Met	1195
GAA Glu 150	GCC Ala	GGT Gly	CTT Leu	GTC Val	GAT Asp	CAG Gln	GAT Asp	GAT Asp	CTG Leu	GAC Asp	GAA Glu	GAG Glu	CAT His	CAG Gln	GGG Gly	1243
CTC Leu 165	GCG Ala	CCA Pro	GCC Ala	GAA Glu	CTG Leu	TTC Phe	GCC Ala	AGG Arg	CTC Leu	AAG Lys	GCG Ala	CGC Arg	ATG Met	CCC Pro	GAC Asp 180	1291
GGC Gly	GAG Glu	GAT Asp	CTC Leu	GTC Val	GTG Val	ACC Thr	CAT His	GGC Gly	GAT Asp	GCC Ala	TGC Cys	TTG Leu	CCG Pro	AAT Asn	ATC Ile	1339
ATG Met	GTG Val	GAA Glu	AAT Asn	GGC Gly	CGC Arg	TTT Phe	TCT Ser	GGA Gly	TTC Phe	ATC Ile	GAC Asp	TGT Cys	GGC Gly	CGG Arg	CTG Leu	1387
GGT Gly	GTG Val	GCG Ala	GAC Asp	CGC Arg	TAT Tyr	CAG Gln	GAC Asp	ATA Ile	GCG Ala	TTG Leu	GCT Ala	ACC Thr	CGT Arg	GAT Asp	ATT Ile	1435

GCT GAA GAG CTT GGC GGC GAA TGG GCT GAC CGC TTC CTC GTG CTT TAC Ala Glu Glu Leu Gly Gly Glu Trp Ala Asp Arg Phe Leu Val Leu Tyr 230 235 240	1483
GGT ATC GCC GCT CCC GAT TCG CAG CGC ATC GCC TTC TAT CGC CTT CTT Gly Ile Ala Ala Pro Asp Ser Gln Arg Ile Ala Phe Tyr Arg Leu Leu 245 250 255 260	1531
GAC GAG TTC TTC TGA GCGGGACTCT GGGGTTCGAA ATGACCGACC AAGCGACGCC Asp Glu Phe Phe 264	1586
CA TTG AGG GCG CAA GAG GAG AAA TGG ATT GAC CAA GAG ATC GTG GCT Leu Arg Ala Gln Glu Glu Lys Trp Ile Asp Gln Glu Ile Val Ala 197 200 205 210	1633
GTT ACG GAT GAA CAG TTC GAT TTA GAG GGC TAC AAC AGT CGA GCA ATT Val Thr Asp Glu Gln Phe Asp Leu Glu Gly Tyr Asn Ser Arg Ala Ile 215 220 225	1681
GAA CTG CCT CGG AAG GCA AAA TTG TTG ATC GTG ACA GTC ATC CGC GGC Glu Leu Pro Arg Lys Ala Lys Leu Leu Ile Val Thr Val Ile Arg Gly 230 235 240	1729
CTA GCA GTC TTT GAA GCC CTT TCC CGA TTG AAG CCT GTT CAT TCT GGC Leu Ala Val Phe Glu Ala Leu Ser Arg Leu Lys Pro Val His Ser Gly 245 250 255	1777
GGG GTG CAG ACT GCG GGC AAC AGC TGT GCC GTA GTG GAC GGC GCC GCG Gly Val Gln Thr Ala Gly Asn Ser Cys Ala Val Val Asp Gly Ala Ala 260 265 270 275	1825
GCG GCT TTG GTG GCT CGA GAG TCG TCT GCG ACA CAG CCG GTC TTG GCT Ala Ala Leu Val Ala Arg Glu Ser Ser Ala Thr Gln Pro Val Leu Ala 280 285 290	1873
AGG ATA CTG GCT ACC TCC GTA GTC GGG ATC GAG CCC GAG CAT ATG GGG Arg Ile Leu Ala Thr Ser Val Val Gly Ile Glu Pro Glu His Met Gly 295 300 305	1921
CTC GGC CCT GCG CCC GCG ATT CGC CTG CTG CTT GCG CGT AGT GAT CTT Leu Gly Pro Ala Pro Ala Ile Arg Leu Leu Leu Ala Arg Ser Asp Leu 310 315 320	1969
AGT TTG AGG GAT ATC GAC CTC TTT GAG ATA AAC GAG GCG CAG GCC GCC Ser Leu Arg Asp Ile Asp Leu Phe Glu Ile Asn Glu Ala Gln Ala Ala 325 330 335	2017
CAA GTT CTA GCG GTA CAG CAT GAA TTG GGT ATT GAG CAC TCA AAA CTT Gln Val Leu Ala Val Gln His Glu Leu Gly Ile Glu His Ser Lys Leu 340 345 350 355	2065
AAT ATT TGG GGC GGG GCC ATT GCA CTT GGA CAC CCG CTT GCC GCG ACC Asn Ile Trp Gly Gly Ala Ile Ala Leu Gly His Pro Leu Ala Ala Thr 360 365 370	2113

GGA TTG CGT CTC TGC ATG ACC CTC GCT CAC CAA TTG CAA GCT AAT AAC	2161
Gly Leu Arg Leu Cys Met Thr Leu Ala His Gln Leu Gln Ala Asn Asn	
375 380 385	
TTT CGA TAT GGA ATT GCC TCG GCA TGC ATT GGT GGG GGA CAG GGG ATG	2209
Phe Arg Tyr Gly Ile Ala Ser Ala Cys Ile Gly Gly Gly Gln Gly Met	
390 395 400	
GCG GTT CTT TTA GAG AAT CCC CAC TTC GGT TCG TCC TCT GCA CGA AGT	2257
Ala Val Leu Leu Glu Asn Pro His Phe Gly Ser Ser Ser Ala Arg Ser	
405 410 415	
TCG ATG ATT AAC AGA GTT GAC CAC TAT CCA CTG AGC TAA CGGGCATCTC	2306
Ser Met Ile Asn Arg Val Asp His Tyr Pro Leu Ser	
420 425 430 431	
CTTTGTTGCT TTGAGGTGGC GCACGAAGGA GGGCTCGAAA ATCTCTGCTA AAAACAAGAA	2366
GAAGGAACAG GGAACATGAT TAGTTTCGCT CGTATGGCAG AAAGTTTAGG AGTCCAGGCT	2426
AAACTTGCCC TTGCCTTCGC ACTCGTATTA TGTGTCGGGC TGATTGTTAC CGGCACGGGT	2486
TTCTACAGTG TACATACCTT GTCAGGGTTG GTGGGAATTC	2526

FIG. 2p:

GAATTCGCG GTCGGCGAAA GTTGATGCGC TGTATCGTGG TGAAGATCAA TCCATGCTGC	60
GTGACGAGGC CACACT GTG AGT TGG TCA GGG GGG GCT TAC TCG GCG TTT TCC	112
Met Ser Trp Ser Gly Gly Ala Tyr Ser Ala Phe Ser	
1 5 10	
GAC ACT GCG TTG GTT GCG GCA GTG CGC ACC CCC TGG ATT GAT TGC GGG	160
Asp Thr Ala Leu Val Ala Ala Val Arg Thr Pro Trp Ile Asp Cys Gly	
15 20 25	
GGT GCC CTG TCG CTG GTG TCG CCT ATC GAC TTA GGG GTA AAG GTC GCT	208
Gly Ala Leu Ser Leu Val Ser Pro Ile Asp Leu Gly Val Lys Val Ala	
30 35 40	
CGC GAA GTT CTG ATG CGT GCG TCG CTT GAA CCA CAA ATG GTC GAT AGC	256
Arg Glu Val Leu Met Arg Ala Ser Leu Glu Pro Gln Met Val Asp Ser	
45 50 55 60	
GTA CTC GCA GGC TCT ATG GCT CAA GCA AGC TTT GAT GCT TAC CTG CTC	304
Val Leu Ala Gly Ser Met Ala Gln Ala Ser Phe Asp Ala Tyr Leu Leu	
65 70 75	
CCG CGG CAC ATT GGC TTG TAC AGC GGT GTT CCC AAG TCG GTT CCG GCC	352
Pro Arg His Ile Gly Leu Tyr Ser Gly Val Pro Lys Ser Val Pro Ala	
80 85 90	
TTG GGG GTG CAG CGC ATT TGC GGC ACA GGC TTC GAA CTG CTT CGG CAG	400
Leu Gly Val Gln Arg Ile Cys Gly Thr Gly Phe Glu Leu Leu Arg Gln	
95 100 105	
GCC GGC GAG CAG ATT TCC CAA GGC GCT GAT CAC GTG CTG TGT GTC GCG	448
Ala Gly Glu Gln Ile Ser Gln Gly Ala Asp His Val Leu Cys Val Ala	
110 115 120	
GCA GAG TCC ATG TCG CGT AAC CCC ATC GCG TCG TAT ACA CAC CGG GGC	496
Ala Glu Ser Met Ser Arg Asn Pro Ile Ala Ser Tyr Thr His Arg Gly	
125 130 135 140	
GGG TTC CGC CTC GGT GCG CCC GTT GAG TTC AAG GAT TTT TTG TGG GAG	544
Gly Phe Arg Leu Gly Ala Pro Val Glu Phe Lys Asp Phe Leu Trp Glu	
145 150 155	
GCA TTG TTT GAT CCT GCT CCA GGA CTC GAC ATG ATC GCT ACC GCA GAA	592
Ala Leu Phe Asp Pro Ala Pro Gly Leu Asp Met Ile Ala Thr Ala Glu	
160 165 170	
AAC CTG GGGGAGAGGC GGTTTGCCTA TTGGGCGCAT GCATAAAAAC TGTTGTAATT	648
Asn Leu	
174	
CATTAAGCAT TCTGCCGACA TGGAAGCCAT CACAAACGGC ATGATGAACC TGAATCGCCA	708
GCGGCATCAG CACCTTGTCG CCTTGCGTAT AATATTTGCC CATGGACGCA CACCGTGGAA	768
ACGGATGAAG GCACGAACCC AGTTGACATA AGCCTGTTTC GTTCGTAAAC TGTAATGCAA	828
GTAGCGTATG CGCTCAGCA ACTGGTCCAG AACCTTGACC GAACGCAGCG GTGGTAACGG	888

CGCAGTGGCG	GTTTT	CATGG	CTTGTTATGA	CTGTTTTTTT	GTACAGTCTA	TGCCTCGGGC	948
ATCCAAGC	AGCAAGCGCG	TTACGCCGTG	GGTCGATGTTTG	ATGTTATGGA	GCAGCAACG	1007	
ATG TTA CGC AGC AGC AAC GAT GTT ACG CAG CAG GGC AGT CGC CCT AAA	1055						
Met Leu Arg Ser Ser Asn Asp Val Thr Gln Gln Gly Ser Arg Pro Lys							
1 5 10 15							
ACA AAG TTA GGT GGC TCA AGT ATG GGC ATC ATT CGC ACA TGT AGG CTC	1103						
Thr Lys Leu Gly Gly Ser Ser Met Gly Ile Ile Arg Thr Cys Arg Leu							
20 25 30							
GGC CCT GAC CAA GTC AAA TCC ATG CGG GCT GCT CTT GAT CTT TTC GGT	1151						
Gly Pro Asp Gln Val Lys Ser Met Arg Ala Ala Leu Asp Leu Phe Gly							
35 40 45							
CGT GAG TTC GGA GAC GTA GCC ACC TAC TCC CAA CAT CAG CCG GAC TCC	1199						
Arg Glu Phe Gly Asp Val Ala Thr Tyr Ser Gln His Gln Pro Asp Ser							
50 55 60							
GAT TAC CTC GGG AAC TTG CTC CGT AGT AAG ACA TTC ATC GCG CTT GCT	1247						
Asp Tyr Leu Gly Asn Leu Leu Arg Ser Lys Thr Phe Ile Ala Leu Ala							
65 70 75 80							
GCC TTC GAC CAA GAA GCG GTT GTT GGC GCT CTC GCG GCT TAC GTT CTG	1295						
Ala Phe Asp Gln Glu Ala Val Val Gly Ala Leu Ala Ala Tyr Val Leu							
85 90 95							
CCC AGG TTT GAG CAG CCG CGT AGT GAG ATC TAT ATC TAT GAT CTC GCA	1343						
Pro Arg Phe Glu Gln Pro Arg Ser Glu Ile Tyr Ile Tyr Asp Leu Ala							
100 105 110							
GTC TCC GGC GAG CAC CGG AGG CAG GGC ATT GCC ACC GCG CTC ATC AAT	1391						
Val Ser Gly Glu His Arg Arg Gln Gly Ile Ala Thr Ala Leu Ile Asn							
115 120 125							
CTC CTC AAG CAT GAG GCC AAC GCG CTT GGT GCT TAT GTG ATC TAC GTG	1439						
Leu Leu Lys His Glu Ala Asn Ala Leu Gly Ala Tyr Val Ile Tyr Val							
130 135 140							
CAA GCA GAT TAC GGT GAC GAT CCC GCA GTG GCT CTC TAT ACA AAG TTG	1487						
Gln Ala Asp Tyr Gly Asp Asp Pro Ala Val Ala Leu Tyr Thr Lys Leu							
145 150 155 160							
GGC ATA CGG GAA GAA GTG ATG CAC TTT GAT ATC GAC CCA AGT ACC GCC	1535						
Gly Ile Arg Glu Glu Val Met His Phe Asp Ile Asp Pro Ser Thr Ala							
165 170 175							
ACC TAA CAATTCGTTT AAGCCGAGAT CGGCTTCCCA TTG AGG GCG CAA GAG GAG	1589						
Thr Leu Arg Ala Gln Glu Glu							
177 197 200							
AAA TGG ATT GAC CAA GAG ATC GTG GCT GTT ACG GAT GAA CAG TTC GAT	1637						
Lys Trp Ile Asp Gln Glu Ile Val Ala Val Thr Asp Glu Gln Phe Asp							
205 210 215							

TTA GAG GGC TAC AAC AGT CGA GCA ATT GAA CTG CCT CGG AAG GCA AAA Leu Glu Gly Tyr Asn Ser Arg Ala Ile Glu Leu Pro Arg Lys Ala Lys 220 225 230	1685
TTG TTG ATC GTG ACA GTC ATC CGC GGC CTA GCA GTC TTT GAA GCC CTT Leu Leu Ile Val Thr Val Ile Arg Gly Leu Ala Val Phe Glu Ala Leu 235 240 245 250	1733
TCC CGA TTG AAG CCT GTT CAT TCT GGC GGG GTG CAG ACT GCG GGC AAC Ser Arg Leu Lys Pro Val His Ser Gly Gly Val Gln Thr Ala Gly Asn 255 260 265	1781
AGC TGT GCC GTA GTG GAC GGC GCC GCG GCG GCT TTG GTG GCT CGA GAG Ser Cys Ala Val Val Asp Gly Ala Ala Ala Leu Val Ala Arg Glu 270 275 280	1829
TCG TCT GCG ACA CAG CCG GTC TTG GCT AGG ATA CTG GCT ACC TCC GTA Ser Ser Ala Thr Gln Pro Val Leu Ala Arg Ile Leu Ala Thr Ser Val 285 290 295	1877
GTC GGG ATC GAG CCC GAG CAT ATG GGG CTC GGC CCT GCG CCC GCG ATT Val Gly Ile Glu Pro Glu His Met Gly Leu Gly Pro Ala Pro Ala Ile 300 305 310	1925
CGC CTG CTG CTT GCG CGT AGT GAT CTT AGT TTG AGG GAT ATC GAC CTC Arg Leu Leu Leu Ala Arg Ser Asp Leu Ser Leu Arg Asp Ile Asp Leu 315 320 325 330	1973
TTT GAG ATA AAC GAG GCG CAG GCC GCC CAA GTT CTA GCG GTA CAG CAT Phe Glu Ile Asn Glu Ala Gln Ala Ala Gln Val Leu Ala Val Gln His 335 340 345	2021
GAA TTG GGT ATT GAG CAC TCA AAA CTT AAT ATT TGG GGC GGG GCC ATT Glu Leu Gly Ile Glu His Ser Lys Leu Asn Ile Trp Gly Gly Ala Ile 350 355 360	2069
GCA CTT GGA CAC CCG CTT GCC GCG ACC GGA TTG CGT CTC TGC ATG ACC Ala Leu Gly His Pro Leu Ala Ala Thr Gly Leu Arg Leu Cys Met Thr 365 370 375	2117
CTC GCT CAC CAA TTG CAA GCT AAT AAC TTT CGA TAT GGA ATT GCC TCG Leu Ala His Gln Leu Gln Ala Asn Asn Phe Arg Tyr Gly Ile Ala Ser 380 385 390	2165
GCA TGC ATT GGT GGG GGA CAG GGG ATG GCG GTT CTT TTA GAG AAT CCC Ala Cys Ile Gly Gly Gly Gln Gly Met Ala Val Leu Leu Glu Asn Pro 395 400 405 410	2213
CAC TTC GGT TCG TCC TCT GCA CGA AGT TCG ATG ATT AAC AGA GTT GAC His Phe Gly Ser Ser Ser Ala Arg Ser Ser Met Ile Asn Arg Val Asp 415 420 425	2261
CAC TAT CCA CTG AGC TAA CGGGCATCTC CTTTGTTGCT TTGAGGTGGC His Tyr Pro Leu Ser 430 431	2309

GCACGAAGGA GGGCTCGAAA ATCTCTGCTA AAAACAAGAA GAAGGAACAG GGAACATGAT	2369
TAGTTTCGCT CGTATGGCAG AAAGTTTAGG AGTCCAGGCT AAACCTGCCC TTGCCTTCGC	2429
ACTCGTATTA TGTGTCGGGC TGATTGTTAC CGGCACGGGT TTCTACAGTG TACATACCTT	2489
GTCAGGGTTG GTGGGAATTC	2509

FIG. 2q:

GAATTCGCGC GTCGGCGAAA GTTGATGCGC TGTATCGTGG TGAAGATCAA TCCATGCTGC	60
GTGACGAGGC CACACT GTG AGT TGG TCA GGG GGG GCT TAC TCG GCG TTT TCC	112
Met Ser Trp Ser Gly Gly Ala Tyr Ser Ala Phe Ser	
1 5 10	
GAC ACT GCG TTG GTT GCG GCA GTG CGC ACC CCC TGG ATT GAT TGC GGG	160
Asp Thr Ala Leu Val Ala Ala Val Arg Thr Pro Trp Ile Asp Cys Gly	
15 20 25	
GGT GCC CTG TCG CTG GTG TCG CCT ATC GAC TTA GGG GTA AAG GTC GCT	208
Gly Ala Leu Ser Leu Val Ser Pro Ile Asp Leu Gly Val Lys Val Ala	
30 35 40	
CGC GAA GTT CTG ATG CGT GCG TCG CTT GAA CCA CAA ATG GTC GAT AGC	256
Arg Glu Val Leu Met Arg Ala Ser Leu Glu Pro Gln Met Val Asp Ser	
45 50 55 60	
GTA CTC GCA GGC TCT ATG GCT CAA GCA AGC TTT GAT GCT TAC CTG CTC	304
Val Leu Ala Gly Ser Met Ala Gln Ala Ser Phe Asp Ala Tyr Leu Leu	
65 70 75	
CCG CGG CAC ATT GGC TTG TAC AGC GGT GTT CCC AAG TCG GTT CCG GCC	352
Pro Arg His Ile Gly Leu Tyr Ser Gly Val Pro Lys Ser Val Pro Ala	
80 85 90	
TTG GGG GTG CAG CGC ATT TGC GGC ACA GGC TTC GAA CTG CTT CGG CAG	400
Leu Gly Val Gln Arg Ile Cys Gly Thr Gly Phe Glu Leu Leu Arg Gln	
95 100 105	
GCC GGC GAG CAG ATT TCC CAA GGC GCT GAT CAC GTG CTG TGT GTC GCG	448
Ala Gly Glu Gln Ile Ser Gln Gly Ala Asp His Val Leu Cys Val Ala	
110 115 120	
GCA GAG TCC ATG TCG CGT AAC CCC ATC GCG TCG TAT ACA CAC CGG GGC	496
Ala Glu Ser Met Ser Arg Asn Pro Ile Ala Ser Tyr Thr His Arg Gly	
125 130 135 140	
GGG TTC CGC CTC GGT GCG CCC GTT GAG TTC AAG GAT TTT TTG TGG GAG	544
Gly Phe Arg Leu Gly Ala Pro Val Glu Phe Lys Asp Phe Leu Trp Glu	
145 150 155	
GCA TTG TTT GAT CCT GCT CCA GGA CTC GAC ATG ATC GCT ACC GCA GAA	592
Ala Leu Phe Asp Pro Ala Pro Gly Leu Asp Met Ile Ala Thr Ala Glu	
160 165 170	
AAC CTG GCG CGC A TTG AGG GCG CAA GAG GAG AAA TGG ATT GAC CAA GAG	641
Asn Leu Ala Arg Leu Arg Ala Gln Glu Lys Trp Ile Asp Gln Glu	
175 176 197 200 205	
ATC GTG GCT GTT ACG GAT GAA CAG TTC GAT TTA GAG GGC TAC AAC AGT	689
Ile Val Ala Val Thr Asp Glu Gln Phe Asp Leu Glu Gly Tyr Asn Ser	
210 215 220	
CGA GCA ATT GAA CTG CCT CGG AAG GCA AAA TTG TTG ATC GTG ACA GTC	737
Arg Ala Ile Glu Leu Pro Arg Lys Ala Lys Leu Leu Ile Val Thr Val	
225 230 235 240	

ATC CGC GGC CTA GCA GTC TTT GAA GCC CTT TCC CGA TTG AAG CCT GTT Ile Arg Gly Leu Ala Val Phe Glu Ala Leu Ser Arg Leu Lys Pro Val 245 250 255	785
CAT TCT GGC GGG GTG CAG ACT GCG GGC AAC AGC TGT GCC GTA GTG GAC His Ser Gly Gly Val Gln Thr Ala Gly Asn Ser Cys Ala Val Val Asp 260 265 270	833
GGC GCC GCG GCG GCT TTG GTG GCT CGA GAG TCG TCT GCG ACA CAG CCG Gly Ala Ala Ala Leu Val Ala Arg Glu Ser Ser Ala Thr Gln Pro 275 280 285	881
GTC TTG GCT AGG ATA CTG GCT ACC TCC GTA GTC GGG ATC GAG CCC GAG Val Leu Ala Arg Ile Leu Ala Thr Ser Val Val Gly Ile Glu Pro Glu 290 295 300	929
CAT ATG GGG CTC GGC CCT GCG CCC GCG ATT CGC CTG CTG CTT GCG CGT His Met Gly Leu Gly Pro Ala Pro Ala Ile Arg Leu Leu Leu Ala Arg 305 310 315 320	977
AGT GAT CTT AGT TTG AGG GAT ATC GAC CTC TTT GAG ATA AAC GAG GCG Ser Asp Leu Ser Leu Arg Asp Ile Asp Leu Phe Glu Ile Asn Glu Ala 325 330 335	1025
CAG GCC GCC CAA GTT CTA GCG GTA CAG CAT GAA TTG GGT ATT GAG CAC Gln Ala Ala Gln Val Leu Ala Val Gln His Glu Leu Gly Ile Glu His 340 345 350	1073
TCA AAA CTT AAT ATT TGG GGC GGG GCC ATT GCA CTT GGA CAC CCG CTT Ser Lys Leu Asn Ile Trp Gly Gly Ala Ile Ala Leu Gly His Pro Leu 355 360 365	1121
GCC GCG ACC GGA TTG CGT CTC TGC ATG ACC CTC GCT CAC CAA TTG CAA Ala Ala Thr Gly Leu Arg Leu Cys Met Thr Leu Ala His Gln Leu Gln 370 375 380	1169
GCT AAT AAC TTT CGA TAT GGA ATT GCC TCG GCA TGC ATT GGT GGG GGA Ala Asn Asn Phe Arg Tyr Gly Ile Ala Ser Ala Cys Ile Gly Gly Gly 385 390 395 400	1217
CAG GGG ATG GCG GTT CTT TTA GAG AAT CCC CAC TTC GGT TCG TCC TCT Gln Gly Met Ala Val Leu Leu Glu Asn Pro His Phe Gly Ser Ser Ser 405 410 415	1265
GCA CGA AGT TCG ATG ATT AAC AGA GTT GAC CAC TAT CCA CTG AGC TAA Ala Arg Ser Ser Met Ile Asn Arg Val Asp His Tyr Pro Leu Ser 420 425 430 431	1313
CGGGCATCTC CTTTGTGCT TTGAGGTGGC GCACGAAGGA GGGCTCGAAA ATCTCTGCTA	1373
AAAACAAGAA GAAGGAACAG GGAACATGAT TAGTTTCGCT CGTATGGCAG AAAGTTTAGG	1433
AGTCCAGGCT AAACCTGCCC TTGCCTTCGC ACTCGTATTA TGTGTCGGGC TGATTGTTAC	1493
CGGCACGGGT TTCTACAGTG TACATACCTT GTCAGGGTTG GTGGGAATTC	1543

FIG. 2r:

Sequenz 1

CTGCAGCCAG	GGCTGAAAAG	GAGGGATTCA	GTGAGGTCAT	GAAGGGAGGG	GACGGCGCCT	60
GGCTCCAATT	GCTCGATGGC	GCCGCGATTG	AGTGTCTTGG	GCGCGGTCTT	GGAGAGTTCCG	120
GCTAGGGAGA	TAAATTTGCT	GGCCATGGTG	GCGGCCCTTG	ATGGGTTGGA	TGATTTTCTG	180
CATTCTGCAT	CATGAAATTC	ATGAAATCAT	CACTTTTCGG	GGGGTGGGTG	CACGGGATTG	240
AAGGTTGCTA	GGAGAGTGCA	TTGCTCGTAA	GCCCAGGAAG	CACGCGGGTT	TCAGGATGGT	300
GCATGGAAAT	GGCATGAGCT	TTGCTGGATA	TGATTAGAGA	CATTAACAT	TTTGGCGGAA	360
TGGAAGCACG	ATTCCTCGCC	CGGTAGAGCG	GTAACCGCGA	CATTCAGGAC	CGTAAAAAGG	420
AAAGAGCATG	CAACTGACCA	ACAAGAAAAT	CGTCGTCACC	GGAGTGTCTT	CCGGTATCGG	480
TGCCGAAACT	GCCCGCGTTC	TGCGCTCTCA	CGGCGCCACA	GTGATTGGCG	TAGATCGCAA	540
CATGCCGAGC	CTGACTCTGG	ATGCTTTTCT	TCAGGCTGAC	CTGAGCCATC	CTGAAGGCAT	600
CGATAAGGCC	ATCGGGACAG	CAAGCGAACC	GGAATTGCCA	GCTGGGGCGC	CCTCTGGTAA	660
GGTTGGGAAG	CCCTGCAAAG	TAAACTGGAT	GGCTTTCTTG	CCGCCAAGGA	TCTGATGGCG	720
CAGGGGATCA	AGATCTGATC	AAGAGACAGG	ATGAGGATCG	TTTCGCATGA	TTGAACAAGA	780
TGGATTGCAC	GCAGGTTCTC	CGGCCGCTTG	GGTGGAGAGG	CTATTTCGGCT	ATGACTGGGC	840
ACAACAGACA	ATCGGCTGCT	CTGATGCCGC	CGTGTTCGGG	CTGTCAGCGC	AGGGGCGCCC	900
GGTTCTTTTT	GTCAAGACCG	ACCTGTCCGG	TGCCCTGAAT	GAAGTGCAGG	ACGAGGCAGC	960
GCGGCTATCG	TGGCTGGCCA	CGACGGGCGT	TCCTTGCGCA	GCTGTGCTCG	ACGTTGTCAC	1020
TGAAGCGGGA	AGGGACTGGC	TGCTATTGGG	CGAAGTGCCG	GGGCAGGATC	TCCTGTATC	1080
TCACCTTGCT	CCTGCCGAGA	AAGTATCCAT	CATGGCTGAT	GCAATGCGGC	GGCTGCATAC	1140
GCTTGATCCG	GCTACCTGCC	CATTCGACCA	CCAAGCGAAA	CATCGCATCG	AGCGAGCACG	1200
TACTCGGATG	GAAGCCGGTC	TTGTCGATCA	GGATGATCTG	GACGAAGAGC	ATCAGGGGCT	1260
CGCGCCAGCC	GAAGTGTTCG	CCAGGCTCAA	GGCGCGCATG	CCCGACGGCG	AGGATCTCGT	1320
CGTGACCCAT	GGCGATGCCT	GCTTGCCGAA	TATCATGGTG	GAAAATGGCC	GCTTTTCTGG	1380
ATTCATCGAC	TGTGGCCGGC	TGGGTGTGGC	GGACCGCTAT	CAGGACATAG	CGTTGGCTAC	1440
CCGTGATATT	GCTGAAGAGC	TTGGCGGCGA	ATGGGCTGAC	CGCTTCCTCG	TGCTTTACGG	1500
TATCGCCGCT	CCCGATTTCG	AGCGCATCGC	CTTCTATCGC	CTTCTTGACG	AGTTCTTCTG	1560
AGCGGGACTC	TGGGGTTTCA	AATGACCGAC	CAAGCGACGC	CCTGGCCGCG	GTGATTGCAT	1620
TCATGTGTGC	TGAGGAGTCA	CGTTGGATCA	ACGGCATAAA	TATTCCAGTG	GACGGAGGTT	1680
TGGCATCGAC	CTACGTGTAA	GTTCTGGGAC	GCCCTTTGCA	CGCGCACTAT	ATCTCTATGC	1740
AGCAGCTGAA	AGCAGCTTTG	GTTTTGATCG	GAGGTAGCGG	GCGGAAAGGT	GCAGAATGTC	1800
TAAATAATAA	AGGATTCTTG	TGAAGCTTTA	GTTGTCCGTA	AACGAAAATA	AAAATAAAGA	1860
GGAATGATAT	GAAAGCAAGT	AGATCAGTCT	GCACCTTCAA	AATAGCTACC	CTGGCAGGCG	1920
CCATTTATGC	AGCGCTGCCA	ATGTCAGCTG	CAAACCTCGAT	GCAGCTGGAT	GTAGGTAGCT	1980
CGGATTGGAC	GGTGCGTTGG	GGACAACACC	CTCAAGTATA	GCCTTGCCCTC	TCGCCTGAAT	2040
GAGCAAGACT	CAAGTCTGAC	AAATGCGCCG	ACTGTCAATG	GTTATATCCG	GATATTCAAA	2100
GTCAGGGTGA	TCGTAACCTT	GACCGGGGGC	TTGGTATCCA	ATCGTCTCGA	TATTCTGGCT	2160
GCAG						2164

Sequenz 2

CTGCAGCCAG	GGCTGAAAAG	GAGGGATTCA	GTGAGGTCAT	GAAGGGAGGG	GACGGCGCCT	60
GGCTCCAATT	GCTCGATGGC	GCCGCGATTG	AGTGTCTTGG	GCGCGGTCTT	GGAGAGTTCC	120
GCTAGGGAGA	TAAATTTGCT	GGCCATGGTG	GCGGCCCTG	ATGGGTGGA	TGATTTTCTG	180
CATTCTGCAT	CATGAAATTC	ATGAAATCAT	CACTTTTTCGG	GGGGTGGGTG	CACGGGATTG	240
AAGGTTGCTA	GGAGAGTGCA	TTGCTCGTAA	GCCCAGGAAG	CACGCGGGTT	TCAGGATGGT	300
GCATGGAAAT	GGCATGAGCT	TTGCTGGATA	TGATTAGAGA	CATTAACAT	TTTGGCGGAA	360
TGGAAGCACG	ATTCCTCGCC	CGGTAGAGCG	GTAACCGCGA	CATTACAGGAC	CGTAAAAAGG	420
AAAGAGCATG	CAACTGACCA	ACAAGAAAAT	CGTCGTCACC	GGAGTGTCTT	CCGGTATCGG	480
TGCCGAAACT	GCCCGCGTTC	TGCGCTCTCA	CGGCGCCACA	GTGATTGGCG	TAGATCGCAA	540
CATGCCGAGC	CTGACTCTGG	ATGCTTTCGT	TCAGGCTGAC	CTGAGCCATC	CTGAGGGGAG	600
AGGCGGTTTG	CGTATTGGGC	GCATGCATAA	AACTGTTGT	AATTCATTAA	GCATTCTGCC	660
GACATGGAAAG	CCATCACAAA	CGGCATGATG	AACCTGAATC	GCCAGCGGCA	TCAGCACCTT	720
GTCGCCTTGC	GTATAATATT	TGCCCATGGA	CGCACACCGT	GGAAACGGAT	GAAGGCACGA	780
ACCCAGTTGA	CATAAGCCTG	TTCGGTTCGT	AACTGTAAT	GCAAGTAGCG	TATGGCGTCA	840
CGCAACTGGT	CCAGAACCCTT	GACCGAACGC	AGCGGTGGTA	ACGGCGCAGT	GGCGGTTTTTC	900
ATGGCTTGTT	ATGACTGTTT	TTTTGTACAG	TCTATGCCTC	GGGCATCCAA	GCAGCAAGCG	960
CGTTACGCCG	TGGGTCGATG	TTTGATGTTA	TGGAGCAGCA	ACGATGTTAC	GCAGCAGCAA	1020
CGATGTTACG	CAGCAGGGCA	GTCGCCCTAA	AACAAAGTTA	GGTGGCTCAA	GTATGGGCAT	1080
CATTGCGACA	TGTAGGCTCG	GCCCTGACCA	AGTCAAATCC	ATGCGGGCTG	CTCTTGATCT	1140
TTTCGGTTCG	GAGTTCGGAG	ACGTAGCCAC	CTACTCCCAA	CATCAGCCGG	ACTCCGATTA	1200
CCTCGGGAAC	TTGCTCCGTA	GTAAGACATT	CATCGCGCTT	GCTGCCTTCG	ACCAAGAAGC	1260
GGTTGTTGGC	GCTCTCGCGG	CTTACGTTCT	GCCCAGGTTT	GAGCAGCCGC	GTAGTGAGAT	1320
CTATATCTAT	GATCTCGCAG	TCTCCGGCGA	GCACCGGAGG	CAGGGCATTG	CCACCGCGCT	1380
CATCAATCTC	CTCAAGCATG	AGGCCAACGC	GCTTGGTGCT	TATGTGATCT	ACGTGCAAGC	1440
AGATTACGGT	GACGATCCCG	CAGTGGCTCT	CTATACAAAG	TTGGGCATAC	GGGAAGAAGT	1500
GATGCACTTT	GATATCGACC	CAAGTACCGC	CACCTAACAA	TTCGTTCAAG	CCGAGATCGG	1560
CTTCCCTGAT	TGCATTATG	TGTGCTGAGG	AGTCACGTTG	GATCAACGGC	ATAAATATTC	1620
CAGTGGACCG	AGGTTTGGCA	TCGACCTACG	TGTAAGTTTCG	TGGACGCCCT	TTGCACGCGC	1680
ACTATATCTC	TATGCAGCAG	CTGAAAGCAG	CTTTGGTTTT	GATCGGAGGT	AGCGGGCGGA	1740
AAGGTGCAGA	ATGTCTAAAT	AATAAAGGAT	TCTTGTTGAAG	CTTTAGTTGT	CCGTAAACGA	1800
AAATAAAAAAT	AAAGAGGAAT	GATATGAAAG	CAAGTAGATC	AGTCTGCACT	TTCAAAATAG	1860
CTACCCTGGC	AGGCGCCATT	TATGCAGCGC	TGCCAATGTC	AGCTGCAAAC	TCGATGCAGC	1920
TGGATGTAGG	TAGCTCGGAT	TGGACGGTGC	GTTGGGGACA	ACACCCTCAA	GTATAGCCTT	1980
GCCTCTCGCC	TGAATGAGCA	AGACTCAAGT	CTGACAAATG	CGCCGACTGT	CAATGGTTAT	2040
ATCCGGATAT	TCAAAGTCAG	GGTGATCGTA	ACTTTGACCG	GGGGCTTGGT	ATCCAATCGT	2100
CTCGATATTC	TGGCTGCAG					2119

Sequenz 3

CTGCAGCCAG	GGCTGAAAAG	GAGGGATTCA	GTGAGGTCAT	GAAGGGAGGG	GACGGCGCCT	60
GGCTCCAATT	GCTCGATGGC	GCCGCGATTG	AGTGTCTTGG	GCGCGGTCTT	GGAGAGTTGG	120
GCTAGGGAGA	TAAATTTGCT	GGCCATGGTG	GCGGCCCCCTG	ATGGGTTGGA	TGATTTTCTG	180
CATTCTGCAT	CATGAAATTC	ATGAAATCAT	CACTTTTCGG	GGGGTGGGTG	CACGGGATTG	240
AAGGTTGCTA	GGAGAGTGCA	TTGCTCGTAA	GCCCAGGAAG	CACGCGGGTT	TCAGGATGGT	300
GCATGGAAAT	GGCATGAGCT	TTGCTGGATA	TGATTAGAGA	CATTAACAT	TTTGGCGGAA	360
TGGAAGCACG	ATTCCTCGCC	CGGTAGAGCG	GTAACCGCGA	CATTCAGGAC	CGTAAAAAGG	420
AAAGAGCATG	CAACTGACCA	ACAAGAAAAT	CGTCGTCACC	GGAGTGTCTT	CCGGTATCGG	480
TGCCGAAACT	GCCCGCGTTC	TGCGCTCTCA	CGGCGCCACA	GTGATTGGCG	TAGATCGCAA	540
CATGCCGAGC	CTGACTCTGG	ATGCTTTCGT	TCAGGCTGAC	CTGAGCCATC	CTGAAGGCAT	600
CGATCAACGG	CATAAATATT	CCAGTGGACG	GAGGTTTGGC	ATCGACCTAC	GTGTAAGTTC	660
GTGGACGCCC	TTTGCACGCG	CACTATATCT	CTATGCAGCA	GCTGAAAGCA	GCTTTGGTTT	720
TGATCGGAGG	TAGCGGGCGG	AAAGGTGCAG	AATGTCTAAA	TAATAAAGGA	TTCTTGTGAA	780
GCTTTAGTTG	TCCGTAAACG	AAAATAAAAA	TAAAGAGGAA	TGATATGAAA	GCAAGTAGAT	840
CAGTCTGCAC	TTTCAAATA	GCTACCCTGG	CAGGCGCCAT	TTATGCAGCG	CTGCCAATGT	900
CAGCTGCAAA	CTCGATGCAG	CTGGATGTAG	GTAGCTCGGA	TTGGACGGTG	CGTTGGGGAC	960
AACACCTCA	AGTATAGCCT	TGCCTCTCGC	CTGAATGAGC	AAGACTCAAG	TCTGACAAAT	1020
GCGCCGACTG	TCAATGGTTA	TATCCGGATA	TTCAAAGTCA	GGGTGATCGT	AACTTTGACC	1080
GGGGGCTTGG	TATCCAATCG	TCTCGATATT	CTGGCTGCAG			1120

Sequenz 4

GAATTCCGCG	TATCGCCCCG	TTCTATCAGC	GGGCCGCTTT	CGAAAGTCAT	GGTGTTAGCC	60
GGTAGGGTCT	TTTTCTTGCC	CATGCTTGTT	GCCTGAACCT	TCGTTGACAT	AGGGCAGAGG	120
TGCGTTTGCC	GCTTCGCTTC	GCGATGAACC	GCATCGAGAT	GCTGAGGTCA	GGATTTTTC	180
TTAACTCGCG	TAAGCATTC	GTCATTTTT	TGGTGGCTTT	GAACAGCCTG	ATGAAAGGTG	240
GTCTCGCCCT	TTGAGGCCGA	TTCTTGGGCG	CTTGGCGGCG	TCGAAGCGAT	GCTCCACTAC	300
CGATTAAGAT	AATTAATAA	AGGAAACCGC	ATGGTTTCTT	ATGTGAATTT	GTCTGGCATA	360
CTCCAGCTCA	AGGGCAATTT	TTGGGCTATT	GGCTGAGCAG	TTGCCTCTAT	ATGGTTATTC	420
AGAATAACAA	TTGACTCCTC	AGGAGGTCAG	CGATGAGCAT	TCTTGGTTTG	AATGGTGCCC	480
CGGTGCGAGC	TGAGCAGCTG	GGCTCGGCTC	TTGATCGCAT	GAAGAAGGCG	CACCTGGAGC	540
AGGGGCCTGC	AAACTTGGAG	CTGCGTCTGA	GTAGGCTGGA	TCGTGCGATT	GCAATGCTTC	600
TGGAATAATCG	TGAAGCAATT	GCCGACGCGG	TTTCTGCTGA	CTTTGGCAAT	CGCAGCCGTG	660
AGCAAACACT	GCTTTGCGAC	ATTGCTGGCT	CGGTGGCAAG	CCTGAAGGAT	AGCCGCGAGC	720
ACGTGGCCAA	ATGGATGGAG	CCCGAACATC	ACAAGGCGAT	GTTTCCAGGG	GCGGAGGCAC	780
GCGTTGAGTT	TCAGCCGCTG	GGTGTCGTTG	GGGTCATTAG	TCCCTGGAAC	TTCCCTATCG	840
TACTGGCCTT	TGGGCCGCTG	GCCGGCATAT	TCGCAGCAGG	TAATCGCGCC	ATGCTCAAGC	900
CGTCCGAGCT	TACCCCGCGG	ACTTCTGCCC	TGCTTGCGGA	GCTAATTGCT	CGTTACTTCG	960
ATGAAACTGA	GCTGACTACA	GTGCTGGGCG	ACGCTGAAGT	CGGTGCGCTG	TTCAGTGCTC	1020
AGCCTTTTGA	TCATCTGATC	TTACCCGGCG	GCACTGCCGT	GGCCAAGCAC	ATCATGCGTG	1080
CCGCGGCGGA	TAACCTAGTG	CCCGTTACCC	TGGAATTGGG	TGGCAAATCG	CCGGTGATCG	1140
TTTCCCGCAG	TGCAGATATG	GCGGACGTTG	CACAACGGGT	GTTGACGGTG	AAAACCTTCA	1200
ATGCCGGGCA	AATCTGTCTG	GCACCGGACT	ATGTGCTGCT	GCCGGAAGGG	ACAGCAAGCG	1260
AACCGGAATT	GCCAGCTGGG	GCGCCCTCTG	GTAAGGTTGG	GAAGCCCTGC	AAAGTAAACT	1320
GGATGGCTTT	CTTGCCGCCA	AGGATCTGAT	GGCGCAGGGG	ATCAAGATCT	GATCAAGAGA	1380
CAGGATGAGG	ATCGTTTCGC	ATGATTGAAC	AAGATGGATT	GCACGCAGGT	TCTCCGGCCG	1440
CTTGGGTGGA	GAGGCTATTC	GGCTATGACT	GGGCACAACA	GACAATCGGC	TGCTCTGATG	1500
CCGCCGTGTT	CCGGCTGTCA	GCGCAGGGGC	GCCCGGTTCT	TTTTGTCAAG	ACCGACCTGT	1560
CCGGTGCCCT	GAATGAACTG	CAGGACGAGG	CAGCGCGGCT	ATCGTGGCTG	GCCACGACGG	1620
GCGTTCCTTG	CGCAGCTGTG	CTCGACGTTG	TCACTGAAGC	GGGAAGGGAC	TGGCTGCTAT	1680
TGGGCGAAGT	GCCGGGGCAG	GATCTCCTGT	CATCTCACCT	TGCTCCTGCC	GAGAAAGTAT	1740
CCATCATGGC	TGATGCAATG	CGGCGGCTGC	ATACGCTTGA	TCCGGCTACC	TGCCCATTCG	1800
ACCACCAAGC	GAAACATCGC	ATCGAGCGAG	CACGTACTCG	GATGGAAGCC	GGTCTTGTCG	1860
ATCAGGATGA	TCTGGACGAA	GAGCATCAGG	GGCTCGCGCC	AGCCGAACTG	TTCCGCCAGG	1920
TCAAGGCGCG	CATGCCCGAC	GGCGAGGATC	TCGTCTGAC	CCATGGCGAT	GCCTGCTTGC	1980
CGAATATCAT	GGTGGAAAAT	GGCCGCTTTT	CTGGATTTCAT	CGACTGTGGC	CGGCTGGGTG	2040
TGGCGGACCG	CTATCAGGAC	ATAGCGTTGG	CTACCCGTGA	TATTGCTGAA	GAGCTTGGCG	2100
GCGAATGGGC	TGACCGCTTC	CTCGTGCTTT	ACGGTATCGC	CGCTCCCGAT	TCGCAGCGCA	2160

TCGCCTTCTA	TCGCCTTCTT	GACGAGTTCT	TCTGAGCGGG	ACTCTGGGGT	TCGAAATGAC	2220
CGACCAAGCG	ACGCCCCGCA	TGCCAAGCCT	GTTCTCGTGC	AAAGTCCTGT	GGGTGAGTCG	2280
AACTTGGCGA	TGCGCGCACC	CTACGGAGAA	GCGATCCACG	GACTGCTCTC	TGTCCTCCTT	2340
TCAACGGAGT	GTTAGAACCG	TTGGTAGTGG	TTTTGGACGG	GCCCAGGAGC	ATGCGCTTCT	2400
GGGCCCCGTT	CTTGAGTATT	CATTGGATAG	TCACGCGTGG	TAGCTTCGAG	CCTGCACAGC	2460
TGATGAGCAC	CCTGGAAGGC	GCGCTGTACG	CGGACGACTG	GGTTCATCTT	CGCCATTTCAT	2520
GACGGAATC	CGTTCCCCAG	TACCGCGATG	ACTATTTTGC	CTCTTCCGAT	GTCCGATTCC	2580
ACGCCGCCTG	ACGCTAAGCG	GGGGCGGGGG	CGCCCGCATC	CCAGCCCAGA	CAGCAACAAA	2640
TGAGTAGGCT	CTTGGAAGCC	GCGGCGGCTG	AGATTGGTAA	CGGCAATTTT	GTCAATGTGA	2700
CGATGGATTC	GATTGCCCCG	GCTGCCGGCG	TCTCAAAAAA	AACGCTGTAC	GTCTTGGTGG	2760
CGAGCAAGGA	AGAACTCATT	TCCCGGTTAG	TGGCTCGAGA	CATGTCCAAC	CTTGAGGAAT	2820
TC						2822

Sequenz 5

GAATTCGCG	TATCGCCCG	TTCTATCAGC	GGGCCGCTTT	CGAAAGTCAT	GGTGTTAGCC	60
GGTAGGGTCT	TTTTCTTGGC	CATGCTTGTT	GCCTGAACCT	TCGTTGACAT	AGGGCAGAGG	120
TGCGTTTGCC	GCTTCGCTTC	GCGATGAACC	GCATCGAGAT	GCTGAGGTCA	GGATTTTTCC	180
TTAACTCGCG	TAAGCATTC	GTCATTTTTT	TGGTGGCTTT	GAACAGCCTG	ATGAAAGGTG	240
GTCTCGCCCT	TTGAGGCCGA	TTCTTGGGCG	CTTGGCGGCG	TCGAAGCGAT	GCTCCACTAC	300
CGATTAAGAT	AATTAATAA	AGGAAACCGC	ATGGTTTCTT	ATGTGAATTT	GTCTGGCATA	360
CTCCAGCTCA	AGGGCAATTT	TTGGGCTATT	GGCTGAGCAG	TTGCCTCTAT	ATGGTTATTC	420
AGAATAACAA	TTGACTCCTC	AGGAGGTGAG	CGATGAGCAT	TCTTGGTTTG	AATGGTGCCC	480
CGGTCGGAGC	TGAGCAGCTG	GGCTCGGCTC	TTGATCGCAT	GAAGAAGGCG	CACCTGGAGC	540
AGGGGCCTGC	AAACTTGGAG	CTGCGTCTGA	GTAGGCTGGA	TCGTGCGATT	GCAATGCTTC	600
TGGAAAATCG	TGAAGCAATT	GCCGACGCGG	TTTCTGCTGA	CTTTGGCAAT	CGCAGCCGTG	660
AGCAAACACT	GCTTTGCGAC	ATTGCTGGCT	CGGTGGCAAG	CCTGAAGGAT	AGCCGCGAGC	720
ACGTGGCCAA	ATGGATGGAG	CCCGAACATC	ACAAGGCGAT	GTTTCCAGGG	GCGGAGGCAC	780
GCGTTGAGTT	TCAGCCGCTG	GGTGTGCTTG	GGGTGATTAG	TCCCTGGAAC	TTCCCTATCG	840
TACTGGCCTT	TGGGCCGCTG	GCCGGCATAT	TCGCAGCAGG	TAATCGCGCC	ATGCTCAAGC	900
CGTCCGAGCT	TACCCCGCGG	ACTTCTGCCC	TGCTTGCGGA	GCTAATTGCT	CGTTACTTCG	960
ATGAAACTGA	GCTGACTACA	GTGCTGGGCG	ACGCTGAAGT	CGGTGCGCTG	TTCAGTGCTC	1020
AGCCTTTTGA	TCATCTGATC	TTCAACGGCG	GCACTGCCGT	GGCCAAGCAC	ATCATGCGTG	1080
CCGCGGCGGA	TAACCTAGTG	CCCGTTACCC	TGGAATTGGG	TGGCAAATCG	CCGGTGATCG	1140
TTTCCCGCAG	TGCAGATATG	GCGGACGTTG	CACAACGGGT	GTTGACGGTG	AAAACCTTCA	1200
ATGCCGGGCA	AATCTGTCTG	GCACCGGACT	ATGTGCTGGG	GGAGAGGCGG	TTTGCCTATT	1260
GGGCGCATGC	ATAAAAACTG	TTGTAATTCA	TTAAGCATTC	TGCCGACATG	GAAGCCATCA	1320
CAAACGGCAT	GATGAACCTG	AATCGCCAGC	GGCATCAGCA	CCTTGTCGCC	TTGCGTATAA	1380
TATTTGCCCA	TGGACGCACA	CCGTGGAAAC	GGATGAAGGC	ACGAACCCAG	TTGACATAAG	1440
CCTGTTCCGT	TCGTAAACTG	TAATGCAAGT	AGCGTATGCG	CTCACGCAAC	TGGTCCAGAA	1500
CCTTGACCGA	ACGCAGCGGT	GGTAACGGCG	CAGTGCCGGT	TTTCATGGCT	TGTTATGACT	1560
GTTTTTTTGT	ACAGTCTATG	CCTCGGGCAT	CCAAGCAGCA	AGCGCGTTAC	GCCGTGGGTC	1620
GATGTTTGAT	GTTATGGAGC	AGCAACGATG	TTACGCAGCA	GCAACGATGT	TACGCAGCAG	1680
GGCAGTCGCC	CTAAACAAA	GTTAGGTGGC	TCAAGTATGG	GCATCATTCG	CACATGTAGG	1740
CTCGGCCCTG	ACCAAGTCAA	ATCCATGCGG	GCTGCTCTTG	ATCTTTTCGG	TCGTGAGTTC	1800
GGAGACGTAG	CCACCTACTC	CCAACATCAG	CCGGACTCCG	ATTACCTCGG	GAACCTTGCTC	1860
CGTAGTAAGA	CATTCATCGC	GCTTGCTGCC	TTCGACCAAG	AAGCGGTTGT	TGGCGCTCTC	1920
GCGGCTTACG	TTCTGCCAG	GTTTGAGCAG	CCGCGTAGTG	AGATCTATAT	CTATGATCTC	1980
GCAGTCTCCG	GCGAGCACCG	GAGGCAGGGC	ATTGCCACCG	CGCTCATCAA	TCTCCTCAAG	2040
CATGAGGCCA	ACGCGCTTGG	TGCTTATGTG	ATCTACGTGC	AAGCAGATTA	CGGTGACGAT	2100
CCCGCAGTGG	CTCTCTATAC	AAAGTTGGGC	ATACGGGAAG	AAGTGATGCA	CTTTGATATC	2160

GACCCAAGTA	CCGCCACCTA	ACAATTCGTT	CAAGCCGAGA	TCGGCTTCCC	TGCAAAGTCC	2220
TGTGGGTGAG	TCGAACTTGG	CGATGCGCGC	ACCCTACGGA	GAAGCGATCC	ACGGACTGCT	2280
CTCTGTCCTC	CTTCAACGG	AGTGTTAGAA	CCGTTGGTAG	TGGTTTTGGA	CGGGCCCAGG	2340
AGCATGCGCT	TCTGGGCCCCG	TTTCTTGAGT	ATTCATTGGA	TAGTCACGCG	TGGTAGCTTC	2400
GAGCCTGCAC	AGCTGATGAG	CACCCTGGAA	GGCGCGCTGT	ACGCGGACGA	CTGGGTTTCAT	2460
CTTCGCCATT	CATGACGGAA	CTCCGTTCCC	CAGTACCGCG	ATGACTATTT	TGCCTCTTCC	2520
GATGTCCGAT	TCCACGCCGC	CTGACGCTAA	GCGGGGGCGG	GGGCGCCCCG	ATCCCAGCCC	2580
AGACAGCAAC	AAATGAGTAG	GCTCTTGAT	GCCGCGGCGG	CTGAGATTGG	TAACGGCAAT	2640
TTCGTCAATG	TGACGATGGA	TTCGATTGCC	CGTGCTGCCG	GCGTCTCAA	AAAAACGCTG	2700
TACGTCTTGG	TGGCGAGCAA	GGAAGAACTC	ATTCCCCGGT	TAGTGGCTCG	AGACATGTCC	2760
AACCTTGAGG	AATTC					2775

Sequenz 6

GAATTCCGCG	TATCGCCCGG	TTCTATCAGC	GGGCCGCTTT	CGAAAGTCAT	GGTGTTAGCC	60
GGTAGGGTCT	TTTTCTTGCC	CATGCTTGTT	GCCTGAACCT	TCGTTGACAT	AGGGCAGAGG	120
TGCGTTTGCC	GCTTCGCTTC	GCGATGAACC	GCATCGAGAT	GCTGAGGTCA	GGATTTTTC	180
TTAACTCGCG	TAAGCATTCT	GTCATTTTTT	TGGTGGCTTT	GAACAGCCTG	ATGAAAGGTG	240
GTCTCGCCCT	TTGAGGCCGA	TTCTTGGGCG	CTTGGCGGCG	TCGAAGCGAT	GCTCCACTAC	300
CGATTAAGAT	AATTAAAATA	AGGAAACCGC	ATGGTTTCTT	ATGTGAATTT	GTCTGGCATA	360
CTCCAGCTCA	AGGGCAATTT	TTGGGCTATT	GGCTGAGCAG	TTGCCTCTAT	ATGGTTATTC	420
AGAATAACAA	TTGACTCCTC	AGGAGGTCAG	CGATGAGCAT	TCTTGGTTTG	AATGGTGCCC	480
CGGTCGGAGC	TGAGCAGCTG	GGCTCGGCTC	TTGATCGCAT	GAAGAAGGCG	CACCTGGAGC	540
AGGGGCCTGC	AAACTTGGAG	CTGCGTCTGA	GTAGGCTGGA	TCGTGCGATT	GCAATGCTTC	600
TGGAAAATCG	TGAAGCAATT	GCCGACGCGG	TTTCTGCTGA	CTTTGGCAAT	CGCAGCCGTG	660
AGCAAACACT	GCTTTGCGAC	ATTGCTGGCT	CGGTGGCAAG	CCTGAAGGAT	AGCCGCGAGC	720
ACGTGGCCAA	ATGGATGGAG	CCCGAACATC	ACAAGGCGAT	GTTTCCAGGG	GCGGAGGCAC	780
GCGTTGAGTT	TCAGCCGCTG	GGTGTGCTTG	GGGTCATTAG	TCCCTGGAAC	TTCCCTATCG	840
TACTGGCCTT	TGGGCCGCTG	GCCGGCATAT	TCGCAGCAGG	TAATCGCGCC	ATGCTCAAGC	900
CGTCCGAGCT	TACCCGCGCG	ACTTCTGCCC	TGCTTGCGGA	GCTAATTGCT	CGTTACTTCG	960
ATGAAACTGA	GCTGACTACA	GTGCTGGGCG	ACGCTGAAGT	CGGTGCGCTG	TTCAGTGCTC	1020
AGCCTTTCGA	TCATCTGATC	TTCACCGGCG	GCACTGCCGT	GGCCAAGCAC	ATCATGCGTG	1080
CCGCGGCGGA	TAACCTAGTG	CCCGTTACCC	TGGAATTGGG	TGGCAAATCG	CCGGTGATCG	1140
TTTCCCGCAG	TGCAGATATG	GCGGACGTTG	CACAACGGGT	GTTGACGGTG	AAAACCTTCA	1200
ATGCCGGGCA	AATCTGTCTG	GCACCGTGCG	TGAGTCGAAC	TTGGCGATGC	GCGCACCCCTA	1260
CGGAGAAGCG	ATCCACGGAC	TGCTCTCTGT	CCTCCTTTCA	ACGGAGTGTT	AGAACCGTTG	1320
GTAGTGTTT	TGGACGGGCC	CAGGAGCATG	CGCTTCTGGG	CCCGTTTCTT	GAGTATTCAT	1380
TGGATAGTCA	CGCGTGGTAG	CTTCGAGCCT	GCACAGCTGA	TGAGCACCCCT	GGAAGGCGCG	1440
CTGTACGCGG	ACGACTGGGT	TCATCTTCGC	CATTTCATGAC	GGAACCTCCGT	TCCCCAGTAC	1500
CGCGATGACT	ATTTTGCCCTC	TTCCGATGTC	CGATTCCACG	CCGCCTGACG	CTAAGCGGGG	1560
GCGGGGGCGC	CCGCATCCCA	GCCCAGACAG	CAACAAATGA	GTAGGCTCTT	GGATGCCGCG	1620
GCGGCTGAGA	TTGGTAACGG	CAATTTTCGTC	AATGTGACGA	TGGATTCGAT	TGCCCCGTGCT	1680
GCCGGCGTCT	CAAAAAAAC	GCTGTACGTC	TTGGTGGCGA	GCAAGGAAGA	ACTCATTTCC	1740
CGGTTAGTGG	CTCGAGACAT	GTCCAACCTT	GAGGAATTC			1779

Sequenz 7

CTGCAGCCGA	GCATCGATTG	AGCACTTTAC	CCAGCTGCGC	TGGCTGACCA	TTCAGAATGG	60
CCCGCGGCAC	TATCCAATCT	AAATCGATCT	TCGGGCGCCG	CGGGCATCAT	GCCCGCGGCG	120
CTCGCCTCAT	TTCAATCTCT	AACTTGATAA	AAACAGAGCT	GTTCTCCGGT	CTTGGTGGAT	180
CAAGGCCAGT	CGCGGAGAGT	CTCGAAGAGG	AGAGTACAGT	GAACGCCGAG	TCCACATTGC	240
AACCGCAGGC	ATCATCATGC	TCTGCTCAGC	CACGCTACCG	CAGTGTGTCG	ATTGGTCATC	300
CTCCGGTTGA	GGTTACGCAA	GACGCTGGAG	GTATTGTCCG	GATGCGTTCT	CTCGAGGCCG	360
TTCTTCCCTT	CCCGGGTCGA	ATTCTTGAGC	GTCTCGAGCA	TTGGGCTAAG	ACCCGTCCAG	420
AACAAACCTG	CGTTGCTGCC	AGGGCGGCAA	ATGGGGAATG	GCGTCGTATC	AGCTACGCGG	480
AAATGTTCCA	CAACGTCCGC	GCCATCGCAC	AGAGCTTGCT	TCCTTACGGA	CTATCGGCAG	540
AGCGTCCGCT	GCTTATCGTC	TCTGGAAATG	ACCTGGAACA	TCTTCAGCTG	GCAATTTGGGG	600
CTATGTATGC	GGGCATTCCC	TATTGCCCGG	TGTCTCCTGC	TTATTCACTG	CTGTCGCAAG	660
ATTTGGCGAA	GTCGCGTCAC	ATCGTAGGTC	TTCTGCAACC	GGGACTGGTC	TTTGCTGCCG	720
ATGCAGCACC	TTTCCAGGGG	ACAGCAAGCG	AACCGGAATT	GCCAGCTGGG	GCGCCCTCTG	780
GTAAGGTTGG	GAAGCCCTGC	AAAGTAAACT	GGATGGCTTT	CTTGCCGCCA	AGGATCTGAT	840
GGCGCAGGGG	ATCAAGATCT	GATCAAGAGA	CAGGATGAGG	ATCGTTTCGC	ATGATTGAAC	900
AAGATGGATT	GCACGCAGGT	TCTCCGGCCG	CTTGGGTGGA	GAGGCTATTC	GGCTATGACT	960
GGGCACAACA	GACAATCGGC	TGCTCTGATG	CCGCCGTGTT	CCGGCTGTCA	GCGCAGGGGC	1020
GCCCGGTTCT	TTTTGTCAAG	ACCGACCTGT	CCGGTGCCCT	GAATGAACTG	CAGGACGAGG	1080
CAGCGCGGCT	ATCGTGGCTG	GCCACGACGG	GCGTTCCTTG	CGCAGCTGTG	CTCGACGTTG	1140
TCACTGAAGC	GGGAAGGGAC	TGGCTGCTAT	TGGGCGAAGT	GCCGGGGCAG	GATCTCCTGT	1200
CATCTCACCT	TGCTCCTGCC	GAGAAAGTAT	CCATCATGGC	TGATGCAATG	CGGCGGCTGC	1260
ATACGCTTGA	TCCGGCTACC	TGCCCCATTG	ACCACCAAGC	GAAACATCGC	ATCGAGCGAG	1320
CACGTACTCG	GATGGAAGCC	GGTCTTGTCG	ATCAGGATGA	TCTGGACGAA	GAGCATCAGG	1380
GGCTCGCGCC	AGCCGAAGTG	TTGCGCCAGG	TCAAGGCGCG	CATGCCCGAC	GGCGAGGATC	1440
TCGTGCTGAC	CCATGGCGAT	GCCTGCTTGC	CGAATATCAT	GGTGGAAAAT	GGCCGCTTTT	1500
CTGGATTGAT	CGACTGTGGC	CGGCTGGGTG	TGGCGGACCG	CTATCAGGAC	ATAGCGTTGG	1560
CTACCCGTGA	TATTGCTGAA	GAGCTTGCGG	GCGAATGGGC	TGACCGCTTC	CTCGTGCTTT	1620
ACGGTATCGC	CGCTCCCGAT	TCGCAGCGCA	TCGCCTTCTA	TCGCCTTCTT	GACGAGTTCT	1680
TCTGAGCGGG	ACTCTGGGGT	TCGAAATGAC	CGACCAAGCG	ACGCCCCCTGT	TTTGCAATGG	1740
CGGTGCGCGA	AAGTTGATGC	GCTGTATCGT	GGTGAAGATC	AATCCATGCT	GCGTGACGAG	1800
GCCACACTGT	GAGTTGGTCA	GGGGGGGCTT	ACTCGGCGTT	TTCCGACACT	GCGTTGGTTG	1860
CGGCAGTGCG	CACCCCTGGG	ATTGATTGCG	GGGGTGCCCT	GTCGCTGGTG	TCCGCTATCG	1920
ACTTAGGGGT	AAAGGTCGCT	CGCGAAGTTC	TGATGCGTGC	GTCGCTTGAA	CCACAAATGG	1980
TCGATAGCGT	ACTCGCAGGC	TCTATGGCTC	AAGCAAGCTT	TGATGCTTAC	CTGCTCCCGC	2040
GGCACATTGG	CTTGATACAGC	GGTGTTCCTA	AGTCGGTTCC	GGCCTTGGGG	GTGCAGCGCA	2100
TTTGCGGCAC	AGGCTTCGAA	CTGCTTCGGC	AGGCCGGCGA	GCAGATTTC	CAAGGCGCTG	2160
ATCACGTGCT	GTGTGTCGCG	GGCTGCAG				2188

Sequenz 8

CTGCAGCCGA	GCATCGATTG	AGCACTTTAC	CCAGCTGCGC	TGGCTGACCA	TTCAGAATGG	60
CCCGCGGCAC	TATCCAATCT	AAATCGATCT	TCGGGCGCCG	CGGGCATCAT	GCCCCGCGCG	120
CTCGCCTCAT	TTCAATCTCT	AACTTGATAA	AAACAGAGCT	GTTCTCCGGT	CTTGGTGGAT	180
CAAGGCCAGT	CGCGGAGAGT	CTCGAAGAGG	AGAGTACAGT	GAACGCCGAG	TCCACATTGC	240
AACCGCAGGC	ATCATCATGC	TCTGCTCAGC	CACGCTACCG	CAGTGTGTCT	ATTGGTTCATC	300
CTCCGGTTGA	GGTTACGCAA	GACGCTGGAG	GTATTGTCCG	GATGCGTTCT	CTCGAGGCGC	360
TTCTTCCCTT	CCCGGGTCGA	ATTCTTGAGC	GTCTCGAGCA	TTGGGCTAAG	ACCCGTCCAG	420
AACAAACCTG	CGTTGCTGCC	AGGGCGGCAA	ATGGGGAATG	GCGTCGTATC	AGCTACGCGG	480
AAATGTTCCA	CAACGTCCGC	GCCATCGCAC	AGAGCTTGCT	TCCTTACGGA	CTATCGGCAG	540
AGCGTCCGCT	GCTTATCGTC	TCTGGAAATG	ACCTGGAACA	TCTTCAGCTG	GCATTGTTGGG	600
CTATGTATGC	GGGCATTCCC	TATTGCCCGG	TGTCTCCTGC	TTATTCACTG	CTGTGCGAAG	660
ATTTGGCGAA	GCTGCGTCAC	ATCGTAGGTC	TTCTGCAACC	GGGACTGGTC	TTTGCTGCCG	720
ATGCAGCACC	TTTCCAGGGG	GAGAGGCGGT	TTGCGTATTG	GGCGCATGCA	TAAAAACTGT	780
TGTAATTCAT	TAAGCATTCT	GCCGACATGG	AAGCCATCAC	AAACGGCATG	ATGAACCTGA	840
ATCGCCAGCG	GCATCAGCAC	CTTGTCGCTT	TGCGTATAAT	ATTTGCCCAT	GGACGCACAC	900
CGTGGAACG	GATGAAGGCA	CGAACCCAGT	TGACATAAGC	CTGTTCCGGT	CGTAAACTGT	960
AATGCAAGTA	GCGTATGCGC	TCACGCAACT	GGTCCAGAAC	CTTGACCGAA	CGCAGCGGTG	1020
GTAACGGCGC	AGTGGCGGTT	TTCATGGCTT	GTTATGACTG	TTTTTTTGTA	CAGTCTATGC	1080
CTCGGGCATC	CAAGCAGCAA	GCGCGTTACG	CCGTGGGTCTG	ATGTTTGATG	TTATGGAGCA	1140
GCAACGATGT	TACGCAGCAG	CAACGATGTT	ACGCAGCAGG	GCAGTCGCCC	TAAAACAAAG	1200
TTAGGTGGCT	CAAGTATGGG	CATCATTCGC	ACATGTAGGC	TCGGCCCTGA	CCAAGTCAAA	1260
TCCATGCGGG	CTGCTCTTGA	TCTTTTCGGT	CGTGAGTTCG	GAGACGTAGC	CACCTACTCC	1320
CAACATCAGC	CGGACTCCGA	TTACCTCGGG	AACTTGCTCC	GTAGTAAGAC	ATTTCATCGCG	1380
CTTGCTGCCT	TCGACCAAGA	AGCGGTTGTT	GGCGCTCTCG	CGGCTTACGT	TCTGCCCAGG	1440
TTTGAGCAGC	CGCGTAGTGA	GATCTATATC	TATGATCTCG	CAGTCTCCGG	CGAGCACCGG	1500
AGGCAGGGCA	TTGCCACCGC	GCTCATCAAT	CTCCTCAAGC	ATGAGGCCAA	CGCGCTTGGT	1560
GCTTATGTGA	TCTACGTGCA	AGCAGATTAC	GGTGACGATC	CCGCAGTGGC	TCTCTATACA	1620
AAGTTGGGCA	TACGGGAAGA	AGTGATGCAC	TTTGATATCG	ACCCAAGTAC	CGCCACCTAA	1680
CAATTCGTTT	AAGCCGAGAT	CGGCTTCCCC	TGTTTTGCAA	TGGCGGTCTGG	CGAAAGTTGA	1740
TGCGCTGTAT	CGTGGTGAAG	ATCAATCCAT	GCTGCGTGAC	GAGGCCACAC	TGTGAGTTGG	1800
TCAGGGGGGG	CTTACTCGGC	GTTTTCCGAC	ACTGCGTTGG	TTGCGGCAGT	GCGCACCCCC	1860
TGGATTGATT	GCGGGGGTGC	CCTGTCGCTG	GTGTCGCCTA	TCGACTTAGG	GGTAAAGGTC	1920
GCTCGCGAAG	TTCTGATGCG	TGCGTCGCTT	GAACCACAAA	TGGTCGATAG	CGTACTCGCA	1980
GGCTCTATGG	CTCAAGCAAG	CTTTGATGCT	TACCTGCTCC	CGCGGCACAT	TGGCTTGATC	2040
AGCGGTGTTT	CCAAGTCGGT	TCCGGCCTTG	GGGGTGACG	GCATTTGCGG	CACAGGCTTC	2100
GAAGTGTTC	GGCAGGCCGG	CGAGCAGATT	TCCCAAGGCG	CTGATCACGT	GCTGTGTGTC	2160
GCGGGCTGCA	G					2171

Sequenz 9

CTGCAGCCGA	GCATCGATTG	AGCACTTTAC	CCAGCTGCGC	TGGCTGACCA	TTCAGAATGG	60
CCCCGCGCAC	TATCCAATCT	AAATCGATCT	TCGGGCGCCG	CGGGCATCAT	GCCCCGCGCG	120
CTCGCCTCAT	TTCAATCTCT	AACTTGATAA	AAACAGAGCT	GTTCTCCGGT	CTTGGTGGAT	180
CAAGGCCAGT	CGCGGAGAGT	CTCGAAGAGG	AGAGTACAGT	GAACGCCGAG	TCCACATTGC	240
AACCGCAGGC	ATCATCATGC	TCTGCTCAGC	CACGCTACCG	CAGTGTGTCG	ATTGGTCATC	300
CTCCGGTTGA	GGTTACGCAA	GACGCTGGAG	GTATTGTCCG	GATGCGTTCT	CTCGAGGCGC	360
TTCTTCCCTT	CCCCGGGTCGA	ATTCTTGAGC	GTCTCGAGCA	TTGGGCTAAG	ACCCGTCCAG	420
AACAAACCTG	CGTTGCTGCC	AGGGCGGCAA	ATGGGGAATG	GCGTCGTATC	AGCTACGCGG	480
AAATGTTCCA	CAACGTCCGC	GCCATCGCAC	AGAGCTTGCT	TCCTTACGGA	CTATCGGCAG	540
AGCGTCCGCT	GCTTATCGTC	TCTGGAATG	ACCTGGAACA	TCTTCAGCTG	GCATTTGGGG	600
CTATGTATGC	GGGCATTCCC	TATTGCCCCG	TGTCTCCTGC	TTATTCACTG	CTGTGCAAG	660
ATTTGGCGAA	GCTGCGTCAC	ATCGTAGGTC	TTCTGCAACC	GGGACTGGTC	TTTGCTGCCG	720
ATGCAGCACC	TTTCCAGCGC	GCTGTTTTGC	AATGGCGGTC	GGCGAAAGTT	GATGCGCTGT	780
ATCGTGGTGA	AGATCAATCC	ATGCTGCGTG	ACGAGGCCAC	ACTGTGAGTT	GGTCAGGGGG	840
GGCTTACTCG	GCGTTTTCCG	ACACTGCGTT	GGTTGCGGCA	GTGCGCACCC	CCTGGATTGA	900
TTGCGGGGGT	GCCCTGTGCG	TGGTGTGCGC	TATCGACTTA	GGGGTAAAGG	TCGCTCGCGA	960
AGTTCTGATG	CGTGCGTCGC	TTGAACCACA	AATGGTCGAT	AGCGTACTCG	CAGGCTCTAT	1020
GGCTCAAGCA	AGCTTTGATG	CTTACCTGCT	CCCCGCGCAC	ATTGGCTTGT	ACAGCGGTGT	1080
TCCCAAGTCG	GTTCCGGCCT	TGGGGGTGCA	GCGCATTTGC	GGCACAGGCT	TCGAACTGCT	1140
TCGGCAGGCC	GGCGAGCAGA	TTTCCCAAGG	CGCTGATCAC	GTGCTGTGTG	TCGCGGGCTG	1200
CAG						1203

Sequenz 10

GAATCCCCCT	GGCGACGAAA	GGGCGGCAGG	CCGCATGGCC	ACGGCTGGGC	GGTAACTGAT	60
GCTTGCGTTA	ATCGTTAACC	GTTTGAAATT	CCTTGCCAAA	TTTCGGCGAG	AGAATCATGC	120
GGGTACGCCT	TTCCGTGCGC	TTTGATCTGC	GCTTCCGTGC	CTTGAATCAG	AAAAATAGTT	180
AATTGACAGA	ACTATAGGTT	CGCAGTAGCT	TTTGCTCACC	CACCAAATCC	ACAGCACTGG	240
GGTGACGAT	GAATAGCTAC	GATGGCCGTT	GGTCTACCGT	TGATGTGAAG	GTTGAAGAAG	300
GTATCGCTTG	GGTCACGCTG	AACCGCCCGG	AGAAGCGCAA	CGCAATGAGC	CCAACTCTCA	360
ATCGAGAGAT	GGTCGAGGTT	CTGGAGGTGC	TGGAGCAGGA	CGCAGATGCT	CGCGTGCTTG	420
TTCTGACTGG	TGCAGGCGAA	TCCTGGACCG	CGGGCATGGA	CCTGAAGGAG	TATTTCCGCG	480
AGACCGATGC	TGGCCCCGAA	ATTCTGCAAG	AGAAGATTCTG	TCGGGGACAG	CAAGCGAACC	540
GGAATTGCCA	GCTGGGGCGC	CCTCTGGTAA	GGTTGGGAAG	CCCTGCAAAG	TAAACTGGAT	600
GGCTTTCTTG	CCGCCAAGGA	TCTGATGGCG	CAGGGGATCA	AGATCTGATC	AAGAGACAGG	660
ATGAGGATCG	TTTCGCATGA	TTGAACAAGA	TGGATTGCAC	GCAGGTTCTC	CGGCCGCTTG	720
GGTGGAGAGG	CTATTCCGCT	ATGACTGGGC	ACAACAGACA	ATCGGCTGCT	CTGATGCCCG	780
CGTGTTCCGG	CTGTCAAGCG	AGGGGCGCCC	GGTTCTTTTT	GTCAAGACCG	ACCTGTCCGG	840
TGCCCTGAAT	GAAGTGCAGG	ACGAGGCAGC	GCGGCTATCG	TGGCTGGCCA	CGACGGGCGT	900
TCCTTGCGCA	GCTGTGCTCG	ACGTTGTAC	TGAAGCGGGA	AGGGACTGGC	TGCTATTGGG	960
CGAAGTGCCG	GGGCAGGATC	TCCTGTATC	TCACCTTGCT	CCTGCCGAGA	AAGTATCCAT	1020
CATGGCTGAT	GCAATGCGGC	GGCTGCATAC	GCTTGATCCG	GCTACCTGCC	CATTCGACCA	1080
CCAAGCGAAA	CATCGCATCG	AGCGAGCAG	TACTCGGATG	GAAGCCGGTC	TTGTCGATCA	1140
GGATGATCTG	GACGAAGAGC	ATCAGGGGCT	CGCGCCAGCC	GAAGTGTTCG	CCAGGCTCAA	1200
GGCGCGCATG	CCCGACGGCG	AGGATCTCGT	CGTGACCCAT	GGCGATGCCT	GCTTGCCGAA	1260
TATCATGGTG	GAAAATGGCC	GCTTTTCTGG	ATTCATCGAC	TGTGGCCGGC	TGGGTGTGGC	1320
GGACCGCTAT	CAGGACATAG	CGTTGGCTAC	CCGTGATATT	GCTGAAGAGC	TTGGCGGCGA	1380
ATGGGCTGAC	CGCTTCCTCG	TGCTTTACGG	TATCGCCGCT	CCCGATTTCG	AGCGCATCGC	1440
CTTCTATCGC	CTTCTTGACG	AGTTCTTCTG	AGCGGGACTC	TGGGGTTCGA	AATGACCGAC	1500
CAAGCGACGC	CCCGAGCAGG	GCATGAAGCA	GTTCTTGTAC	GAGAAAAGCA	TCAAGCCGGG	1560
CTTGCAGACC	TACAAGCGCT	GATAAATGCG	CCGGGGCCCT	CGCTGCGCCC	CCGGCCTTCC	1620
AATAATGACA	ATAATGAGGA	GTGCCCAATG	TTTCACGTGC	CCCTGCTTAT	TGGTGGTAAG	1680
CCTTGTTTCAG	CATCTGATGA	GCGCACCTTC	GAGCGTCGTA	GCCCGCTGAC	CGGAGAAGTG	1740
GTATCGCGCG	TCGCTGCTGC	CAGTTTGGAA	GATGCGGACG	CCGCAGTGGC	CGCTGCACAG	1800
GCTGCGTTTC	CTGAATGGGC	GGCGCTTGCT	CCGAGCGAAC	GCCGTGCCCC	ACTGCTGCGA	1860
GCGGCGGATC	TTCTAGAGGA	CCGTTCTTCC	GAGTTACCCG	CCGCAGCGAG	TGAAACTGGC	1920
GCAGCGGGAA	ACTGGTATGG	GTTTAACGTT	TACCTGGCGG	CGGGCATGTT	GCGGGGAATT	1980
C						1981

Sequenz 11

GAATTCCCCCT	GGCGACGAAA	GGGCGGCAGG	CCGCATGGCC	ACGGCTGGGC	GGTAACTGAT	60
GCTTGCGTTA	ATCGTTAACC	GTTTGAAATT	CCTTGCCAAA	TTTCGGCGAG	AGAATCATGC	120
GGGTACGCCT	TCCGTTGCGC	TTTGATCTGC	GCTTCCGTGC	CTTGAATCAG	AAAAATAGTT	180
AATTGACAGA	ACTATAGGTT	CGCAGTAGCT	TTTGCTCACC	CACCAAATCC	ACAGCACTGG	240
GGTGACAGAT	GAATAGCTAC	GATGGCCGTT	GGTCTACCGT	TGATGTGAAG	GTTGAAGAAG	300
GTATCGCTTG	GGTCACGCTG	AACCGCCCGG	AGAAGCGCAA	CGCAATGAGC	CCAACTCTCA	360
ATCGAGAGAT	GGTCGAGGTT	CTGGAGGTGC	TGGAGCAGGA	CGCAGATGCT	CGCGTGCTTG	420
TTCTGACTGG	TGCAGGCGAA	TCCTGGACCG	CGGGCATGGA	CCTGAAGGAG	TATTTCCGCG	480
AGACCGATGC	TGGCCCCGAA	ATTCTGCAAG	AGAAGATTCTG	TCGGGGGAGA	GGCGGTTTGC	540
GTATTGGGCG	CATGCATAAA	AACTGTTGTA	ATTCAATTAAG	CATTCTGCCG	ACATGGAAGC	600
CATCACAAAC	GGCATGATGA	ACCTGAATCG	CCAGCGGCAT	CAGCACCTTG	TCGCCTTGCG	660
TATAATATTT	GCCCATGGAC	GCACACCGTG	GAAACGGATG	AAGGCACGAA	CCCAGTTGAC	720
ATAAGCCTGT	TCGGTTCGTA	AACTGTAATG	CAAGTAGCGT	ATGCGCTCAC	GCAACTGGTC	780
CAGAACCTTG	ACCGAACGCA	GCGGTGGTAA	CGGCGCAGTG	GCGGTTTTCA	TGGCTTGTTA	840
TGACTGTTTT	TTGTACAGT	CTATGCCTCG	GGCATCCAAG	CAGCAAGCGC	GTTACGCCGT	900
GGGTCGATGT	TTGATGTTAT	GGAGCAGCAA	CGATGTTACG	CAGCAGCAAC	GATGTTACGC	960
AGCAGGGCAG	TCGCCCTAAA	ACAAAGTTAG	GTGGCTCAAG	TATGGGCATC	ATTGCGACAT	1020
GTAGGCTCGG	CCCTGACCAA	GTCAAATCCA	TGCGGGCTGC	TCTTGATCTT	TTCGGTCGTG	1080
AGTTCGGAGA	CGTAGCCACC	TACTCCCAAC	ATCAGCCGGA	CTCCGATTAC	CTCGGGAAC	1140
TGCTCCGTAG	TAAGACATTC	ATCGCGCTTG	CTGCCTTCGA	CCAAGAAGCG	GTTGTTGGCG	1200
CTCTCGCGGC	TTACGTTCTG	CCCAGTTTG	AGCAGCCGCG	TAGTGAGATC	TATATCTATG	1260
ATCTCGCAGT	CTCCGGCGAG	CACCGGAGGC	AGGGCATTGC	CACCGCGCTC	ATCAATCTCC	1320
TCAAGCATGA	GGCCAACGCG	CTTGGTGCTT	ATGTGATCTA	CGTGCAAGCA	GATTACGGTG	1380
ACGATCCCGC	AGTGGCTCTC	TATACAAAGT	TGGGCATACG	GGAAGAAGTG	ATGCACTTTG	1440
ATATCGACCC	AAGTACCGCC	ACCTAACAAT	TCGTTCAAGC	CGAGATCGGC	TTCCCCGAGC	1500
AGGGCATGAA	GCAGTTCCTT	GACGAGAAAA	GCATCAAGCC	GGGCTTGCA	ACCTACAAGC	1560
GCTGATAAAT	GCGCCGGGGC	CCTCGCTGCG	CCCCCGGCCT	TCCAATAATG	ACAATAATGA	1620
GGAGTGCCCA	ATGTTTCACG	TGCCCCCTGCT	TATTGGTGGT	AAGCCTTGTT	CAGCATCTGA	1680
TGAGCGCACC	TTCGAGCGTC	GTAGCCCGCT	GACCGGAGAA	GTGGTATCGC	GCGTCGCTGC	1740
TGCCAGTTTG	GAAGATGCGG	ACGCCGCGAGT	GGCCGCTGCA	CAGGCTGCGT	TTCCTGAATG	1800
GGCGGCGCTT	GCTCCGAGCG	AACGCCGTGC	CCGACTGCTG	CGAGCGGCGG	ATCTTCTAGA	1860
GGACCGTTCT	TCCGAGTTCA	CCGCCGCAGC	GAGTGAAACT	GGCGCAGCGG	GAAACTGGTA	1920
TGGGTTTAAC	GTTTACCTGG	CGGCGGGCAT	GTTGCGGGGA	ATTC		1964

Sequenz 12

GAATTCCT	GGCGACGAAA	GGGCGGCAGG	CCGCATGGCC	ACGGCTGGGC	GGTAACTGAT	60
GCTTGCGTTA	ATCGTTAACC	GTTTGAAATT	CCTTGCCAAA	TTTCGGCGAG	AGAATCATGC	120
GGGTACGCCT	TTCCGTGCGC	TTTGATCTGC	GCTTCCGTGC	CTTGAATCAG	AAAAATAGTT	180
AATTGACAGA	ACTATAGGTT	CGCAGTAGCT	TTTGCTCACC	CACCAAATCC	ACAGCACTGG	240
GGTGACGAT	GAATAGCTAC	GATGGCCGTT	GGTCTACCGT	TGATGTGAAG	GTTGAAGAAG	300
GTATCGCTTG	GGTCACGCTG	AACCGCCCGG	AGAAGCGCAA	CGCAATGAGC	CCAACTCTCA	360
ATCGAGAGAT	GGTCGAGGTT	CTGGAGGTGC	TGGAGCAGGA	CGCAGATGCT	CGCGTGCTTG	420
TTCTGACTGG	TGCAGGCGAA	TCCTGGACCG	CGGGCATGGA	CCTGAAGGAG	TATTTCCGCG	480
AGACCGATGC	TGGCCCCGAA	ATTCTGCAAG	AGAAGATTCG	TCGCGAGCAG	GGCATGAAGC	540
AGTTCCTTGA	CGAGAAAAGC	ATCAAGCCGG	GCTTGCAGAC	CTACAAGCGC	TGATAAATGC	600
GCCGGGGCCC	TCGCTGCGCC	CCCGGCCCTC	CAATAATGAC	AATAATGAGG	AGTGCCCAAT	660
GTTTCACGTG	CCCCTGCTTA	TTGGTGGTAA	GCCTTGTTCA	GCATCTGATG	AGCGCACCTT	720
CGAGCGTCGT	AGCCCGCTGA	CCGGAGAAGT	GGTATCGCGC	GTCGCTGCTG	CCAGTTTGGG	780
AGATGCGGAC	GCCGCAGTGG	CCGCTGCACA	GGCTGCGTTT	CCTGAATGGG	CGGCGCTTGC	840
TCCGAGCGAA	CGCCGTGCCC	GA CTGCTGCG	AGCGGCGGAT	CTTCTAGAGG	ACCGTTCTTC	900
CGAGTTCACC	GCCGCAGCGA	GTGAACTGG	CGCAGCGGGA	AACTGGTATG	GGTTTAACGT	960
TTACCTGGCG	GCGGGCATGT	TGCGGGGAAT	TC			992

Sequenz 13

GAATTCCAAT	AATGACAATA	ATGAGGAGTG	CCCAATGTTT	CACGTGCCCC	TGCTTATTGG	60
TGGTAAGCCT	TGTTACAGCAT	CTGATGAGCG	CACCTTCGAG	CGTCGTAGCC	CGCTGACCGG	120
AGAAAGTGGTA	TCGCGCGTCG	CTGCTGCCAG	TTTGGAAGAT	GCGGACGCCG	CAGTGGCCCG	180
TGCACAGGCT	GCGTTTCCTG	AATGGGCGGC	GCTTGCTCCG	AGCGAACGCC	GTGCCCGACT	240
GCTGCGAGCG	GCGGATCTTC	TAGAGGACCG	TTCTTCCGAG	TTCACCGCCG	CAGCGAGTGA	300
AACTGGCGCA	GCGGGAAACT	GGTATGGGTT	TAACGTTTAC	CTGGCGGCGG	GCATGTTGCG	360
GGAAGCCGCG	GCCATGACCA	CACAGATTCA	GGGCGATGTC	ATTCCGTCCA	ATGTGCCCGG	420
TAGCTTTGCC	ATGGCGGTTT	GACAGCCATG	TGGCGTGTTG	CTCGGTATTG	CGCCTTGGAA	480
TGCTCCCGTA	ATCCTTGGCG	TACGGGCTGT	TGCGATGCCG	TTGGCATGCG	GCAATACCGT	540
GGTGTTGAAA	AGCTCTGAGC	TGAGTCCCTT	TACCCATCGC	CTGATTGGTC	AGGTGTTGCA	600
TGATGCTGGT	CTGGGGGATG	GCGTGGTGAA	TGTCATCAGC	AATGCCCCGC	AAGACGCTCC	660
TGCGGTGGTG	GAGCGACTGA	TTGCAAATCC	TGCGGTACGT	CGAGTGAAGT	TCACCGGTTT	720
GACCCACGTT	GGACGGATCA	TTGGTGAGCT	GTCTGCGCGT	CATCTGAAGC	CTGCTGTGCT	780
GGAATTAGGT	GGTAAGGCTC	CGTTCTTGGT	CTTGACGAT	GCCGACCTCG	ATGCGGCGGT	840
CGAAGCGGCG	GCCTTTGGTG	CCTACTTCAA	TCAGGTCATA	ATCTGCATGT	CCACTGAGCG	900
TCTGATTGTG	ACAGCAGTCG	CAGACGCCTT	TGTTGAAAAG	CTGGCGAGGA	AGGTCGCCAC	960
ACTGCGTGCT	GGCGATCCTA	ATGATCCGCA	ATCGGTCTTG	GGTTCGTTGA	TTGATGCCAA	1020
TGCAGGTCAT	CGCATCCAGG	TTCTGGTCTG	TGATGCGCTC	GGGGACAGCA	AGCGAACCGG	1080
AATTGCCAGC	TGGGGCGCCC	TCTGGTAAGG	TTGGGAAGCC	CTGCAAAGTA	AACTGGATGG	1140
CTTTCTTGCC	GCCAAGGATC	TGATGGCGCA	GGGGATCAAG	ATCTGATCAA	GAGACAGGAT	1200
GAGGATCGTT	TCGCATGATT	GAACAAGATG	GATTGCACGC	AGGTTCTCCG	GCCGCTTGGG	1260
TGGAGAGGCT	ATTGCGCTAT	GAATGGGAC	AACAGACAAT	CGGCTGCTCT	GATGCCGCCG	1320
TGTTCCGGCT	GTCAGCGCAG	GGGCGCCCGG	TTCTTTTGT	CAAGACCGAC	CTGTCCGGTG	1380
CCCTGAATGA	ACTGCAGGAC	GAGGCAGCGC	GGCTATCGTG	GCTGGCCACG	ACGGGCGGTC	1440
CTTGCGCAGC	TGTGCTCGAC	GTTGTCACTG	AAGCGGGAAG	GGACTGGCTG	CTATTGGGCG	1500
AAGTGCCGGG	GCAGGATCTC	CTGTCATCTC	ACCTTGCTCC	TGCCGAGAAA	GTATCCATCA	1560
TGGCTGATGC	AATGCGGCGG	CTGCATACGC	TTGATCCGGC	TACCTGCCCC	TTGACCAACC	1620
AAGCGAAACA	TCGCATCGAG	CGAGCACGTA	CTCGGATGGA	AGCCGGTCTT	GTCGATCAGG	1680
ATGATCTGGA	CGAAGAGCAT	CAGGGGCTCG	CGCCAGCCGA	ACTGTTCCGC	AGGCTCAAGG	1740
CGCGCATGCC	CGACGGCGAG	GATCTCGTCG	TGACCCATGG	CGATGCCTGC	TTGCCGAATA	1800
TCATGGTGGA	AAATGGCCGC	TTTTCTGGAT	TCATCGACTG	TGGCCGGCTG	GGTGTGGCGG	1860
ACCGCTATCA	GGACATAGCG	TTGGCTACCC	GTGATATTGC	TGAAGAGCTT	GGCGGCGAAT	1920
GGGCTGACCG	CTTCCTCGTG	CTTTACGGTA	TCGCCGCTCC	CGATTGCGAG	CGCATCGCCT	1980
TCTATCGCCT	TCTTGACGAG	TTCTTCTGAG	CGGGACTCTG	GGGTTGAAA	TGACCGACCA	2040
AGCGACGCC	GGCCAGCGC	GTCGATTCCG	GCATTTGCCA	TATCAATGGA	CCGACTGTGC	2100
ATGACGAGGC	TCAGATGCCA	TTGCGTGGGG	TGAAGTCCAG	CGGCTACGGC	AGCTTCGGCA	2160

GTCGAGCATC	GATTGAGCAC	TTTACCCAGC	TGCGCTGGCT	GACCATTTCAG	AATGGCCCCG	2220
GGCACTATCC	AATCTAAATC	GATCTTCGGG	CGCCGCGGGC	ATCATGCCCCG	CGGCGCTCGC	2280
CTCATTTCAA	TCTCTAACTT	GATAAAAACA	GAGCTGTTCT	CCGGTCTTGG	TGGATCAAGG	2340
CCAGTCGCGG	AGAGTCTCGA	AGAGGAGAGT	ACAGTGAACG	CCGAGTCCAC	ATTGCAACCG	2400
CAGGCATCAT	CATGCTCTGC	TCAGCCACGC	TACCGCAGTG	TGTCGATTGG	TCATCCTCCG	2460
GTTGAGGTTA	CGCAAGACGC	TGGAGGTATT	GTCCGGATGC	GTTCTCTCGA	GGCGCTTCTT	2520
CCCTTCCCGG	GTGGAATTC					2539

Sequenz 14

GAATTCCAAT	AATGACAATA	ATGAGGAGTG	CCCAATGTTT	CACGTGCCCC	TGCTTATTGG	60
TGGTAAGCCT	TGTTCAGCAT	CTGATGAGCG	CACCTTCGAG	CGTCGTAGCC	CGCTGACCGG	120
AGAAGTGGTA	TCGCGCGTCG	CTGCTGCCAG	TTTGGAAGAT	GCGGACGCCG	CAGTGGCCCG	180
TGCACAGGCT	GCGTTTCCTG	AATGGGCGGC	GCTTGCTCCG	AGCGAACGCC	GTGCCCAGCT	240
GCTGCGAGCG	GCGGATCTTC	TAGAGGACCG	TTCTTCCGAG	TTCACCGCCG	CAGCGAGTGA	300
AACTGGCGCA	GCGGGAAACT	GGTATGGGTT	TAACGTTTAC	CTGGCGGCGG	GCATGTTGCG	360
GGAAGCCCGC	GCCATGACCA	CACAGATTCA	GGGCGATGTC	ATTCCGTCCA	ATGTGCCCCG	420
TAGCTTTGCC	ATGGCGGTTT	GACAGCCATG	TGGCGTGGTG	CTCGGTATTG	CGCCTTGGA	480
TGCTCCGGTA	ATCCTTGCGG	TACGGGCTGT	TGCGATGCCG	TTGGCATGCG	GCAATACCGT	540
GGTGTTGAAA	AGCTCTGAGC	TGAGTCCCTT	TACCCATCGC	CTGATTGGTC	AGGTGTTGCA	600
TGATGCTGGT	CTGGGGGATG	GCGTGGTGAA	TGTCATCAGC	AATGCCCCGC	AAGACGCTCC	660
TGCGGTGGTG	GAGCGACTGA	TTGCAAAATC	TGCGGTACGT	CGAGTGAAC	TCACCGGTTT	720
GACCCACGTT	GGACGGATCA	TTGGTGAGCT	GTCTGCGCGT	CATCTGAAGC	CTGCTGTGCT	780
GGAATTAGGT	GGTAAGGCTC	CGTTCTTGGT	CTTGACGAT	GCCGACCTCG	ATGCGGCGGT	840
CGAAGCGGCG	GCCTTTGGTG	CCTACTTCAA	TCAGGGTCAA	ATCTGCATGT	CCACTGAGCG	900
TCTGATTGTG	ACAGCAGTCG	CAGACGCCTT	TGTTGAAAAG	CTGGCGAGGA	AGGTCGCCAC	960
ACTGCGTGCT	GGCGATCCTA	ATGATCCGCA	ATCGGTCTTG	GGTTCGTTGA	TTGATGCCAA	1020
TGCAGGTCAA	CGCATCCAGG	TGGGGAGAGG	CGGTTTGCGT	ATTGGGCGCA	TGCATAAAAA	1080
CTGTTGTAAT	TCATTAAAGCA	TTCTGCCGAC	ATGGAAGCCA	TCACAAACGG	CATGATGAAC	1140
CTGAATCGCC	AGCGGCATCA	GCACCTTGTC	GCCTTGCGTA	TAATATTTGC	CCATGGACGC	1200
ACACCGTGGA	AACGGATGAA	GGCAGCAACC	CAGTTGACAT	AAGCCTGTTT	GGTTCGTAAA	1260
CTGTAATGCA	AGTAGCGTAT	GCGCTCACGC	AACCTGGTCCA	GAACCTTGAC	CGAACGCAGC	1320
GGTGGTAACG	GCGCAGTGGC	GGTTTTTCATG	GCTTGTTATG	ACTGTTTTTT	TGTACAGTCT	1380
ATGCCTCGGG	CATCCAAGCA	GCAAGCGCGT	TACGCCGTGG	GTCGATGTTT	GATGTTATGG	1440
AGCAGCAACG	ATGTTACGCA	GCAGCAACGA	TGTTACGCAG	CAGGGCAGTC	GCCCTAAAAC	1500
AAAGTTAGGT	GGCTCAAGTA	TGGGCATCAT	TCCGACATGT	AGGCTCGGCC	CTGACCAAGT	1560
CAAATCCATG	CGGGCTGCTC	TTGATCTTTT	CGGTCGTGAG	TTCCGAGACG	TAGCCACCTA	1620
CTCCCAACAT	CAGCCGGACT	CCGATTACCT	CGGGAACCTG	CTCCGTAGTA	AGACATTTCAT	1680
CGCGCTTGCT	GCCTTCGACC	AAGAAGCGGT	TGTTGGCGCT	CTCGCGGCTT	ACGTTCTGCC	1740
CAGGTTTGAG	CAGCCGCGTA	GTGAGATCTA	TATCTATGAT	CTCGCAGTCT	CCGGCGAGCA	1800
CCGGAGGCAG	GGCATTGCCA	CCGCGCTCAT	CAATCTCCTC	AAGCATGAGG	CCAACGCGCT	1860
TGGTGCTTAT	GTGATCTACG	TGCAAGCAGA	TTACGGTGAC	GATCCCGCAG	TGGCTCTCTA	1920
TACAAAGTTG	GGCATACGGG	AAGAAGTGAT	GCACTTTGAT	ATCGACCCAA	GTACCGCCAC	1980
CTAACAATTC	GTTCAAGCCG	AGATCGGCTT	CCCAATTGGC	CCAGCGCGTC	GATTGCGGCA	2040
TTTGCCATAT	CAATGGACCG	ACTGTGCATG	ACGAGGCTCA	GATGCCATTC	GGTGGGGTGA	2100
AGTCCAGCGG	CTACGGCAGC	TTCGGCAGTC	GAGCATCGAT	TGAGCACTTT	ACCCAGCTGC	2160

GCTGGCTGAC	CATTCAGAAT	GGCCCGCGGC	ACTATCCAAT	CTAAATCGAT	CTTCGGGCGC	2220
CGCGGGCATC	ATGCCCCGGG	CGCTCGCCTC	ATTTCAATCT	CTAACTTGAT	AAAAACAGAG	2280
CTGTTCTCCG	GTCTTGGTGG	ATCAAGGCCA	GTCGCGGAGA	GTCTCGAAGA	GGAGAGTACA	2340
GTGAACGCCG	AGTCCACATT	GCAACCGCAG	GCATCATCAT	GCTCTGCTCA	GCCACGCTAC	2400
CGCAGTGTGT	CGATTGGTCA	TCCTCCGGTT	GAGGTTACGC	AAGACGCTGG	AGGTATTGTC	2460
CGGATGCGTT	CTCTCGAGGC	GCTTCTTCCC	TTCCCGGGTG	GAATTC		2506

Sequenz 15

GAATTCCAAT	AATGACAATA	ATGAGGAGTG	CCCAATGTTT	CACGTGCCCC	TGCTTATTGG	60
TGGTAAGCCT	TGTTCAGCAT	CTGATGAGCG	CACCTTCGAG	CGTCGTAGCC	CGCTGACCGG	120
AGAAGTGGTA	TCGCGCGTCG	CTGCTGCCAG	TTTGGAAGAT	GCGGACGCCG	CAGTGCCCGC	180
TGCACAGGCT	GCGTTTCCTG	AATGGGCGGC	GCTTGCTCCG	AGCGAACGCC	GTGCCCCGACT	240
GCTGCGAGCG	GCGGATCTTC	TAGAGGACCG	TTCTTCCGAG	TTCACCGCCG	CAGCGAGTGA	300
AACTGGCGCA	GCGGGAAACT	GGTATGGGTT	TAACGTTTAC	CTGGCGGCGG	GCATGTTGCG	360
GGAAGCCGCG	GCCATGACCA	CACAGATTCA	GGGCGATGTC	ATTCCGTCCA	ATGTGCCCCG	420
TAGCTTTGCC	ATGGCGGTTT	GACAGCCATG	TGGCGTGGTG	CTCGGTATTG	CGCCTTGGA	480
TGCTCCGGTA	ATCCTTGCGG	TACGGGCTGT	TGCGATGCCG	TTGGCATGCG	GCAATACCGT	540
GGTGTGAAA	AGCTCTGAGC	TGAGTCCCTT	TACCCATCGC	CTGATTGGTC	AGGTGTTGCA	600
TGATGCTGGT	CTGGGGGATG	GCGTGGTGAA	TGTCATCAGC	AATGCCCCGC	AAGACGCTCC	660
TGCGGTGGTG	GAGCGACTGA	TTGCAAATCC	TGCGGTACGT	CGAGTGAAGT	TCACCGGTTT	720
GACCCACGTT	GGACGGATCA	TTGGTGAGCT	GTCTGCGCGT	CATCTGAAGC	CTGCTGTGCT	780
GGAATTAGGT	GGTAAGGCTC	CGTTCTTGGT	CTTGGACGAT	GCCGACCTCG	ATGCGGCGGT	840
CGAAGCGGCG	GCCTTTGGTG	CCTACTTCAA	TCAGGGTCAA	ATCTGCATGT	CCACTGAGCG	900
TCTGATTGTG	ACAGCAGTCG	CAGACGCCTT	TGTTGAAAAG	CTGGCGAGGA	AGGTCGCCAC	960
ACTGCGTGCT	GGCGATCCTA	ATGATCCGCA	ATCGGTCTTG	GGTTCGTTGA	TTGATGCCAA	1020
TGCAGGTCAA	CGCATCCAGG	TTCTGGTCGA	TGATGCGCTC	GCAAAAGGCG	CGCAATGGAA	1080
TTGGCCCAGC	GCGTCGATTC	GGGCATTTGC	CATATCAATG	GACCGACTGT	GCATGACGAG	1140
GCTCAGATGC	CATTCGGTGG	GGTGAAGTCC	AGCGGCTACG	GCAGCTTCGG	CAGTCGAGCA	1200
TCGATTGAGC	ACTTTACCCA	GCTGCGCTGG	CTGACCATTG	AGAATGGCCC	GCGGCACTAT	1260
CCAATCTAAA	TCGATCTTCG	GGCGCCGCGG	GCATCATGCC	CGCGGCGCTC	GCCTCATTTT	1320
AATCTCTAAC	TTGATAAAAA	CAGAGCTGTT	CTCCGGTCTT	GGTGGATCAA	GGCCAGTCGC	1380
GGAGAGTCTC	GAAGAGGAGA	GTACAGTGAA	CGCCGAGTCC	ACATTGCAAC	CGCAGGCATC	1440
ATCATGCTCT	GCTCAGCCAC	GCTACCGCAG	TGTGTCGATT	GGTCATCCTC	CGGTTGAGGT	1500
TACGCAAGAC	GCTGGAGGTA	TTGTCCGGAT	GCGTTCTCTC	GAGGCGCTTC	TTCCCTTCCC	1560
GGGTGGAATT	C					1571

Sequenz 16

GAATTCGCG	GTCGGCGAAA	GTTGATGCGC	TGTATCGTGG	TGAAGATCAA	TCCATGCTGC	60
GTGACGAGGC	CACACTGTGA	GTTGGTCAGG	GGGGGCTTAC	TCGGCGTTTT	CCGACACTGC	120
GTTGGTTGCG	GCAGTGCGCA	CCCCCTGGAT	TGATTGCGGG	GGTGCCCTGT	CGCTGGTGTC	180
GCCTATCGAC	TTAGGGGTAA	AGGTCGCTCG	CGAAGTTCTG	ATGCGTGCGT	CGCTTGAACC	240
ACAAATGGTC	GATAGCGTAC	TCGCAGGCTC	TATGGCTCAA	GCAAGCTTTG	ATGCTTACCT	300
GCTCCCGCGG	CACATTGGCT	TGTACAGCGG	TGTTCCCAAG	TCGGTTCCGG	CCTTGGGGGT	360
GCAGCGCATT	TGCGGCACAG	GCTTCGAACT	GCTTCGGCAG	GCCGGCGAGC	AGATTTCCCA	420
AGGCGCTGAT	CACGTGCTGT	GTGTCGCGGC	AGAGTCCATG	TCGCGTAACC	CCATCGCGTC	480
GTATACACAC	CGGGGCGGGT	TCCGCCTCGG	TGCGCCCGTT	GAGTTCAAGG	ATTTTTTGTG	540
GGAGGCATTG	TTTGATCCTG	CTCCAGGACT	CGACATGATC	GCTACCGCAG	AAAACCTGGG	600
GACAGCAAGC	GAACCGGAAT	TGCCAGCTGG	GGCGCCCTCT	GGTAAGGTTG	GGAAGCCCTG	660
CAAAGTAAAC	TGGATGGCTT	TCTTGCCGCC	AAGGATCTGA	TGGCGCAGGG	GATCAAGATC	720
TGATCAAGAG	ACAGGATGAG	GATCGTTTCG	CATGATTGAA	CAAGATGGAT	TGCACGCAGG	780
TTCTCCGGCC	GCTTGGGTGG	AGAGGCTATT	CGGCTATGAC	TGGGCACAAC	AGACAATCGG	840
CTGCTCTGAT	GCCGCCGTGT	TCCGGCTGTC	AGCGCAGGGG	CGCCCGGTTT	TTTTTGTC	900
GACCGACCTG	TCCGGTGCCC	TGAATGAACT	GCAGGACGAG	GCAGCGCGGC	TATCGTGGCT	960
GGCCACGACG	GGCGTTCCCT	GCGCAGCTGT	GCTCGACGTT	GTCACGTAAG	CGGGAAGGGA	1020
CTGGCTGCTA	TTGGGCGAAG	TGCCGGGGCA	GGATCTCCTG	TCATCTCACC	TTGCTCCTGC	1080
CGAGAAAGTA	TCCATCATGG	CTGATGCAAT	GCGGCGGCTG	CATACGCTTG	ATCCGGCTAC	1140
CTGCCCATTG	GACCACCAAG	CGAAACATCG	CATCGAGCGA	GCACGTAATC	GGATGGAAGC	1200
CGGTCTTGTC	GATCAGGATG	ATCTGGACGA	AGAGCATCAG	GGGCTCGCGC	CAGCCGAACT	1260
GTTGCGCCAGG	CTCAAGGCGC	GCATGCCCCA	CGGCGAGGAT	CTCGTCGTGA	CCCATGGCGA	1320
TGCCTGCTTG	CCGAATATCA	TGGTGGAATA	TGGCCGCTTT	TCTGGATTCA	TCGACTGTGG	1380
CCGGCTGGGT	GTGGCGGACC	GCTATCAGGA	CATAGCGTTG	GCTACCCGTG	ATATTGCTGA	1440
AGAGCTTGGC	GGCGAATGGG	CTGACCGCTT	CCTCGTGCTT	TACGGTATCG	CCGCTCCCGA	1500
TTGCGAGCGC	ATCGCCTTCT	ATCGCCTTCT	TGACGAGTTC	TTCTGAGCGG	GACTCTGGGG	1560
TTCGAAATGA	CCGACCAAGC	GACGCCCAT	GAGGGCGCAA	GAGGAGAAAT	GGATTGACCA	1620
AGAGATCGTG	GCTGTTACGG	ATGAACAGTT	CGATTTAGAG	GGCTACAACA	GTCGAGCAAT	1680
TGAACTGCCT	CGGAAGGCAA	AATTGTTGAT	CGTGACAGTC	ATCCGCGGCC	TAGCAGTCTT	1740
TGAAGCCCTT	TCCCGATTGA	AGCCTGTTCA	TTCTGGCGGG	GTGCAGACTG	CGGGCAACAG	1800
CTGTGCCGTA	GTGGACGGCG	CCGCGGCGGC	TTTGGTGGCT	CGAGAGTCGT	CTGCGACACA	1860
GCCGGTCTTG	GCTAGGATAC	TGGCTACCTC	CGTAGTCGGG	ATCGAGCCCC	AGCATATGGG	1920
GCTCGGCCCT	GCGCCCGCGA	TTGCGCTGCT	GCTTGCGCGT	AGTGATCTTA	GTTTGAGGGA	1980
TATCGACCTC	TTTGAGATAA	ACGAGGCGCA	GGCCGCCCAA	GTTCTAGCGG	TACAGCATGA	2040
ATTGGGTATT	GAGCACTCAA	AACTTAATAT	TTGGGGCGGG	GCCATTGCAC	TTGGACACCC	2100
GCTTGCCGCG	ACCGGATTGC	GTCTCTGCAT	GACCCTCGCT	CACCAATTGC	AAGCTAATAA	2160

CTTTCGATAT	GGAATTGCCT	CGGCATGCAT	TGGTGGGGGA	CAGGGGATGG	CGGTTCTTTT	2220
AGAGAATCCC	CACTTCGGTT	CGTCCTCTGC	ACGAAGTTCG	ATGATTAACA	GAGTTGACCA	2280
CTATCCACTG	AGCTAACGGG	CATCTCCTTT	GTTGCTTTGA	GGTGGCGCAC	GAAGGAGGGC	2340
TCGAAAATCT	CTGCTAAAAA	CAAGAAGAAG	GAACAGGGAA	CATGATTAGT	TTCGCTCGTA	2400
TGGCAGAAAG	TTTAGGAGTC	CAGGCTAAAC	TTGCCCTTGC	CTTCGCACTC	GTATTATGTG	2460
TCGGGCTGAT	TGTTACCGGC	ACGGGTTTCT	ACAGTGTACA	TACCTTGTC	GGGTTGGTGG	2520
GAATTC						2526

Sequenz 17

GAATTCGCG	GTCGGCGAAA	GTTGATGCGC	TGTATCGTGG	TGAAGATCAA	TCCATGCTGC	60
GTGACGAGGC	CACACTGTGA	GTTGGTCAGG	GGGGGCTTAC	TCGGCGTTTT	CCGACACTGC	120
GTTGGTTGCG	GCAGTGCGCA	CCCCCTGGAT	TGATTGCGGG	GGTGCCCTGT	CGCTGGTGTC	180
GCCTATCGAC	TTAGGGGTAA	AGGTCGCTCG	CGAAGTTCTG	ATGCGTGCGT	CGCTTGAACC	240
ACAAATGGTC	GATAGCGTAC	TCGCAGGCTC	TATGGCTCAA	GCAAGCTTTG	ATGCTTACCT	300
GCTCCCGCGG	CACATTGGCT	TGTACAGCGG	TGTTCCCAAG	TCGGTTCCGG	CCTTGGGGGT	360
GCAGCGCATT	TGCGGCACAG	GCTTCCAAGT	GCTTCGGCAG	GCCGGCGAGC	AGATTTCCCA	420
AGGCGCTGAT	CACGTGCTGT	GTGTCGCGGC	AGAGTCCATG	TCGCGTAACC	CCATCGCGTC	480
GTATACACAC	CGGGGCGGGT	TCCGCCTCGG	TGCGCCCGTT	GAGTTCAAGG	ATTTTTTGTG	540
GGAGGCATTG	TTTGATCCTG	CTCCAGGACT	CGACATGATC	GCTACCGCAG	AAAACCTGGG	600
GGAGAGGCGG	TTTGCGTATT	GGGCGCATGC	ATAAAACTG	TTGTAATTCA	TTAAGCATTC	660
TGCCGACATG	GAAGCCATCA	CAAACGGCAT	GATGAACCTG	AATCGCCAGC	GGCATCAGCA	720
CCTTGTCGCC	TTGCGTATAA	TATTTGCCCA	TGGACGCACA	CCGTGGAAAC	GGATGAAGGC	780
ACGAACCCAG	TTGACATAAG	CCTGTTCCGT	TCGTAAACTG	TAATGCAAGT	AGCGTATGCG	840
CTCACGCAAC	TGGTCCAGAA	CCTTGACCGA	ACGCAGCGGT	GGTAACGGCG	CAGTGGCGGT	900
TTTCATGGCT	TGTTATGACT	GTTTTTTTGT	ACAGTCTATG	CCTCGGGCAT	CCAAGCAGCA	960
AGCGCGTTAC	GCCGTGGGTC	GATGTTTGAT	GTTATGGAGC	AGCAACGATG	TTACGCAGCA	1020
GCAACGATGT	TACGCAGCAG	GGCAGTCGCC	CTAAAACAAA	GTTAGGTGGC	TCAAGTATGG	1080
GCATCATTCG	CACATGTAGG	CTCGGCCCTG	ACCAAGTCAA	ATCCATGCGG	GCTGCTCTTG	1140
ATCTTTTCGG	TCGTGAGTTC	GGAGACGTAG	CCACCTACTC	CCAACATCAG	CCGGACTCCG	1200
ATTACCTCGG	GAACTTGCTC	CGTAGTAAGA	CATTTCATCG	GCTTGCTGCC	TTCGACCAAG	1260
AAGCGTTTGT	TGGCGCTCTC	GCGGCTTACG	TTCTGCCCAG	GTTTGAGCAG	CCGCGTAGTG	1320
AGATCTATAT	CTATGATCTC	GCAGTCTCCG	GCGAGCACCG	GAGGCAGGGC	ATTGCCACCG	1380
CGCTCATCAA	TCTCCTCAAG	CATGAGGCCA	ACGCGCTTGG	TGCTTATGTG	ATCTACGTGC	1440
AAGCAGATTA	CGGTGACGAT	CCCGCAGTGG	CTCTCTATAC	AAAGTTGGGC	ATACGGGAAG	1500
AAGTGATGCA	CTTTGATATC	GACCCAAGTA	CCGCCACCTA	ACAATTCGTT	CAAGCCGAGA	1560
TCGGCTTCCC	ATTGAGGGCG	CAAGAGGAGA	AATGGATTGA	CCAAGAGATC	GTGGCTGTTA	1620
CGGATGAACA	GTTCGATTTA	GAGGGCTACA	ACAGTCGAGC	AATTGAACTG	CCTCGGAAGG	1680
CAAAATTGTT	GATCGTGACA	GTCATCCGCG	GCCTAGCAGT	CTTTGAAGCC	CTTTCCCGAT	1740
TGAAGCCTGT	TCATTCTGGC	GGGGTGCGAG	CTGCGGGCAA	CAGCTGTGCC	GTAGTGGACG	1800
GCGCCGCGGC	GGCTTTGGTG	GCTCGAGAGT	CGTCTGCGAC	ACAGCCGGTC	TTGGCTAGGA	1860
TACTGGCTAC	CTCCGTAGTC	GGGATCGAGC	CCGAGCATAT	GGGGCTCGGC	CCTGCGCCCG	1920
CGATTGCGCT	GCTGCTTGCG	CGTAGTGATC	TTAGTTTGAG	GGATATCGAC	CTCTTTGAGA	1980
TAAACGAGGC	GCAGGCCGCC	CAAGTTCTAG	CGGTACAGCA	TGAATTGGGT	ATTGAGCACT	2040
CAAAACTTAA	TATTTGGGGC	GGGGCCATTG	CACTTGGACA	CCCGCTTGCC	GCGACCGGAT	2100
TGCGTCTCTG	CATGACCCTC	GCTCACCAAT	TGCAAGCTAA	TAACTTTCGA	TATGGAATTG	2160

CCTCGGCATG	CATTGGTGGG	GGACAGGGGA	TGGCGGTTCT	TTTAGAGAAT	CCCCACTTCG	2220
GTTCGTCCTC	TGCACGAAGT	TCGATGATTA	ACAGAGTTGA	CCACTATCCA	CTGAGCTAAC	2280
GGGCATCTCC	TTTGTTGCTT	TGAGGTGGCG	CACGAAGGAG	GGCTCGAAAA	TCTCTGCTAA	2340
AAACAAGAAG	AAGGAACAGG	GAACATGATT	AGTTTCGCTC	GTATGGCAGA	AAGTTTAGGA	2400
GTCCAGGCTA	AACTTGCCCT	TGCCTTCGCA	CTCGTATTAT	GTGTCGGGCT	GATTGTTACC	2460
GGCACGGGTT	TCTACAGTGT	ACATACCTTG	TCAGGGTTGG	TGGGAATTC		2509

Sequenz 18

GAATTCGCG	GTCGGCGAAA	GTTGATGCGC	TGTATCGTGG	TGAAGATCAA	TCCATGCTGC	60
GTGACGAGGC	CACACTGTGA	GTTGGTCAGG	GGGGGCTTAC	TCGGCGTTTT	CCGACACTGC	120
GTTGGTTGCG	GCACTGCGCA	CCCCCTGGAT	TGATTGCGGG	GGTGCCCTGT	CGCTGGTGTC	180
GCCTATCGAC	TTAGGGGTAA	AGGTCGCTCG	CGAAGTTCTG	ATGCGTGCGT	CGCTTGAACC	240
ACAAATGGTC	GATAGCGTAC	TCGCAGGCTC	TATGGCTCAA	GCAAGCTTTG	ATGCTTACCT	300
GCTCCCGCGG	CACATTGGCT	TGTACAGCGG	TGTTCCCAAG	TCGGTTCCGG	CCTTGGGGGT	360
GCAGCGCATT	TGCGGCACAG	GCTTCGAACT	GCTTCGGCAG	GCCGGCGAGC	AGATTTCCCA	420
AGGCGCTGAT	CACGTGCTGT	GTGTCGCGGC	AGAGTCCATG	TCGCGTAACC	CCATCGCGTC	480
GTATACACAC	CGGGGCGGGT	TCCGCCTCGG	TGCGCCCGTT	GAGTTCAAGG	ATTTTGTGTG	540
GGAGGCATTG	TTTGATCCTG	CTCCAGGACT	CGACATGATC	GCTACCGCAG	AAAACCTGGC	600
GCGCATTGAG	GGCGCAAGAG	GAGAAATGGA	TTGACCAAGA	GATCGTGGCT	GTTACGGATG	660
AACAGTTCGA	TTTAGAGGGC	TACAACAGTC	GAGCAATTGA	ACTGCCTCGG	AAGGCAAAAT	720
TGTTGATCGT	GACAGTCATC	CGCGGCCTAG	CAGTCTTTGA	AGCCCTTTCC	CGATTGAAGC	780
CTGTTTCATT	TGGCGGGGTG	CAGACTGCGG	GCAACAGCTG	TGCCGTAGTG	GACGGCGCCG	840
CGGCGGCTTT	GGTGGCTCGA	GAGTCGTCTG	CGACACAGCC	GGTCTTGGCT	AGGATACTGG	900
CTACCTCCGT	AGTCGGGATC	GAGCCCGAGC	ATATGGGGCT	CGGCCCTGCG	CCCGCGATTC	960
GCCTGCTGCT	TGCGCGTAGT	GATCTTAGTT	TGAGGGATAT	CGACCTCTTT	GAGATAAACG	1020
AGGCGCAGGC	CGCCCAAGTT	CTAGCGGTAC	AGCATGAATT	GGGTATTGAG	CACTCAAAAC	1080
TTAATATTTG	GGGCGGGGCC	ATTGCACTTG	GACACCCGCT	TGCCGCGACC	GGATTGCGTC	1140
TCTGCATGAC	CCTCGCTCAC	CAATTGCAAG	CTAATAACTT	TCGATATGGA	ATTGCCTCGG	1200
CATGCATTGG	TGGGGGACAG	GGGATGGCGG	TTCTTTTAGA	GAATCCCCAC	TTCGGTTTCGT	1260
CCTCTGCACG	AAGTTCGATG	ATTAACAGAG	TTGACCACTA	TCCACTGAGC	TAACGGGCAT	1320
CTCCTTTGTT	GCTTTGAGGT	GGCGCACGAA	GGAGGGCTCG	AAAATCTCTG	CTAAAAACAA	1380
GAAGAAGGAA	CAGGGAACAT	GATTAGTTTC	GCTCGTATGG	CAGAAAGTTT	AGGAGTCCAG	1440
GCTAAACTTG	CCCTTGCCCT	CGCACTCGTA	TTATGTGTCT	GGCTGATTGT	TACCGGCACG	1500
GGTTTCTACA	GTGTACATAC	CTTGTCAGGG	TTGCTGGGAA	TTC		1543

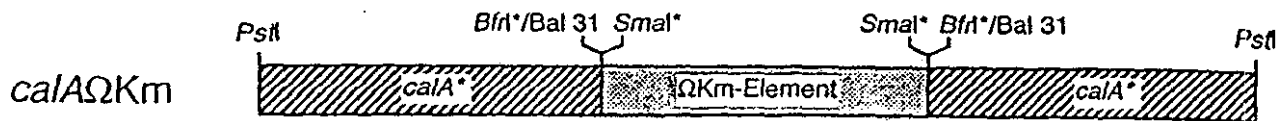


Fig. 1a



Fig. 1b

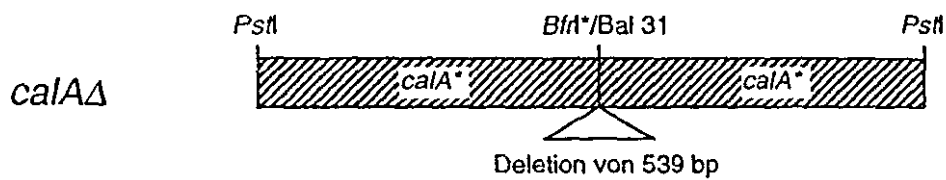


Fig. 1c

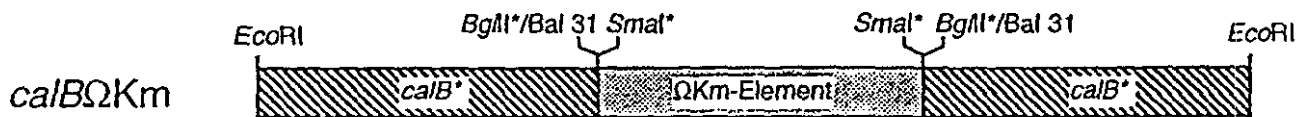


Fig. 1d

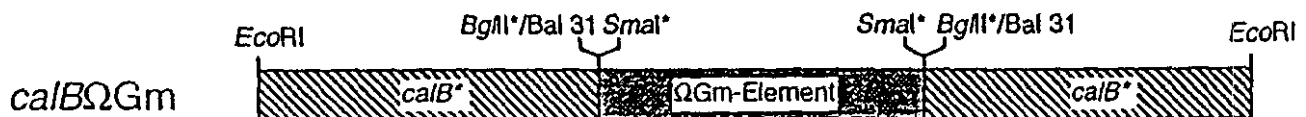


Fig. 1e

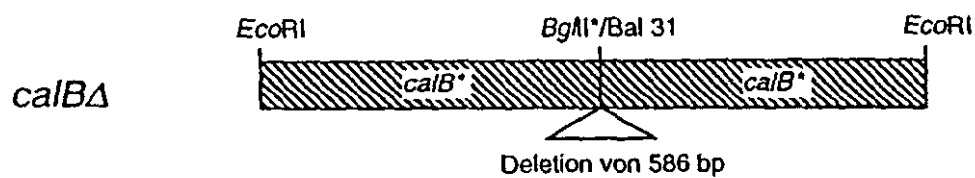


Fig. 1f

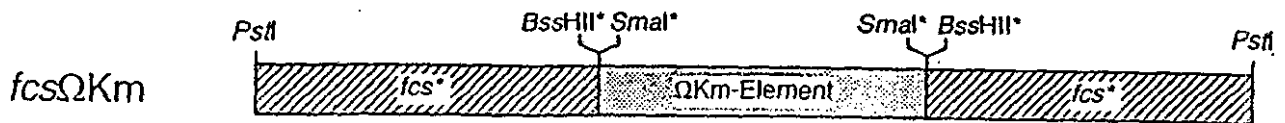


Fig. 1g



Fig. 1h

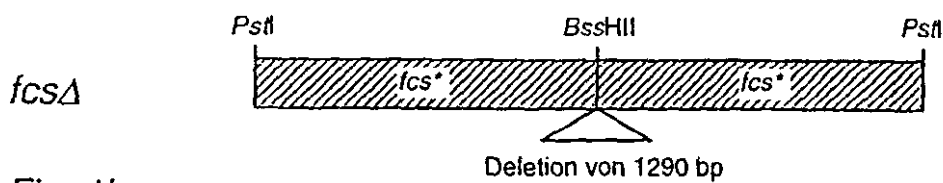


Fig. 1i

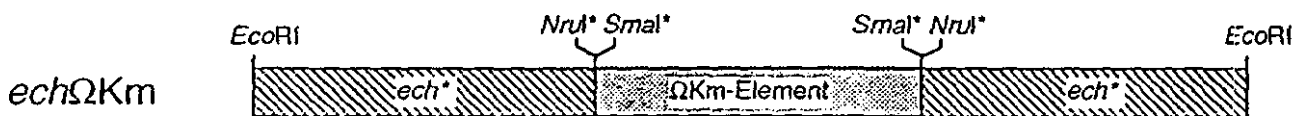


Fig. 1j

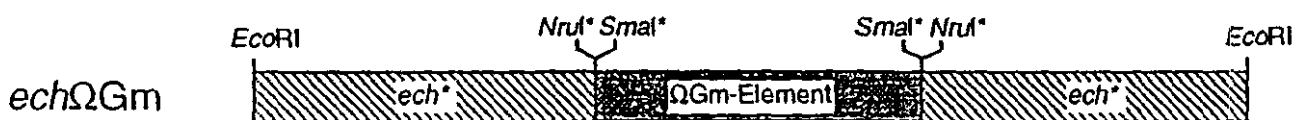


Fig. 1k

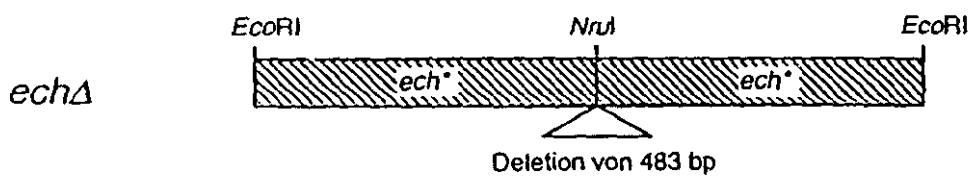


Fig. 1l

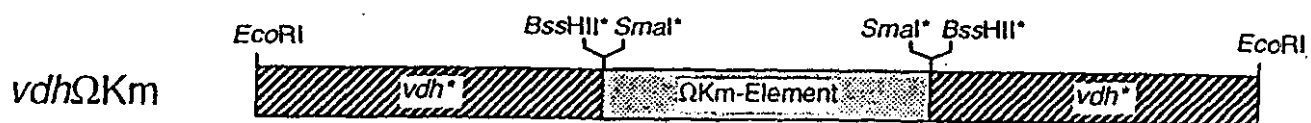


Fig. 1m

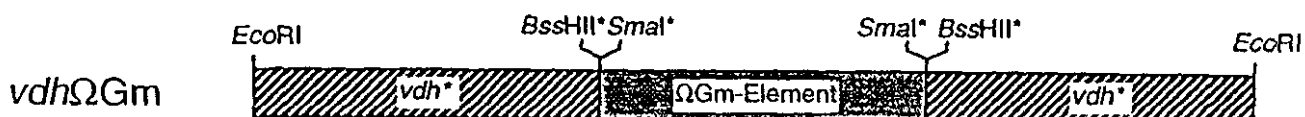


Fig. 1n

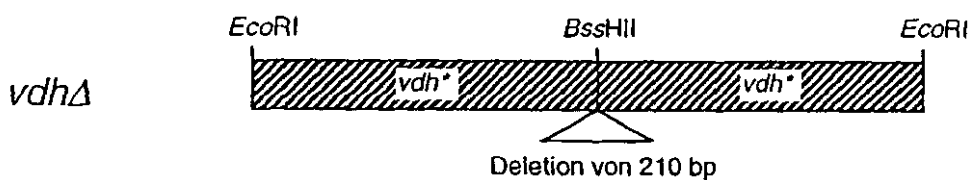


Fig. 1o

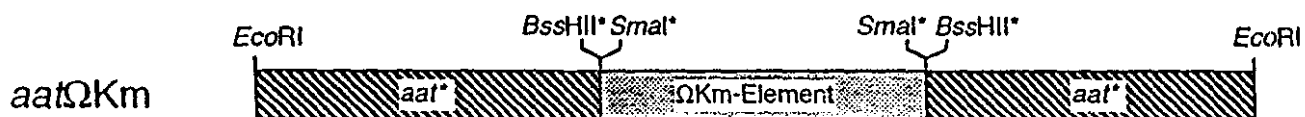


Fig. 1p

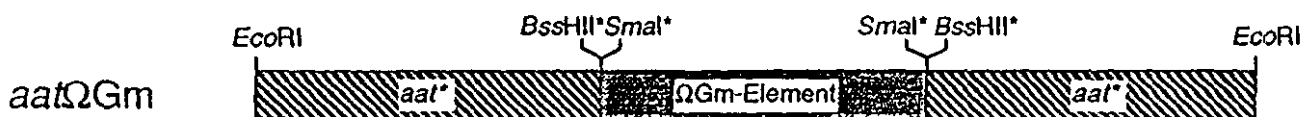


Fig. 1q

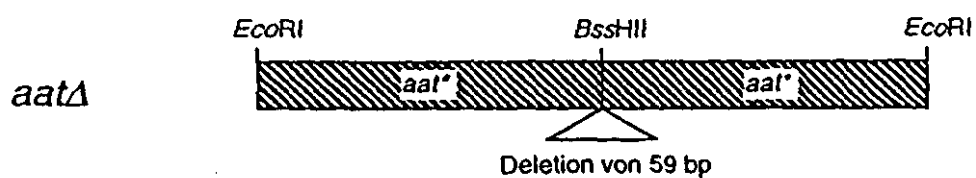


Fig. 1r

HKP0120525

[19]中华人民共和国国家知识产权局

[51]Int. Cl⁷

[12] 发明专利申请公开说明书

[21] 申请号 99812907.0

C12N 15/00

C12N 1/00 C12N 9/02

C12N 9/00 C12N 9/88

C12N 9/10 C12P 7/22

C12P 7/24 C12P 7/42

[43]公开日 2001 年 12 月 5 日

[11]公开号 CN 1325444A

[22]申请日 1999.10.20 [21]申请号 99812907.0

[30]优先权

[32]1998.10.31 [33]DE [31]19850242.7

[86]国际申请 PCT/EP99/07952 1999.10.20

[87]国际公布 WO00/26355 德 2000.5.11

[85]进入国家阶段日期 2001.4.30

[71]申请人 哈尔曼及赖默股份有限公司

地址 德国霍尔兹明登

[72]发明人 J·拉本霍尔斯特 A·斯坦比歇尔

H·普里菲尔特

J·奥维尔哈格

[74]专利代理机构 中国专利代理(香港)有限公司

代理人 罗 宏 谭明胜

权利要求书 2 页 说明书 97 页 附图页数 3 页

[54]发明名称 经过特异性灭活丁子香酚和阿魏酸分解代谢的基因构建制备取代酚的生产菌株

[57]摘要

本发明涉及转化的和/或诱变的单或多细胞生物体,其特征在 于,丁子香酚和/或阿魏酸代谢的酶被灭活,使得中间产物松柏醛,松柏醇,阿魏酸,香草醛和/或香草酸被积累。

ISSN 1008-4274

知识产权出版社出版

权 利 要 求 书

1. 转化的和/或诱变的单细胞或多细胞生物, 其特征在于丁子香酚和/或阿魏酸分解代谢的酶被灭活, 使得中间产物松柏醇, 松柏醛, 阿魏酸, 香草醛和/或香草酸积累。

5 2. 根据权利要求 1 的生物, 其特征在于经过在相应的基因插入 Ω 元件或导入缺失改变丁子香酚和/或阿魏酸分解代谢。

3. 根据权利要求 1 或 2 的生物, 其特征在于编码松柏醇脱氢酶, 松柏醛脱氢酶, 阿魏酰-CoA 合成酶, 烯酰-CoA 水合酶-醛缩酶, β -酮硫解酶, 香草醛脱氢酶或香草酸脱甲基酶的一个或多个基因被改
10 变和/或被灭活。

4. 根据权利要求 1-3 之一的生物, 其特征而在于它是单细胞的, 优选微生物或植物或动物细胞。

5. 根据权利要求 1-4 之一的生物, 其特征而在于它是细菌, 优选假单胞菌属种类。

15 6. 编码松柏醇脱氢酶, 松柏醛脱氢酶, 阿魏酰-CoA 合成酶, 烯酰-CoA 水合酶-醛缩酶, β -酮硫解酶, 香草醛脱氢酶或香草酸脱甲基酶, 或两个或多个该酶的核苷酸序列被改变和/或被灭活的基因结构。

7. 具有图 1a-1r 所给序列的基因结构。

20 8. 具有图 2a-2r 所给序列的基因结构。

9. 含有至少一个根据权利要求 6-8 之一的基因结构的载体。

10. 根据权利要求 1-5 之一的转化的生物, 其特征而在于它含有至少一个根据权利要求 9 的载体。

11. 根据权利要求 1-5 之一的生物, 其特征而在于它含有整合进
25 基因组中代替各完整基因的至少一个根据权利要求 6-8 之一的基因结构。

12. 有机化合物, 特别是醇, 醛和有机酸的生物技术制备方法, 其特征而在于使用根据权利要求 1-5 或 10-11 之一的生物。

13. 制备根据权利要求 1-5 之一的生物的方法, 其特征而在于借
30 助于本身已知的微生物学培养方法实现改变丁子香酚和/或阿魏酸分解代谢。

14. 制备根据权利要求 1-5 或 10-11 之一的生物的方法, 其特

征在于借助于重组 DNA 方法实现改变丁子香酚和/或阿魏酸分解代谢, 和/或灭活相应的基因.

15. 根据权利要求 1-5 或 10-11 之一的生物在制备松柏醇, 松柏醛, 阿魏酸, 香草醛和/或香草酸中的应用.

5 16. 根据权利要求 6-8 之一的基因结构或根据权利要求 9 的载体在制备转化的和/或诱变的生物中的应用.

说明书

经过特异性灭活丁子香酚和阿魏酸分解代谢的
基因构建制备取代酚的生产菌株

5 本发明涉及生产菌株的构建且涉及制备取代的甲氧基苯酚，特别是香草醛的方法。

DE-A 4227076 (制备取代的甲氧基苯酚的方法，及适用于该目的微生物)描述了使用一种新的假单胞菌属种类制备取代的甲氧基苯酚。在该文中起始材料是丁子香酚且产物是阿魏酸、香草酸、松柏醇和松柏醛。

由 Rosazza 等撰写(阿魏酸:一种丰富的芳香族天然产物: J. Ind. Microbiol. 15:457-471)的使用阿魏酸可能的生物转化的综述也于 1995 年出版。

15 在 EP-A 0845532 中描述了来自假单胞菌属种类的合成松柏醇，松柏醛，阿魏酸，香草酸和香草醛的基因和酶。

英国 Norwich 郡食品研究所在 WO 97/35999 中描述了将反式-阿魏酸转化成反式-阿魏酰-SCoA 酯并随后转化成香草醛的酶及裂解该酯的基因。1998 年，该专利的内容以科技文献的形式出版(Gasson 等, 1998, 阿魏酸代谢成香草醛. 生物学化学杂志, 273:4163-4170: Narbad 和 Gasson 1998. 在新分离的荧光假单胞菌菌株中使用新的 CoA-依赖型途径经香草醛代谢阿魏酸. 微生物学, 144:1397-1405).

DE-A-195 32 317 描述了使用拟无枝酸菌属种类经发酵从阿魏酸以高产量获得香草醛。

25 已知方法的缺点在于它们仅能获得极低产量的香草醛或者使用相当昂贵的起始化合物。尽管最后提到的方法(DE-A 195 32 317)确实实现了高产量，但使用假单胞菌属种类 HR199 和拟无枝酸菌属种类 HR167 将丁子香酚生物转化成香草醛需要进行 2 步发酵，因此导致相当昂贵和费时。

30 因此本发明的目的是构建能以一步方法相当廉价地将原材料丁子香酚转化成香草醛的生物。

借助于构建单细胞的生产菌株或多细胞的生物实现该目的，该菌株的特征在于丁子香酚和/或阿魏酸分解代谢的酶被灭活使得积累中

间产物松柏醇，松柏醛，阿魏酸，香草醛和/或香草酸。

该生产菌株可以是单细胞或多细胞。因此，本发明涉及微生物，植物或动物。而且也可使用从生产品种获得的提取物。根据本发明，
5 优选使用单细胞生物。这些所述的生物可以是微生物或动物或植物细胞。根据本发明，特别优选使用真菌和细菌。最优选的是细菌种类。在其丁子香酚和/或阿魏酸分解代谢被改变后特别可使用的细菌是红球菌属，假单胞菌属和埃希氏杆菌属的种类。

在最简单的情况下，已知的常规微生物学方法可用于分离根据本发明使用的生物。

10 因此，可使用酶抑制剂改变丁子香酚和/或阿魏酸分解代谢中涉及的蛋白质的酶活性。而且，经过突变编码这些蛋白质的基因可改变丁子香酚和/或阿魏酸分解代谢中涉及的蛋白质的酶活性。该突变可借助于传统方法以随机方式产生，例如经过使用紫外照射或诱导突变的化学剂。

15 诸如缺失，插入和/或核苷酸交换的重组 DNA 方法同样适合于分离新生物。因此，例如可使用其它的 DNA 元件 (Ω 元件) 灭活该生物的基因。合适的载体可用于以改变和/或灭活的基因结构取代完整基因。在本文中，待灭活的基因，和用于灭活的 DNA 元件可借助于传统克隆技术或借助于聚合酶链式反应 (PCR) 获得。

20 例如，在本发明的一个可能的实施方案中，可经过插入 Ω 元件或导入缺失进入合适的基因中来改变丁子香酚分解代谢和阿魏酸分解代谢。在本文中，可使用上述重组 DNA 方法灭活编码脱氢酶，合成酶，水合酶 - 醛缩酶，硫解酶或脱甲基酶的基因的功能以便阻断相关酶的生产。优选的是，该基因是编码松柏醇脱氢酶，松柏醛脱氢酶，阿魏
25 酰 CoA 合成酶，烯酰 CoA 水合酶 - 醛缩酶， β -酮硫解酶，香草醛脱氢酶或香草酸脱甲基酶的基因。非常特别优选的是编码在 EP - A 0845532 中限定的氨基酸序列的基因和/或编码其等位变异体的核苷酸序列。

因此本发明还涉及制备转化生物和突变体的基因结构。

30 优选采用编码脱氢酶，合成酶，水合酶 - 醛缩酶，硫解酶或脱甲基酶的核苷酸序列被灭活的基因结构来分离该生物和突变体。特别优选的是编码松柏醇脱氢酶，松柏醛脱氢酶，阿魏酰 CoA 合成酶，烯酰

CoA 水合酶 - 醛缩酶, β - 酮硫解酶, 香草醛脱氢酶或香草酸脱甲基酶的核苷酸序列被灭活的基因结构。极其优选的基因结构表现为图 1a 至 1r 给出的结构, 具有图 2a 至 2r 所示的核苷酸序列和/或编码其等位变异体的核苷酸序列。在本文中, 特别优选的是核苷酸序列 1 至

5 18.

本发明还包含这些基因结构的部分序列及其功能等同物。功能等同物的意思应理解为该 DNA 的衍生物, 其中个别的核苷酸碱基被交换 (摆动交换 (Wobbel austausche)) 而不改变其功能。在蛋白质水平上也可交换氨基酸而不导致其功能改变。

10 在该基因结构的上游和/或下游可插入一个或多个 DNA 序列。经过克隆该基因结构, 可获得适用于转化和/或转染生物和/或适用于接合转移进生物的质粒或载体。

本发明还涉及用于制备根据本发明转化的生物和突变体的质粒和/或载体。因而这些生物和突变体含有所述的基因结构。因此本发明还涉及含有所说质粒和/或载体的生物。

质粒和/或载体的性质取决于其应用的对象。例如, 为了用以 Ω 元件灭活的基因取代假单胞菌中丁子香酚和/或阿魏酸分解代谢的完整基因, 所需的载体一方面可转移进假单胞菌中 (经接合可转移的质粒), 但另一方面不能在这些生物中复制, 因此在假单胞菌中是不稳定的 (所谓的自杀质粒)。借助于该质粒系统转移进假单胞菌的 DNA 片段只有当它们经同源重组整合进细菌细胞基因组中时才能保留。

20 所述的基因结构, 载体和质粒可用于制备不同的转化生物或突变体。所说的基因结构可用于以改变的和/或灭活的基因结构取代完整的核苷酸序列。在这种可经转化或转染或接合获得的细胞中, 用改变的和/或灭活的基因结构经同源重组取代完整的基因, 因此所得的细胞最终在其基因组中仅具有改变的和/或灭活的基因结构。以这种方式, 根据本发明优选可改变和/或灭活基因, 以便有关生物能产生松柏醇, 松柏醛, 阿魏酸, 香草醛和/或香草酸。

30 在 DE - A 4227076 和 EP - A 0845532 中详细描述菌株假单胞菌属种类 HR199 (DSM 7063) 的突变体是根据本发明以这种方式构建的生产菌株的例子, 其中相应的基因结构特别是来自图 1a 至 1r 与图 2a 至 2r:

1. 假单胞菌属种类 HR199calA Ω Km, 含有 Ω Km-灭活的 calA 基因代替编码松柏醇脱氢酶的完整 calA 基因 (图 1a; 图 2a)。

2. 假单胞菌属种类 HR199calA Ω Gm, 含有 Ω Gm 灭活的 calA 基因代替编码松柏醇脱氢酶的完整 calA 基因 (图 1b; 图 2b)。

5 3. 假单胞菌属种类 HR199calA Δ , 含有缺失灭活的 calA 基因代替编码松柏醇脱氢酶的完整 calA 基因 (图 1c; 图 2c)。

4. 假单胞菌属种类 HR199calB Ω Km, 含有 Ω Km 灭活的 calB 基因代替编码松柏醛脱氢酶的完整 calB 基因 (图 1d; 图 2d)。

10 5. 假单胞菌属种类 HR199calB Ω Gm, 含有 Ω Gm-灭活的 calB 基因代替编码松柏醛脱氢酶的完整 calB 基因 (图 1e; 图 2e)。

6. 假单胞菌属种类 HR199calB Δ , 含有缺失灭活的 calB 基因代替编码松柏醛脱氢酶的完整 calB 基因 (图 1f; 图 2f)。

7. 假单胞菌属种类 HR199fcs Ω Km, 含有 Ω Km-灭活的 fcs 基因代替编码阿魏酰 CoA 合成酶的完整 fcs 基因 (图 1g; 图 2g)。

15 8. 假单胞菌属种类 HR199fcs Ω Gm, 含有 Ω Gm 灭活的 fcs 基因代替编码阿魏酰-CoA 合成酶的完整 fcs 基因 (图 1h; 图 2h)。

9. 假单胞菌属种类 HR199fcs Δ , 含有缺失灭活的 fcs 基因代替编码阿魏酰 CoA 合成酶的完整 fcs 基因 (图 1i; 图 2i)。

20 10. 假单胞菌属种类 HR199ech Ω Km, 含有 Ω Km 灭活的 ech 基因代替编码烯酰-CoA 水合酶-醛缩酶的完整 ech 基因 (图 1j; 图 2j)。

11. 假单胞菌属种类 HR199ech Ω Gm, 含有 Ω Gm 灭活的 ech 基因代替编码烯酰-CoA 水合酶-醛缩酶的完整 ech 基因 (图 1k; 图 2k)。

12. 假单胞菌属种类 HR199ech Δ , 含有缺失灭活的 ech 基因代替编码烯酰-CoA 水合酶-醛缩酶的完整 ech 基因 (图 1l; 图 2l)。

25 13. 假单胞菌属种类 HR199aat Ω Km, 含有 Ω Km-灭活的 aat 基因代替编码 β -酮硫解酶的完整 aat 基因 (图 1m, 图 2m)。

14. 假单胞菌属种类 HR199aat Ω Gm, 含有 Ω Gm-灭活的 aat 基因代替编码 β -酮硫解酶的完整 aat 基因 (图 1n; 图 2n)。

30 15. 假单胞菌属种类 HR199aat Δ , 含有缺失灭活的 aat 基因代替编码 β -酮硫解酶的完整 aat 基因 (图 1o; 图 2o)。

16. 假单胞菌属种类 HR199vdh Ω Km, 含有 Ω Km-灭活的 vdh 基因代替编码香草醛脱氢酶的完整 vdh 基因 (图 1p; 图 2p)。

17. 假单胞菌属种类 HR199vdh Ω Gm, 含有 Ω Gm-灭活的 vdh 基因代替编码香草醛脱氢酶的完整 vdh 基因 (图 1q; 图 2q)。

18. 假单胞菌属种类 HR199vdh Δ , 含有缺失灭活的 vdh 基因代替编码香草醛脱氢酶的完整 vdh 基因 (图 1r; 图 2r)。

5 19. 假单胞菌属种类 HR199vdhB Ω Km, 含 Ω Km 灭活的 vdhB 基因代替编码香草醛脱氢酶 II 的完整 vdhB 基因。

20. 假单胞菌属种类 HR199vdhB Ω Gm, 含有 Ω Gm-灭活的 vdhB 基因代替编码香草醛脱氢酶 II 的完整 vdhB 基因。

10 21. 假单胞菌属种类 HR199vdhB Δ , 含有缺失灭活的 vdhB 基因代替编码香草醛脱氢酶 II 的完整 vdhB 基因。

22. 假单胞菌属种类 HR199adh Ω Km, 含有 Ω Km 灭活的 adh 基因代替编码乙醇脱氢酶的完整 adh 基因。

23. 假单胞菌属种类 HR199adh Ω Gm, 含有 Ω Gm-灭活的 adh 基因代替编码乙醇脱氢酶的完整 adh 基因。

15 24. 假单胞菌属种类 HR199adh Δ , 含有缺失灭活的 adh 基因代替编码乙醇脱氢酶的完整 adh 基因。

25. 假单胞菌属种类 HR199vanA Ω Km, 含有 Ω Km 灭活的 vanA 基因代替编码香草酸脱甲基酶 α -亚基的完整 vanA 基因。

20 26. 假单胞菌属种类 HR199vanA Ω Gm, 含有 Ω Gm-灭活的 vanA 基因代替编码香草酸脱甲基酶 α -亚基的完整 vanA 基因。

27. 假单胞菌属种类 HR199vanA Δ , 含有缺失灭活的 vanA 基因代替编码香草酸脱甲基酶 α -亚基的完整 vanA 基因。

28. 假单胞菌属种类 HR199vanB Ω Km, 含有 Ω Km 灭活的 vanB 基因代替编码香草酸脱甲基酶 β -亚基的完整 vanB 基因。

25 29. 假单胞菌属种类 HR199vanB Ω Gm, 含有 Ω Gm-灭活的 vanB 基因代替编码香草酸脱甲基酶 β -亚基的完整 vanB 基因。

30. 假单胞菌属种类 HR199vanB Δ , 含有缺失灭活的 vanB 基因代替编码香草酸脱甲基酶 β -亚基的完整 vanB 基因。

30 本发明还涉及有机化合物的生物技术制备方法。特别是该方法可用于制备醇, 醛和有机酸。后者优选是松柏醇, 松柏醛, 阿魏酸, 香草醛和香草酸。

上述生物用于该新方法。特别优选的生物包括细菌, 特别是假单

胞菌属种类。具体地说，上述假单胞菌属种类优选用于下面的方法：

1. 假单胞菌属种类 HR199calA Ω Km，假单胞菌属种类 HR199calA Ω Gm 和假单胞菌属种类 HR199calA Δ 用于从丁子香酚制备松柏醇。

5 2. 假单胞菌属种类 HR199calB Ω Km，假单胞菌属种类 HR199calB Ω Gm 和假单胞菌属种类 HR199calB Δ 用于从丁子香酚或松柏醇制备松柏醛。

3. 假单胞菌属种类 HR199fcs Ω Km，假单胞菌属种类 HR199fcs Ω Gm，假单胞菌属种类 HR199fcs Δ ，假单胞菌属种类
10 HR199ech Ω Km，假单胞菌属种类 HR199ech Ω Gm 和假单胞菌属种类 HR199ech Δ 用于从丁子香酚或松柏醇或松柏醛制备阿魏酸。

4. 假单胞菌属种类 HR199vdh Ω Km，假单胞菌属种类 HR199vdh Ω Gm，假单胞菌属种类 HR199vdh Δ ，假单胞菌属种类 HR199vdh Ω GmvdhB Ω Km，假单胞菌属种类 HR199vdh Ω KmvdhB Ω Gm，假
15 单胞菌属种类 HR199vdh Δ vdhB Ω Gm 和假单胞菌属种类 HR199vdh Δ vdhB Ω Km 用于从丁子香酚或松柏醇或松柏醛或阿魏酸制备香草醛。

5. 假单胞菌属种类 HR199vanA Ω Km，假单胞菌属种类 HR199vanA Ω Gm，假单胞菌属种类 HR199vanA Δ ，假单胞菌属种类
20 HR199vanB Ω Km，假单胞菌属种类 HR199vanB Ω Gm 和假单胞菌属种类 HR199vanB Δ 用于从丁子香酚或松柏醇或松柏醛或阿魏酸或香草醛制备香草酸。

丁子香酚是优选的底物。然而，巴可加入其它底物或甚至用其它底物代替丁子香酚。

25 根据本发明采用的用于该生物合适营养培养基是合成的，半合成的或复合培养基。这些培养基可包括含碳和含氮化合物，无机盐，其中包括合适的痕量元素和维生素。

合适的含碳化合物是碳水化合物，碳氢化合物或有机标准化学剂。可优选使用的化合物的例子是糖，醇或糖醇，有机酸或复合物混
30 合物。

优选的糖是葡萄糖。可优选使用的有机酸是柠檬酸或乙酸。复合混合物的例子是麦芽提取物，酵母提取物，酪蛋白或酪蛋白水解产

物。

无机化合物是合适的含氮物质。其例子有硝酸盐和铵盐。也可使用有机氮源。这些氮源包括酵母提取物，大豆粉、酪蛋白，酪蛋白水解物和玉米浆。

- 5 可使用的无机盐的例子是硫酸盐，硝酸盐，氯化物，碳酸盐和磷酸盐。所说的盐所含的金属优选是钠，钾，镁，锰，钙，锌和铁。

培养温度优选范围是 50 至 100℃。特别优选的范围是从 15 至 60℃，最优选 22 至 37℃

培养基的 pH 优选是 2 至 12。特别优选的范围是 4 至 8。

- 10 原则上，技术人员已知的任何生物反应器均可用于实施该新方法。优先考虑的是适于浸没方法的任何装置。这意味着根据本发明可采用具有或不具有机械混合装置的容器。后者的例子是摇动装置，泡罩塔反应器或循环式反应器。前者优选包括所有已知的装配有任何设计的搅拌器的装置。

- 15 可连续或分批实施该新方法。达到最大产量所需的发酵时间取决于所用生物的具体特性。然而，原则上，发酵时间在 2 至 200 小时之间。

下面更详细地解释本发明，在提及的实施例中：

- 20 利用丁子香酚的菌株假单胞菌属种类 HR199 (DSM 7063) 的突变体是借助于插入 Ω 元件或导入缺失经特异性灭活丁子香酚分解代谢的基因以靶向方式产生的。采用的 Ω 元件是编码对抗生素卡那霉素 (Ω Km) 和庆大霉素 (Ω Gm) 产生抗性的 DNA 片段。这些抗性基因是使用标准方法从 Tn5 和质粒 pBBR 1MCS-5 分离出的。编码松柏醇脱氢酶，松柏醛脱氢酶，阿魏酰-CoA 合成酶，烯酰-CoA 水合酶-醛缩酶， β -酮硫解酶，香草醛脱氢酶，乙醇脱氢酶，香草醛脱氢酶 II 和香草酸脱甲基酶的基因 calA, calB, fcs, ech, aat, vdh, adh, vdhB, vanA 和 vanB 使用标准方法从菌株假单胞菌属种类 HR199 的基因组 DNA 分离出并克隆进 pBluescript SK⁻。借助于用合适的限制性核酸内切酶消化，可从这些基因中去除 DNA 片段（缺失）或用 Ω 元件取代
- 25
- 30 （插入），导致各基因失活。以这种方式突变的基因重新克隆进可接合转移的载体中并随后导入菌株假单胞菌属种类 HR199。使用合适的选择获得用新导入的失活基因取代了各有功能的野生型基因的转接

合子(transkonju-ganten). 以这种方式获得的插入和缺失突变体现在仅具有各失活的基因. 使用这一方法可获得仅具有一种缺陷基因的突变体以及用该方式灭活了几种基因的多重突变体. 这些突变体可用于进行下列生物转化:

- 5 a) 将丁子香酚转化为松柏醇, 松柏醛, 阿魏酸, 香草醛和/或香草酸;
- b) 将松柏醇转化成松柏醛, 阿魏酸, 香草醛和/或香草酸;
- c) 将松柏醛转化成阿魏酸, 香草醛和/或香草酸;
- d) 将阿魏酸转化成香草醛和/或香草酸, 和
- 10 e) 将香草醛转化成香草酸.

材料和方法

生长细菌的条件

- 大肠杆菌的菌株在 37℃ 下在 Luria-Bertani (LB) 或 M9 无机培养基中繁殖 (J. Sambrook, E. F. Fritsch 和 T. Maniatis, 1989. 分子克隆; 实验室手册, 第 2 版, 冷泉港实验室出版, 纽约冷泉港). 假单胞菌属种类的菌株在营养肉汤 (NB, 0.8%, wt/vol) 或无机培养基 (MM) (H. G. Schlegel 等, 1961, 微生物学报 38:209-222) 或 HR 无机培养基 (HR-MM) (J. Rabenhorst, 1996, 应用微生物学及生物技术, 46:470-474) 中 30℃ 下繁殖. 阿魏酸, 香草醛, 香草酸和原儿茶酸溶于二甲基甲砒并加入到各培养基中以得到 0.1% (wt/vol) 的终浓度. 丁子香酚可直接加入到培养基中以得到 0.1% (vol/vol) 的终浓度或加入到 MM 琼脂板盖的滤纸 (圆形滤纸 595, Schleicher & Schuell, Dassel, 德国) 上. 繁殖假单胞菌属种类的转接合子和突变体时, 分别使用终浓度为 25 µg/ml, 100 µg/ml 和 7.5 µg/ml 的四环素, 卡那霉素和庆大霉素.

培养上清液中代谢中间产物的定性和定量检测.

- 用高压液相色谱 (Knauer HPLC) 直接或用双蒸水稀释后分析培养上清液. 在 Nucleosil 100 C18 (7 µm, 250 × 4mm) 上进行色谱. 0.1% (vol/vol) 甲酸和乙腈用作溶剂. 用于洗脱物质的梯度过程如下:

00:00-06:30 → 26% 乙腈

06:30-08:00 → 100% 乙腈

08:00-12:00→100% 乙腈

12:00-13:00→26% 乙腈

13:00-18:00→26% 乙腈

香草醛脱氢酶 II 的纯化.

5 在 4℃ 下进行纯化.

粗制提取物

将在丁子香酚中繁殖的假单胞菌属种类 HR199 细胞在 10 mM 磷酸钠缓冲液 pH6.0 中洗涤, 然后重悬于相同缓冲液中并在 1000 psi 的压力下通过弗氏压碎器 (Amicon, silver Spring, 美国马里兰州) 2
10 次破碎细胞. 对细胞匀浆物进行超速离心 (1h, 100, 000 xg, 4℃), 所得的粗提物的可溶性部分从上清液中获得.

在 DEAE 葡聚糖纤维素上进行阴离子交换色谱:

粗提物的可溶部分在 10 mM 磷酸钠缓冲液, pH6.0 中透析过夜. 将透析产物上样到用 10 mM 磷酸钠缓冲液 pH6.0 平衡过的 DEAE -
15 Sephacel 柱 (2.6cm × 35cm, 柱容器 [BV]: 186ml) 上, 其流速为 0.8 ml/分. 用 2 倍 BV 的 10 mM 磷酸钠缓冲液 pH6.0 冲洗柱. 香草醛脱氢酶 II (VDH II) 用在 10 mM 磷酸钠缓冲液 pH6.0 (750 ml) 中从 0 到 400 mM NaCl 的线性盐梯度洗脱. 收集 10 ml 馏份. 合并具有高 VDH II 活性的馏份以形成 DEAE 合并液.

20 测定香草醛脱氢酶活性

使用光学酶试验在 30℃ 下测定 VDH 活性. 体积为 1 ml 的反应混合物含 0.1 mmol 磷酸钾 (pH 7.1), 0.125 μmol 香草醛, 0.5 μmol NAD, 1.2 μmol 丙酮酸 (钠盐), 乳酸脱氢酶 (IU; 来自猪心) 和酶溶液. 在 $\lambda = 340 \text{ nm}$ 下监测香草醛的氧化 ($\epsilon_{\text{香草醛}} = 11.6 \text{ cm}^2/\mu\text{mol}$). 酶活
25 性以单位 (U) 表示, 1U 相应于每分钟转变 1 μmol 香草醛的酶的量. 使用 Lowry 等 (O. H. Lowry, N. J. Rosebrough, A. L. Farr 和 R. J. Randall, 1951, 生物学化学杂志, 193: 265-275) 的方法测定样品中的蛋白质浓度.

测定松柏醇脱氢酶活性

30 根据 Jaeger 等 (E. L. Jaeger, Eggeling 和 H. Sahm, 1981. 普通微生物学, 6:333-336) 使用光学酶试验在 30℃ 下测定 CADH 活性. 体积为 1ml 的反应混合物含有 0.2 mmol tris/HCL (pH9.0), 0.4 μmol

松柏醇, 2 μmol NAD, 0.1 mmol 氨基脲和酶溶液. 在 $\lambda = 340 \text{ nm}$ 下监测 NAD 的减少 ($\epsilon = 6.3 \text{ cm}^2/\mu\text{mol}$). 酶活性用单位 (U) 来表示, 1U 相应于每分钟转化 1 μmol 底物的酶量. 按 Lowry 等 (O. H. Lowry, N. J. Rosebrough, A. L. Farr 和 R. J. Randall, 1951, 生物学化学杂志, 193:265-275) 的方法测定样品中的蛋白质浓度.

测定松柏醛脱氢酶的活性

使用光学酶试验在 30°C 下测定 CALDH 活性. 体积为 1ml 的反应混合物含有 0.1 mmol 的 tris/HCL (pH 8.8), 0.08 μmol 的松柏醛, 2.7 μmol 的 NAD 和酶溶液. 在 $\lambda = 400 \text{ nm}$ 下监测松柏醛氧化成阿魏酸 ($\epsilon = 34 \text{ cm}^2/\mu\text{mol}$). 酶活性以单位 (U) 表示, 1U 相应于每分钟转变 1 μmol 底物的酶量. 按 Lowry 等 (O. H. Lowry, N. J. Rosebrough, A. L. Farr 和 Randall, 1951, 生物学化学杂志, 193:265-275) 的方法测定样品中的蛋白质浓度.

测定阿魏酰 CoA 合成酶 (阿魏酸硫激酶) 活性.

使用 Zenk 等 (Zenk 等, 1980, 生物化学年鉴, 101:182-187) 改良的光学酶试验在 30°C 下测定 FCS 活性. 体积为 1 ml 的反应混合物含有 0.09 mmol 的磷酸钾 (pH 7.0), 2.1 μmol 的 MgCl_2 , 0.7 μmol 的阿魏酸, 2 μmol 的 ATP, 0.4 μmol 的辅酶 A 和酶溶液. 在 $\lambda = 345 \text{ nm}$ 下监测从阿魏酸形成 CoA 酯 ($\epsilon = 10 \text{ cm}^2/\mu\text{mol}$). 酶活性以单位 (U) 表示, 1U 相应于每分钟转化 1 μmol 底物的酶量. 使用 Lowry 等 (O. H. Lowry, N. J. Rosebrough, A. L. Farr 和 R. J. Randall, 1951, 生物学化学杂志, 193:265-275) 的方法测定样品中的蛋白质浓度.

电泳方法

使用 Stegemann 等 (Stegemann 等, 1973, Z. Naturforsch. 28c:722-732) 的方法在 7.4% (wt/vol) 聚丙烯酰胺凝胶中天然条件下或者使用 Laemmli (Laemmli, U. K. 1970, 自然 (伦敦) 227:680-685) 的方法在 11.5% (wt/vol) 聚丙烯酰胺凝胶中在变性条件下分离含蛋白质的提取物. Serva Blue R 用于非特异性蛋白质染色. 为了特异性染色松柏醇脱氢酶, 松柏醛脱氢酶和香草醛脱氢酶, 将凝胶在 100 mM KP 缓冲液 (pH 7.0) 中重新缓冲 20 分钟并随后在加入了 0.08% (wt/vol) NAD, 0.04% (wt/vol) 对硝基蓝四唑氯, 0.003% (wt/vol) 吩嗪硫酸甲酯和 1 mM 的各底物的相同缓冲液中 30°C 下温育直到看见

相应的颜色带。

将蛋白质从聚丙烯酰胺凝胶上转移到 PVDF 膜上。

使用半干快速印迹器具 (B32/33, Biometra, Göttingen, 德国) 按照厂家说明将蛋白质从 SDS - 聚丙烯酰胺凝胶转移到 PVDF 膜

5 (Waters-Millipore, Bedford, Mass, USA) 上。

测定 N 端氨基酸序列

使用蛋白质肽测序仪 (477A 型, 应用生物系统, Foster City, USA) 和 PTH 分析仪按照厂商说明测定 N - 端氨基酸序列。

分离和操作 DNA

10 使用 Marmur 的方法 (J. Marmur, 1961, 分子生物学杂志, 3:208-218) 分离基因组 DNA。使用标准方法 (J. E. Sambrook, F. Fritsch 和 T. Maniatis, 1989, 分子克隆: 实验室手册, 第 2 版, 冷泉港实验室出版, 纽约冷泉港) 分离和分析其它质粒 DNA 和/或 DNA 限制性片段。

15 转移 DNA

使用 Hanahan 的方法 (D. Hanahan, 1983, 分子生物学杂志, 166:557-580) 制备和转化感受态的大肠杆菌细胞。在含有质粒的大肠杆菌 S17-1 菌株 (供体) 和假单胞菌属种类菌株 (受体) 之间的接合质粒转移在 NB 琼脂板上按 Friedrich 等的方法 (B. Friedrich 等, 1981, 细菌学杂志, 147: 198-205) 进行或借助于“微量互补法”在含 0.5% (wt/vol) 葡萄糖酸作碳源及 25 μ g 四环素/ml 或 100 μ g 卡那霉素/ml 的 MM 琼脂平板上进行。在这种情况下, 将受体细胞以一个方向进行接种划线。5 分钟后, 将供体菌株的细胞进行接种划线, 这些划线与受体接种划线相交叉。在 30℃ 培养 48h 后, 转接合子直接位于交叉点下方生长, 而供体菌株和受体菌株都不能生长。

25

杂交实验

在 0.8% (wt/vol) 琼脂糖凝胶上在 50 mM tris-50 mM 硼酸 - 1.25 mM EDTA 缓冲液 (pH 8.5) 中电泳分离 DNA 限制性酶切片段 (J. E. Sambrook, F. Fritsch 和 T. Maniatis, 1989, 分子克隆: 实验室手册, 第 2 版, 冷泉港实验室出版, 纽约冷泉港)。将变性的 DNA 从凝胶转移到带正电的尼龙膜 (孔径: 0.45 μ m, Pall Filtrationstechnik, Dreieich, 德国) 上, 随后与生物素标记的

30

或洋地黄毒苷标记的 DNA 探针的杂交, 及这些 DNA 探针的制备均使用标准方法进行 (J. E. Sambrook, F. Fritsch 和 T. Maniatis, 1989, 分子克隆: 实验室手册, 第 2 版, 冷泉港实验室出版社, 纽约冷泉港).

DNA 测序:

- 5 按 Sanger 等 (Sanger 等 1977, 美国科学院学报 74:5463-5467) 的双脱氧链终止法使用 "LI-COR" DNA 测序仪 400L 型 (LI-COR 公司, 生物技术部, Lincoln, NE, USA) 并使用 "含 7-脱氮-2-脱氧鸟苷三磷酸的热测序酶荧光标记的引物循环测序试剂盒" (Amersham 生命科学, Amersham International pls., Little Chalfont, Buckinghamshire, 英国) "非放射活性" 测定核苷酸序列, 在每种情况按厂商说明书进行.

按 Strauss 等 (E. C. Strauss 等, 1986, 生物化学年鉴, 154:353-360) 的 "引物跳跃方法" 使用合成寡核苷酸进行测序.

化学试剂, 生化试剂和酶

- 15 从 C. F. Boehringer & Söhne (Mannheim, 德国) 或从 GIBCO/BRL (Eggenstein, 德国) 获得限制性酶, T4 DNA 连接酶, λ DNA 和酶及光学酶试验的底物. $[\gamma - ^{32}\text{P}]$ ATP 来自 Amersham/Buchler (Braunschweig, 德国). 寡核苷酸从 MWG - Btotech GmbH (Ebersberg, 德国) 获得. NA 型琼脂糖从 Pharmacia-LKB (Uppsala, Schweden) 获得. 所有其它的化学试剂来自 Haarmann & Reimer (Holzminden, 德国), E. Merck AG (Darmstadt, 德国), Fluka Chemie (Buchs, 瑞士), Serva Feinbiochemica (Heidelberg, 德国) 或 Sigma Chemie (Deisenhofen, 德国).

25 实施例

实施例 1

构建介导卡那霉素 (ΩKm) 或庆大霉素 (ΩGm) 抗性的 Ω 元件

- 为丁构建 ΩKm 元件, 以制备规模分离 2099 bp 的转座子 Tn5 的 BglI 片段 (E. A. Auerswald, G. Ludwig 和 H. Schaller, 1981, Cold Spring Harb. Symp. Quant. Biol. 45:107-113; E. Beck, G. Ludwig, E. A. Auerswald, B. Reiss 和 H. Schaller, 1982, 基因, 19:327-336; P. Mazodier, P. Cossart, E. Giraud 和 F. Gasser, 1985, 核酸研究,

13:195-205)。经过用 Bal 31 核酸酶处理将片段缩短到大约 990 bp。然后将这一仅含有卡那霉素抗性基因（编码氨基糖苷-3'-O-磷酸转移酶）的片段连接到 SmaI 切割的 pSKsym DNA（含对称结构的多克隆位点 [Sal I, Hind III, EcoRI, SmaI, EcoRI, Hind III, Sal I] 的 pBluescript SK-衍生物）上。可作为 SmaI 片段, EcoRI 片段, Hind III 片段或 SalI 片段从所得的质粒中重新分离出 Ω Km 元件。

为了构建 Ω Gm 元件, 在制备规模上分离出质粒 pBRIMCS-5 的 983 bp EaeI 片段 (M. E. Kovach, P. H. Elzer, D. S. Hill, G. T. Robertson, M. A. Farris, R. M. Roop 和 K. M. Peterson, 1995, 基因 166:175-176), 然后用绿豆核酸酶处理 (逐渐消化单链 DNA 分子末端)。然后将仅含庆大霉素抗性基因 (编码庆大霉素-3-乙酰转移酶) 的片段连接到 SmaI-裂解的 pSKsym DNA (见上文) 上。可作为 SmaI 片段, EcoRI 片段, HindIII 片段或 SalI 片段从所得质粒重新分离出 Ω Gm 元件。

15 实施例 2

从假单胞菌属种类 HR199 (DSM 7063) 克隆待插入 Ω 元件或缺失灭活的基因。

单独从大肠杆菌 S17-1 菌株 DSM 10439 和 DSM 10440 克隆 fcs, ech, vdh 和 aat 基因并使用质粒 pE207 和 pE5-1 (见 EP-A-0845532)。从这些质粒在制备规模上分离所述片段并按如下所述处理:

为了克隆 fcs 基因, 将来自质粒 pE207 的 2350 bp SalI/EcoRI 片段和来自质粒 pE5-1 的 3700 bp EcoRI/SalI 片段一起在 pBluescript SK- 中克隆使得 2 片段借助于 EcoRI 端连接在一起。从所得的杂种质粒在制备规模上分离 6050 bp SalI 片段并经过用 Bal 31 核酸酶处理缩短到大约 2480 bp。PstI 连接子随后连接到该片段的末端, 用 PstI 消化后, 将该片段克隆进 pBluescript SK-(pSKfcs)。转化大肠杆菌 XL1 blue 后, 所得的克隆表达 fcs 基因并表现出 0.2 U/mg 蛋白质的 Fcs 活性。

30 为了克隆 ech 基因, 在制备规模上从质粒 pE 207 分离 3800bp HindIII/EcoRI 片段并经过用 Bal 31 核酸酶处理缩短到大约 1470 bp。然后, 将 EcoRI 连接子连接到该片段的末端, 用 EcoRI 消化后,

将该片段克隆进 pBluescript SK⁻(pSKech).

为了克隆 vdh 基因, 在制备规模上从质粒 pE207 分离 2350 bp 的 SalI/EcoRI 片段. 克隆进 pBluescript SK⁻后, 使用外切核酸酶 III/绿豆核酸酶系统将该片段在一端截短大约 1530 bp. 然后将 EcoRI 连接子连接到该片段的末端, 用 EcoRI 消化后, 将该片段克隆进 pBluescript SK⁻(pSKvdh). 转化大肠杆菌 XL1 blue 后, 获得的克隆表达 VDH 基因且表现出 0.01 U/mg 蛋白质的 VDH 活性.

为了克隆 aat 基因, 在制备规模上从质粒 pE5-1 分离 3700 bp EcoRI/SalI 片段并经过用 Bal 31 核酸酶处理缩短到大约 1590bp. 然后将 EcoRI 连接子连接到该片段的末端, 用 EcoRI 消化后, 将该片段克隆进 pBluescript SK⁻(pSKaat).

实施例 3

经过插入 Ω 元件或经过缺失这些基因的部分区域灭活上述基因.

用 BssHII 消化含 fcs 基因的质粒 pSKfcs, 导致从 fcs 基因切割出 1290bp 的片段. 随后重新连接, 在 pBluescript SK⁻中以克隆的形式 (pSKfcs Δ) 获得 fcs 基因的缺失衍生物 (fcs Δ) (见图 1i 和 2i). 另外, 切除该片段后, 将 Ω 元件 Ω Km 和 Ω Gm 代替其连接进去. 这样产生了 fcs 基因的 Ω -灭活衍生物 (fcs Ω Km, 见图 1g 和 2g) 及 (fcs Ω Gm, 见图 1h 和 2h), 它们均在 pBluescript SK⁻中以克隆的形式获得 (pSKfcs Ω Km 和 pSKfcs Ω Gm). 在其杂交质粒具有经缺失或经 Ω 元件插入而失活的 fcs 基因的大肠杆菌克隆粗提物中不可能检测到任何 FCS 活性.

用 NruI 消化含 ech 基因的质粒 pSKeach, 导致从 ech 基因切除了 53 bp 的片段及 430 bp 的片段. 重新连接后, ech 基因的缺失衍生物 (ech Δ , 见图 11 和 21) 在 pBluescript SK⁻中以克隆的形式获得 (pSKeach Δ). 另外, 切除该片段后, 将 Ω 元件 Ω Km 和 Ω Gm 代替其连接进其中. 结果产生 ech 基因的 Ω -灭活衍生物 (ech Ω Km 和 ech Ω Gm), 它们在 pBluescript SK⁻中以克隆的形式获得 (pSKeach Ω Km 和 pSKeach Ω Gm).

用 BssHII 消化含 vdh 基因的质粒 pSKvdh, 导致从 vdh 基因切除了 210bp 的片段. 重新连接后, vdh 基因的缺失衍生物 (vdh Δ , 见图 10 和 20) 以克隆的形式在 pBluescript SK⁻(pSK vdh Δ) 中获得. 另外,

切除该片段后，将 Ω 元件 Ω Km和 Ω Gm代替其连接进入，结果产生 vdh 基因的 Ω -灭活衍生物(vdh Ω Km和 vdh Ω Gm)，在 pBluescript SK-中以克隆的形式获得 (pSKvdh Ω Km，见图 1m 和 2m) 及 (pSKvdh Ω Gm，见图 1n 和 2n)。在所得的大肠杆菌克隆的粗提物中不能检测到任何 VDH 活性，该克隆中的杂交质粒具有经缺失或 Ω 元件插入灭活的 vdh 基因。

用 BssHII 消化含 aat 基因的质粒 pSKaat，导致从 aat 基因中切除了 59bp 的片段。重新连接后，aat 基因的缺失衍生物(aat Δ ，见图 1r 和 2r)在 pBluescript SK-中以克隆的形式获得(pSKaat Δ)。另外，切除该片段后，将 Ω 元件 Ω Km和 Ω Gm代替其连接进入其中，结果产生 aat 基因的 Ω -灭活衍生物(aat Ω Km，见图 1p 和 2p)和(aat Ω Gm，见图 1q 和 2q)，在 pBluescript SK-中以克隆的形式获得(pSKaat Ω Km 和 pSKaat Ω Gm)。

15 实施例 4

将 Ω 元件灭活的基因亚克隆进可接合转移的“自杀质粒”pSUP202 中。

为了用 Ω -元件灭活的基因取代假单胞菌属种类 HR199 中的完整基因，需要一种载体，一方面该载体可转移进假单胞菌中(可接合转移的质粒)，另一方面它在这些细菌中不能复制，因此在假单胞菌中不稳定(“自杀质粒”)。使用这种质粒系统转移进假单胞菌中的 DNA 片段只有借助于同源重组(RecA-依赖性重组)整合进细菌细胞的基因组中才能保留。在这种情况下，使用“自杀质粒”pSUP202(Simon 等，1983，见 A. Pühler, 细菌-植物相互作用的分子遗传学, Springer Verlag, Berlin, Heidelberg, 纽约, p98-106)。

用 PstI 消化后，以质粒 pSKfcs Ω Km 和 pSKfcs Ω Gm 分离灭活的基因 fcs Ω Km 和 fcs Ω Gm 并连接到 PstI-裂解的 pSUP202DNA 上。将连接混合物转化进大肠杆菌 S17-1。在分别还含卡那霉素或庆大霉素的含四环素 LB 培养基中进行选择。获得其杂交质粒(pSUPfcs Ω Km)含灭活基因 fcs Ω Km 的卡那霉素抗性转化子。庆大霉素抗性转化子的相应杂交质粒(pSUPfcs Ω Gm)含灭活的基因 fcs Ω Gm。

EcoRI 消化后，从质粒 pSKech Ω Km 和 pSKech Ω Gm 分离灭活的基

因 ech Ω Km 和 ech Ω Gm 并连接到 EcoRI 裂解的 pSUP202 DNA 上。将连接混合物转化进大肠杆菌 S17-1。在分别还含卡那霉素或庆大霉素的含四环素 LB 培养基上进行选择。获得其杂交质粒 (pSUPech Ω Km) 含灭活基因 ech Ω Km 的卡那霉素抗性转化子。庆大霉素抗性转化子的相应杂交质粒 (pSUPech Ω Gm) 含灭活的基因 ech Ω Gm。

EcoRI 消化后, 从质粒 pSKvdh Ω Km 和 pSKvdh Ω Gm 分离灭活的基因 vdh Ω Km 和 vdh Ω Gm 并连接到 EcoRI-裂解的 pSUP202DNA 上。将连接混合物转化进大肠杆菌 S17-1。在分别还含卡那霉素或庆大霉素的含四环素 LB 培养基上进行选择。获得其杂交质粒 (pSUPvdh Ω Km) 含灭活基因 vdh Ω Km 的卡那霉素抗性转化子。庆大霉素抗性转化子的相应杂交质粒 (pSUPvdh Ω Gm) 含灭活基因 vdh Ω Gm。

EcoRI 消化后, 从质粒 pSKaat Ω Km 和 pSKaat Ω Gm 分离灭活基因 aat Ω Km 和 aat Ω Gm 并连接到 EcoRI 裂解的 pSUP202DNA 上。将连接混合物转化进大肠杆菌 S17-1 中。在分别还含卡那霉素或庆大霉素的含四环素 LB 培养基上进行选择。获得其杂交质粒 (pSUPaat Ω Km) 含灭活基因 aat Ω Km 的卡那霉素抗性转化子。庆大霉素抗性转化子的相应杂交质粒 (pSUPaat Ω Gm) 含灭活的基因 aat Ω Gm。

实施例 5

将缺失灭活的基因亚克隆进具有“sacB 选择系统”的可接合转移“自杀质粒”PHE55 中

为了用缺失灭活的基因取代假单胞菌属种类 HR199 中的完整基因, 需要具有在 pSUP202 的例子中已描述过的特性的载体。在缺失失活基因的情况下, 由于不存在选择假单胞菌属种类 HR199 中基因成功取代的可能性, 与 Ω 元件灭活基因相反, 使用了另一选择系统。在“sacB 选择系统”中, 取代的缺失失活基因被克隆进一种质粒中, 该质粒除抗生素抗性基因外还具有 sacB 基因。将该杂交质粒接合转移进假单胞菌后, 在完整基因在基因组中所处的位点处经同源重组整合质粒 (第一次交换)。这种产生的“杂基因型”菌株同时具有完整基因和缺失失活基因, 且被 pHE55 DNA 将这些基因互相分隔开。这些菌株表现出由载体编码的抗性且也具有活性 sacB 基因。随后的目的是借助于第二次同源重组事件将 pHE55 DNA 与完整基因一起分离出基因组 DNA (第二次交换)。这次重组事件产生的菌株仅具有失活的基因。

而且, pHE55 编码的抗生素抗性和 *sacB* 基因均已丢失. 如果将菌株在含蔗糖的培养基上划线培养, 见表达 *sacB* 基因的菌株的生长受抑制, 因为该基因产物将蔗糖转变成聚合物, 而该聚合物积累于细胞周质中. 由于发生了第二次重组事件而不再携带 *sacB* 基因的细胞的生长因此不受抑制. 为了具有在表型上选择缺失灭活基因整合的可能性, 该基因不与完整基因交换, 相反, 使用的菌株待替换的基因已通过插入 Ω 元件而被“标记”. 当发生成功替换时, 所得的菌株丧失由 Ω 元件编码的抗生素抗性.

用 PstI 消化后, 从质粒 pSKfcs Δ 分离失活基因 *fcs* Δ 并连接到 PstI-裂解的 pHE55 DNA 上. 将连接混合物转化进大肠杆菌 S17-1. 在含四环素的 LB 培养基上进行选择. 获得四环素抗性转化子, 其杂交质粒 (pHEfcs Δ) 含灭活基因 *fcs* Δ .

用 EcoRI 消化后, 从质粒 pSKech Δ 分离失活基因 *ech* Δ 并用绿豆核酸酶处理 (产生钝末端). 将该片段连接到 BamHI-裂解的且用绿豆核酸酶处理的 pHE55 DNA 上. 将连接混合物转化进大肠杆菌 S17-1. 在含四环素的 LB 培养基上进行选择. 获得四环素抗性转化子, 其杂交质粒 (pHEech Δ) 含有失活基因 *ech* Δ .

用 EcoRI 消化后, 从质粒 pSKvdh Δ 分离失活基因 *vdh* Δ 并用绿豆核酸酶处理. 将该片段连接到 BamHI 裂解的且用绿豆核酸酶处理过的 pHE55 DNA 上. 将连接混合物转化进大肠杆菌 S17-1. 在含四环素的 LB 培养基上进行选择. 获得四环素抗性转化子, 其杂交质粒 (pHEvdh Δ) 含有灭活的基因 *vdh* Δ .

用 EcoRI 消化后, 从质粒 pSKaat Δ 分离出灭活基因 *aat* Δ 并用绿豆核酸酶处理. 将该片段连接到 BamHI 裂解的并用绿豆核酸酶处理过的 pHE55 DNA 上. 将连接混合物转化进大肠杆菌 S17-1. 在含四环素的 LB 培养基上进行选择. 获得四环素抗性转化子, 其杂交质粒 (pHEaat Δ) 含有灭活的基因 *aat* Δ .

实施例 6

产生假单胞菌属种类 HR199 菌株的突变体, 其中经过插入 Ω -元件特异性灭活丁子香酚分解代谢基因.

在接合实验中菌株假单胞菌属种类 HR199 用作受体, 其中下述含

有 pSUP202 杂交质粒的菌株大肠杆菌 S17-1 用作供体。转接合子在含有相应于 Ω 元件之抗生素的含葡萄糖酸无机培养基上选择。可根据 pSUP202- 编码的四环素抗性区分“同基因子”（通过 2 次交换用 Ω 元件插入失活基因取代完整基因）和“杂基因子”（借助于单次交换将杂交质粒整合进基因组）转接合子。

5 将假单胞菌属种类 HR199 分别与大肠杆菌 S17-1 (pSUPfcs Ω Km) 和大肠杆菌 S17-1 (pSUPfcs Ω Gm) 接合后获得突变的假单胞菌属种类 HR199fcs Ω Km 和假单胞菌属种类 HR199fcs Ω Gm。经过 DNA 测序证实完整 fcs 基因被 Ω Km-灭活的或 Ω Gm-灭活的基因（分别为 fcs Ω Km 和 fcs Ω Gm）取代。

10 将假单胞菌属种类 HR199 分别与大肠杆菌 S17-1 (pSUPech Ω Km) 和大肠杆菌 S17-1 (pSUPech Ω Gm) 接合后获得突变的假单胞菌属种类 HR199ech Ω Km 假单胞菌属种类 HR199ech Ω Gm。经过 DNA 测序证实完整的 ech 基因被 Ω Km-灭活的或 Ω Gm-灭活的基因（分别为 ech Ω Km 和 ech Ω Gm）取代。

15 将假单胞菌属种类 HR199 分别与大肠杆菌 S17-1 (pSUPvdh Ω Km) 和大肠杆菌 S17-1 (pSUPvdh Ω Gm) 接合后获得突变的假单胞菌属种类 HR199 vdh Ω Km 和假单胞菌属种类 HR199 vdh Ω Gm。经过 DNA 测序证实完整的 vdh 基因被 Ω Km-灭活的或 Ω Gm-灭活的基因（分别为 vdh Ω Km 和 vdh Ω Gm）取代。

20 将假单胞菌属种类 HR199 分别与大肠杆菌 S17-1 (pSUP aat Ω Km) 和大肠杆菌 S17-1 (pSUP aat Ω Gm) 接合后获得突变的假单胞菌属种类 HR199 aat Ω Km 和假单胞菌属种类 HR199 aat Ω Gm。经过 DNA 测序证实完整的 aat 基因被 Ω Km-灭活的或 Ω Gm-灭活的基因（分别是 aat Ω Km 和 aat Ω Gm）取代。

25 将假单胞菌属种类 HR199 fcs Ω Km 与大肠杆菌 S17-1 (pSUP vdh Ω Gm) 接合后获得突变的假单胞菌属种类 HR199 fcs Ω Km vdh Ω Gm。经过 DNA 测序证实完整的 vdh 基因被 Ω Gm 灭活的基因 (vdh Ω Gm) 取代。

30 将假单胞菌属种类 HR199 vdh Ω Km 大肠杆菌 S17-1 (pSUP aat Ω Gm) 接合后获得突变的假单胞菌属种类 HR199 vdh Ω Kmaat Ω Gm。经过 DNA 测序证实完整的 aat 基因被 Ω Gm-灭活的基因 (aat Ω Gm) 取代。

将假单胞菌属种类 HR199 vdh Ω Km 与大肠杆菌 S17-1 (pSUP ech Ω Gm) 接合后获得突变的假单胞菌属种类 HR199 vdh Ω Kmech Ω Gm. 经过 DNA 测序证实完整的 ech 基因被 Ω Gm 灭活的基因 (ech Ω Gm) 取代.

5

实施例 7

产生菌株假单胞菌属种类 HR199 的突变体,

其中经过缺失部分区域特异性灭活丁子香酚分解代谢的基因.

在接合实验中, 菌株假单胞菌属种类 HR199 fcs Ω Km, 假单胞菌属种类 HR199 ech Ω Km, 假单胞菌属种类 HR199 vdh Ω Km 和假单胞菌属种类 HR199 aat Ω Km 用作受体, 下述含有 pHE55 杂交质粒的大肠杆菌 S17-1 菌株用作供体. 在除四环素 (pHE55 编码的抗性) 外还含有相应于 Ω 元件的抗生素的含葡萄糖酸无机培养基上选择“杂因子”转接合子. 在含蔗糖的无机培养基上划线培养后, 获得的转接合子经第二次重组事件 (第二次交换) 删除了载体 DNA. 经过在不含抗生素或含有相应于 Ω 元件的抗生素的无机培养基中划线培养, 可鉴定出用缺失灭活的基因 (无抗生素抗性) 取代了 Ω 元件灭活的基因的突变体.

将假单胞菌属种类 HR199fcs Ω Km 与大肠杆菌 S17-1 (pHEfcs Δ) 接合后获得突变的假单胞菌属种类 HR199 fcs Δ . 经 DNA 测序证实 Ω Km 灭活的基因 (fcs Ω Km) 被缺失灭活的基因 (fcs Δ) 取代.

将假单胞菌属种类 HR199 ech Ω Km 与大肠杆菌 S17-1 (pHEech Δ) 接合后获得突变的假单胞菌属种类 HR199 ech Δ . 经 DNA 测序证实 Ω Km 灭活的基因 (ech Ω Km) 被缺失灭活基因 (ech Δ) 取代.

将假单胞菌属种类 HR199 vdh Ω Km 与大肠杆菌 S17-1 (pHEvdh Δ) 接合后获得突变的假单胞菌属种类 HR199 vdh Δ . 经 DNA 测序证实 Ω Km 灭活的基因 (vdh Ω Km) 被缺失灭活的基因 (vdh Δ) 取代.

将假单胞菌属种类 HR199 aat Ω Km 与大肠杆菌 S17-1 (pHEaat Δ) 接合后获得突变的假单胞菌属种类 HR199 aat Δ . 经过 DNA 测序证实 Ω Km-灭活的基因 (aat Ω Km) 被缺失灭活的基因 (aat Δ) 取代.

30

实施例 8

使用突变的假单胞菌属种类 HR199 vdh Ω Km 将丁子香酚生物转化

成香草醛。

菌株假单胞菌属种类 HR199 vdh Ω Km 在 50ml 含 6mM 丁子香酚的 HR-MM 中繁殖到光密度为大约 OD_{600nm} = 0.6。17 小时后，在培养上清中可检测到 2.9mM 的香草醛，1.4mM 阿魏酸和 0.4mM 香草酸。

5

实施例 9

使用突变的假单胞菌属种类 HR199 vdh Ω Gmaat Ω Km 将丁子香酚生物转化成阿魏酸

10 菌株假单胞菌属种类 HR199 vdh Ω Gmaat Ω Km 在 50ml 含 6mM 丁子香酚的 HR-MM 中繁殖到光密度为大约 OD_{600nm} = 0.6。18 小时后，在培养上清中可检测到 1.9mM 香草醛，2.4mM 阿魏酸和 0.6mM 香草酸。

实施例 10

15 使用突变假单胞菌属种类 HR199 vdh Ω Gmaat Ω Km 将丁子香酚生物转化成松柏醇。

菌株假单胞菌属种类 HR199 vdh Ω Gmaat Ω Km 在 50ml 含 60mM 丁子香酚的 HR-MM 中繁殖到光密度大约为 OD_{600nm} = 0.4。15 小时后，在培养物上清中可检测到 1.7mM 松柏醇，1.4mM 香草醛，1.4mM 阿魏酸和 0.2mM 香草酸。

20

实施例 11

使用突变的假单胞菌属种类 HR199 vdh Ω Km 在 10 l 发酵罐中从丁子香酚发酵生产天然香草醛。

25 用 100ml 培养了 24 小时的预培养物接种生产发酵罐，其中该预培养物在调至 pH 7.0 且由 12.5g 甘油/l, 10g 酵母提取物/l 和 0.37g 乙酸/l 组成的培养基中在摇床 (120rpm) 上 32℃ 繁殖过。发酵罐含有 9.9 l 下列组成的培养基：1.5g 酵母提取物/l, 1.6g KH₂PO₄/l, 0.2g NaCl/l, 0.2g MgSO₄/l。用氢氧化钠溶液将 pH 调节至 pH7.0。灭菌后，向培养基中加入 4g 丁子香酚。温度为 32℃，通气为 3NL/分，
30 搅拌速度为 600rpm。用氢氧化钠溶液将 pH 维持在 pH6.5。

在接种后 4 小时，开始连续加入丁子香酚使得在 65 小时后发酵终止时，向培养物中加入了 255g 的丁子香酚。在发酵期间还加入 40g

酵母提取物。发酵终止时，丁子香酚的浓度为 0.2g/l。香草醛的含量为 2.6g/l。也存在 3.4g 的阿魏酸/l。

用诸如色谱，蒸馏和/或提取的已知物理方法分离以这种方式获得的香草醛并用于制备天然调味品。

5

附图的说明

图 1a 至 1r:

分离生物和突变体的基因结构

calA*: 松柏醇脱氢酶的部分失活基因

10

calB*: 松柏醛脱氢酶的部分失活基因

fcs*: 阿魏酰-CoA 合成酶的部分失活基因

ech*: 烯酰-CoA 水合酶-醛缩酶的部分失活基因

vdh*: 香草醛脱氢酶的部分失活基因

aat*: β -酮硫解酶的部分失活基因

15

尽管标记“*”的限制性酶裂解位点用于构建，但它们在所得的构建体上不再具有功能。

图 2a: $calA\Omega Km$ 基因结构的核苷酸序列

图 2b: $calA\Omega Gm$ 基因结构的核苷酸序列

图 2c: $calA\Delta$ 基因结构的核苷酸序列

图 2d: $calB\Omega Km$ 基因结构的核苷酸序列

5 图 2e: $calB\Omega Gm$ 基因结构的核苷酸序列

图 2f: $calB\Delta$ 基因结构的核苷酸序列

图 2g: $fcs\Omega Km$ 基因结构的核苷酸序列

图 2h: $fcs\Omega Gm$ 基因结构的核苷酸序列

图 2i: $fcs\Delta$ 基因结构的核苷酸序列

10 图 2j: $ech\Omega Km$ 基因结构的核苷酸序列

图 2k: $ech\Omega Gm$ 基因结构的核苷酸序列

图 2l: $ech\Delta$ 基因结构的核苷酸序列

图 2m: $vdh\Omega Km$ 基因结构的核苷酸序列

图 2n: $vdh\Omega Gm$ 基因结构的核苷酸序列

15 图 2o: $vdh\Delta$ 基因结构的核苷酸序列

图 2p: $aat\Omega Km$ 基因结构的核苷酸序列

图 2q: $aat\Omega Gm$ 基因结构的核苷酸序列

图 2r: $aat\Delta$ 基因结构的核苷酸序列

序列

CTGCAGCCAG GGCTGAAAAG GAGGGATTCA GTGAGGTCAT GAAGGGAGGG GACGGCGCCT	60
GGCTCCAATT GCTCGATGGC GCCGCGATTG AGTGTCTTGG GCGCGGTCTT GGAGAGTTCTG	120
GCTAGGGAGA TAAATTTGCT GGCCATGGTG GCGGCCCCCTG ATGGGTTGGA TGATTTTCTG	180
CATTCTGCAT CATGAAATTC ATGAAATCAT CACTTTTTCGG GGGGTGGGTG CACGGGATTG	240
AAGGTTGCTA GGAGAGTGCA TTGCTCGTAA GCCCAGGAAG CACGCGGGTT TCAGGATGGT	300
GCATGGAAAT GGCATGAGCT TTGCTGGATA TGATTAGAGA CATTAACAT TTTGGCGGAA	360
TGGAAGCACG ATTCCTCGCC CGGTAGAGCG GTAACCGCGA CATTCAGGAC CGTAAAAAGG	420
AAAGAGCATG CAA CTG ACC AAC AAG AAA ATC GTC GTC ACC GGA GTG TCC TCC	472
Met Gln Leu Thr Asn Lys Lys Ile Val Val Thr Gly Val Ser Ser	
1 5 10 15	
GGT ATC GGT GCC GAA ACT GCC CGC GTT CTG CGC TCT CAC GGC GCC ACA	520
Gly Ile Gly Ala Glu Thr Ala Arg Val Leu Arg Ser His Gly Ala Thr	
20 25 30	
GTG ATT GGC GTA GAT CGC AAC ATG CCG AGC CTG ACT CTG GAT GCT TTC	568
Val Ile Gly Val Asp Arg Asn Met Pro Ser Leu Thr Leu Asp Ala Phe	
35 40 45	
GTT CAG GCT GAC CTG AGC CAT CCT GAA GGC ATC GAT AAG GCC ATC GGG	616
Val Gln Ala Asp Leu Ser His Pro Glu Gly Ile Asp Lys Ala Ile	
50 55 60 62	
ACAGCAAGCG AACCAGGAATT GCCAGCTGGG GCGCCCTCTG GTAAGGTTGG GAAGCCCTGC	676
AAAGTAACT GGATGGCTTT CTTGCCGCCA AGGATCTGAT GGCGCAGGGG ATCAAGATCT	736
GATCAAGAGA CAGGATGAGG ATCGTTTCGC ATG ATT GAA CAA GAT GGA TTG CAC	790
Met Ile Glu Gln Asp Gly Leu His	
1 5	
GCA GGT TCT CCG GCC GCT TGG GTG GAG AGG CTA TTC GGC TAT GAC TGG	838
Ala Gly Ser Pro Ala Ala Trp Val Glu Arg Leu Phe Gly Tyr Asp Trp	
10 15 20	
GCA CAA CAG ACA ATC GGC TGC TCT GAT GCC GCC GTG TTC CGG CTG TCA	886
Ala Gln Gln Thr Ile Gly Cys Ser Asp Ala Ala Val Phe Arg Leu Ser	
25 30 35 40	
GCG CAG GGG CGC CCG GTT CTT TTT GTC AAG ACC GAC CTG TCC GGT GCC	934
Ala Gln Gly Arg Pro Val Leu Phe Val Lys Thr Asp Leu Ser Gly Ala	
45 50 55	
CTG AAT GAA CTG CAG GAC GAG GCA GCG CGG CTA TCG TGG CTG GCC ACG	982
Leu Asn Glu Leu Gln Asp Glu Ala Ala Arg Leu Ser Trp Leu Ala Thr	
60 65 70	

ACG GGC GTT CCT TGC GCA GCT GTG CTC GAC GTT GTC ACT GAA GCG GGA	1030
Thr Gly Val Pro Cys Ala Ala Val Leu Asp Val Val Thr Glu Ala Gly	
75 80 85	
AGG GAC TGG CTG CTA TTG GGC GAA GTG CCG GGG CAG GAT CTC CTG TCA	1078
Arg Asp Trp Leu Leu Leu Gly Glu Val Pro Gly Gln Asp Leu Leu Ser	
90 95 100	
TCT CAC CTT GCT CCT GCC GAG AAA GTA TCC ATC ATG GCT GAT GCA ATG	1126
Ser His Leu Ala Pro Ala Glu Lys Val Ser Ile Met Ala Asp Ala Met	
105 110 115 120	
CGG CGG CTG CAT ACG CTT GAT CCG GCT ACC TGC CCA TTC GAC CAC CAA	1174
Arg Arg Leu His Thr Leu Asp Pro Ala Thr Cys Pro Phe Asp His Gln	
125 130 135	
GCG AAA CAT CGC ATC GAG CGA GCA CGT ACT CGG ATG GAA GCC GGT CTT	1222
Ala Lys His Arg Ile Glu Arg Ala Arg Thr Arg Met Glu Ala Gly Leu	
140 145 150	
GTC GAT CAG GAT GAT CTG GAC GAA GAG CAT CAG GGG CTC GCG CCA GCC	1270
Val Asp Gln Asp Asp Leu Asp Glu Glu His Gln Gly Leu Ala Pro Ala	
155 160 165	
GAA CTG TTC GCC AGG CTC AAG GCG CGC ATG CCC GAC GGC GAG GAT CTC	1318
Glu Leu Phe Ala Arg Leu Lys Ala Arg Met Pro Asp Gly Glu Asp Leu	
170 175 180	
GTC GTG ACC CAT GGC GAT GCC TGC TTG CCG AAT ATC ATG GTG GAA AAT	1366
Val Val Thr His Gly Asp Ala Cys Leu Pro Asn Ile Met Val Glu Asn	
185 190 195 200	
GGC CGC TTT TCT GGA TTC ATC GAC TGT GGC CGG CTG GGT GTG GCG GAC	1414
Gly Arg Phe Ser Gly Phe Ile Asp Cys Gly Arg Leu Gly Val Ala Asp	
205 210 215	
CGC TAT CAG GAC ATA GCG TTG GCT ACC CGT GAT ATT GCT GAA GAG CTT	1462
Arg Tyr Gln Asp Ile Ala Leu Ala Thr Arg Asp Ile Ala Glu Glu Leu	
220 225 230	
GGC GGC GAA TGG GCT GAC CGC TTC CTC GTG CTT TAC GGT ATC GCC GCT	1510
Gly Gly Glu Trp Ala Asp Arg Phe Leu Val Leu Tyr Gly Ile Ala Ala	
235 240 245	
CCC GAT TCG CAG CGC ATC GCC TTC TAT CGC CTT CTT GAC GAG TTC TTC	1558
Pro Asp Ser Gln Arg Ile Ala Phe Tyr Arg Leu Leu Asp Glu Phe Phe	
250 255 260 264	
TGAGCGGGAC TCTGGGGTTC GAAATGACCG ACCAAGCGAC GCCCTG GCC GCG GTG	1613
Ala Ala Val	
225	
ATT GCA TTC ATG TGT GCT GAG GAG TCA CGT TGG ATC AAC GGC ATA AAT	1661
Ile Ala Phe Met Cys Ala Glu Glu Ser Arg Trp Ile Asn Gly Ile Asn	
230 235 240	

ATT CCA GTG GAC GGA GGT TTG GCA TCG ACC TAC GTG TAA GTTCGTGGAC	1710
Ile Pro Val Asp Gly Gly Leu Ala Ser Thr Tyr Val	
245 250 255	
GCCTTTTGCA CGCGCACTAT ATCTCTATGC AGCAGCTGAA AGCAGCTTTG GTTTTGATCG	1770
GAGGTAGCGG GCGGAAAGGT GCAGAATGTC TAAATAATAA AGGATTCTTG TGAAGCTTTA	1830
GTGTGTCGTA AACGAAAATA AAAATAAAGA GGAATGATAT GAAAGCAAGT AGATCAGTCT	1890
GCACTTTCAA AATAGCTACC CTGGCAGGCG CCATTTATGC AGCGCTGCCA ATGTCAGCTG	1950
CAAACCTCGAT GCAGCTGGAT GTAGGTAGCT CGGATTGGAC GGTGCGTTGG GGACAACACC	2010
CTCAAGTATA GCCTTGCCCTC TCGCCTGAAT GAGCAAGACT CAAGTCTGAC AAATGCGCCG	2070
ACTGTCAATG GTTATATCCG GATATTCAAA GTCAGGGTGA TCGTAACTTT GACCGGGGGC	2130
TTGGTATCCA ATCGTCTCGA TATTCTGGCT GCAG	2164

图 2a:

CTGCAGCCAG GGCTGAAAAG GAGGGATTCA GTGAGGTCAT GAAGGGAGGG GACGGCGCCT 60
 GGCTCCAATT GCTCGATGGC GCCGCGATTG AGTGTCTTGG GCGCGGTCTT GGAGAGTTCTG 120
 GCTAGGGAGA TAAATTTGCT GGCCATGGTG GCGGCCCCCTG ATGGGTTGGA TGATTTTCTG 180
 CATTCTGCAT CATGAAATTC ATGAAATCAT CACTTTTCGG GGGGTGGGTG CACGGGATTG 240
 AAGGTTGCTA GGAGAGTGCA TTGCTCGTAA GCCCAGGAAG CACGCGGGTT TCAGGATGGT 300
 GCATGGAAAT GGCATGAGCT TTGCTGGATA TGATTAGAGA CATTAACAT TTTGGCGGAA 360
 TGGAAGCACG ATTCCTCGCC CGGTAGAGCG GTAACCGCGA CATTCAGGAC CGTAAAAAGG 420
 AAAGAGCATG CAA CTG ACC AAC AAG AAA ATC GTC GTC ACC GGA GTG TCC TCC 472
 Met Gln Leu Thr Asn Lys Lys Ile Val Val Thr Gly Val Ser Ser
 1 5 10 15
 GGT ATC GGT GCC GAA ACT GCC CGC GTT CTG CGC TCT CAC GGC GCC ACA 520
 Gly Ile Gly Ala Glu Thr Ala Arg Val Leu Arg Ser His Gly Ala Thr
 20 25 30
 GTG ATT GGC GTA GAT CGC AAC ATG CCG AGC CTG ACT CTG GAT GCT TTC 568
 Val Ile Gly Val Asp Arg Asn Met Pro Ser Leu Thr Leu Asp Ala Phe
 35 40 45
 GTT CAG GCT GAC CTG AGC CAT CCT GAGGGGAGAG GCGGTTTGCG TATTGGGCGC 622
 Val Gln Ala Asp Leu Ser His Pro
 50 55
 ATGCATAAAA ACTGTTGTAA TTCATTAAGC ATTCTGCCGA CATGGAAGCC ATCACAAAACG 682
 GCATGATGAA CCTGAATCGC CAGCGGCATC AGCACCTTGT CGCCTTGCGT ATAATATTTG 742
 CCCATGGACG CACACCGTGG AAACGGATGA AGGCACGAAC CCAGTTGACA TAAGCCTGTT 802
 CGGTTTCGTAA ACTGTAATGC AAGTAGCGTA TGCCTCACG CAACTGGTCC AGAACCTTGA 862
 CCGAACGCAG CGGTGGTAAC GGCGCAGTGG CGGTTTTTCAT GGCTTGTTAT GACTGTTTTT 922
 TTGTACAGTC TATGCCTCGG GCATCCAAGC AGCAAGCGCG TTACGCCGTG GGTGATGTT 982
 TGATGTTATG GAGCAGCAAC G ATG TTA CGC AGC AGC AAC GAT GTT ACG CAG 1033
 Met Leu Arg Ser Ser Asn Asp Val Thr Gln
 1 5 10
 CAG GGC AGT CGC CCT AAA ACA AAG TTA GGT GGC TCA AGT ATG GGC ATC 1081
 Gln Gly Ser Arg Pro Lys Thr Lys Leu Gly Gly Ser Ser Met Gly Ile
 15 20 25
 ATT CGC ACA TGT AGG CTC GGC CCT GAC CAA GTC AAA TCC ATG CGG GCT 1129
 Ile Arg Thr Cys Arg Leu Gly Pro Asp Gln Val Lys Ser Met Arg Ala
 30 35 40
 GCT CTT GAT CTT TTC GGT CGT GAG TTC GGA GAC GTA GCC ACC TAC TCC 1177
 Ala Leu Asp Leu Phe Gly Arg Glu Phe Gly Asp Val Ala Thr Tyr Ser
 45 50 55

CAA CAT CAG CCG GAC TCC GAT TAC CTC GGG AAC TTG CTC CGT AGT AAG Gln His Gln Pro Asp Ser Asp Tyr Leu Gly Asn Leu Leu Arg Ser Lys 60 65 70	1225
ACA TTC ATC GCG CTT GCT GCC TTC GAC CAA GAA GCG GTT GTT GGC GCT Thr Phe Ile Ala Leu Ala Ala Phe Asp Gln Glu Ala Val Val Gly Ala 75 80 85 90	1273
CTC GCG GCT TAC GTT CTG CCC AGG TTT GAG CAG CCG CGT AGT GAG ATC Leu Ala Ala Tyr Val Leu Pro Arg Phe Glu Gln Pro Arg Ser Glu Ile 95 100 105	1321
TAT ATC TAT GAT CTC GCA GTC TCC GGC GAG CAC CGG AGG CAG GGC ATT Tyr Ile Tyr Asp Leu Ala Val Ser Gly Glu His Arg Arg Gln Gly Ile 110 115 120	1369
GCC ACC GCG CTC ATC AAT CTC CTC AAG CAT GAG GCC AAC GCG CTT GGT Ala Thr Ala Leu Ile Asn Leu Leu Lys His Glu Ala Asn Ala Leu Gly 125 130 135	1417
GCT TAT GTG ATC TAC GTG CAA GCA GAT TAC GGT GAC GAT CCC GCA GTG Ala Tyr Val Ile Tyr Val Gln Ala Asp Tyr Gly Asp Asp Pro Ala Val 140 145 150	1465
GCT CTC TAT ACA AAG TTG GGC ATA CGG GAA GAA GTG ATG CAC TTT GAT Ala Leu Tyr Thr Lys Leu Gly Ile Arg Glu Glu Val Met His Phe Asp 155 160 165 170	1513
ATC GAC CCA AGT ACC GCC ACC TAA CAATTCGTTC AAGCCGAGAT CGGCTTCCCT Ile Asp Pro Ser Thr Ala Thr 175 177	1567
G ATT GCA TTC ATG TGT GCT GAG GAG TCA CGT TGG ATC AAC GGC ATA AAT Ile Ala Phe Met Cys Ala Glu Glu Ser Arg Trp Ile Asn Gly Ile Asn 228 230 235 240	1616
ATT CCA GTG GAC GGA GGT TTG GCA TCG ACC TAC GTG TAA GTTCGTGGAC Ile Pro Val Asp Gly Gly Leu Ala Ser Thr Tyr Val 245 250 255	1665
GCCCTTTGCA CGCGCACTAT ATCTCTATGC AGCAGCTGAA AGCAGCTTTG GTTTTGATCG	1725
GAGGTAGCGG GCGGAAAGGT GCAGAATGTC TAAATAATAA AGGATTCTTG TGAAGCTTTA	1785
GTGTGCCGTA AACGAAAATA AAAATAAAGA GGAATGATAT GAAAGCAAGT AGATCAGTCT	1845
GCACTTTCAA AATAGCTACC CTGGCAGGCG CCATTTATGC AGCGCTGCCA ATGTCAGCTG	1905
CAAACCTCGAT GCAGCTGGAT GTAGGTAGCT CGGATTGGAC GGTGCGTTGG GGACAACACC	1965
CTCAAGTATA GCCTTGCCCTC TCGCCTGAAT GAGCAAGACT CAAGTCTGAC AAATGCGCCG	2025
ACTGTCAATG GTTATATCCG GATATTCAAA GTCAGGGTGA TCGTAACTTT GACCGGGGGC	2085
TTGGTATCCA ATCGTCTCGA TATTCTGGCT GCAG	2119

图 2b:

CTGCAGCCAG GGCTGAAAAG GAGGGATTCA GTGAGGTCAT GAAGGGAGGG GACGGCGCCT	60
GGCTCCAATT GCTCGATGGC GCCGCGATTG AGTGTCTTGG GCGCGGTCTT GGAGAGTTCTG	120
GCTAGGGAGA TAAATTTGCT GGCCATGGTG GCGGCCCCCTG ATGGGTTGGA TGATTTTCTG	180
CATTCTGCAT CATGAAATTC ATGAAATCAT CACTTTTTCGG GGGGTGGGTG CACGGGATTG	240
AAGGTTGCTA GGAGAGTGCA TTGCTCGTAA GCCCAGGAAG CACGCGGGTT TCAGGATGGT	300
GCATGGAAAT GGCATGAGCT TTGCTGGATA TGATTAGAGA CATTAACTAT TTTGGCGGAA	360
TGGAAGCACG ATTCCTCGCC CGGTAGAGCG GTAACCGCGA CATTCAAGGAC CGTAAAAAGG	420
AAAGAGCATG CAA CTG ACC AAC AAG AAA ATC GTC GTC ACC GGA GTG TCC TCC	472
Met Gln Leu Thr Asn Lys Lys Ile Val Val Thr Gly Val Ser Ser	
1 5 10 15	
GGT ATC GGT GCC GAA ACT GCC CGC GTT CTG CGC TCT CAC GGC GCC ACA	520
Gly Ile Gly Ala Glu Thr Ala Arg Val Leu Arg Ser His Gly Ala Thr	
20 25 30	
GTG ATT GGC GTA GAT CGC AAC ATG CCG AGC CTG ACT CTG GAT GCT TTC	568
Val Ile Gly Val Asp Arg Asn Met Pro Ser Leu Thr Leu Asp Ala Phe	
35 40 45	
GTT CAG GCT GAC CTG AGC CAT CCT GAA GGC ATC GATC AAC GGC ATA AAT	617
Val Gln Ala Asp Leu Ser His Pro Glu Gly Ile Asn Gly Ile Asn	
50 55 58 240	
ATT CCA GTG GAC GGA GGT TTG GCA TCG ACC TAC GTG TAA GTTCGTGGAC	666
Ile Pro Val Asp Gly Gly Leu Ala Ser Thr Tyr Val	
245 250 255	
GCCCTTTGCA CGCGCACTAT ATCTCTATGC AGCAGCTGAA AGCAGCTTTG GTTTTGATCG	726
GAGGTAGCGG GCGGAAAGGT GCAGAATGTC TAAATAATAA AGGATTCTTG TGAAGCTTTA	786
GTTGTCCGTA AACGAAAATA AAAATAAAGA GGAATGATAT GAAAGCAAGT AGATCAGTCT	846
GCACTTTCAA AATAGCTACC CTGGCAGGCG CCATTTATGC AGCGCTGCCA ATGTCAGCTG	906
CAAACTCGAT GCAGCTGGAT GTAGGTAGCT CGGATTGGAC GGTGCGTTGG GGACAACACC	966
CTCAAGTATA GCCTTGCCCTC TCGCCTGAAT GAGCAAGACT CAAGTCTGAC AAATGCGCCG	1026
ACTGTCAATG GTTATATCCG GATATTCAAA GTCAGGGTGA TCGTAACTTT GACCGGGGGG	1086
TTGGTATCCA ATCGTCTCGA TATTCTGGCT GCAG	1120

图 2c:

GAATTCGCG TATCGCCCGG TTCTATCAGC GGGCCGCTTT CGAAAGTCAT GGTGTTAGCC	60
GGTAGGGTCT TTTTCTTGGC CATGCTTGTT GCCTGAACCT TCGTTGACAT AGGGCAGAGG	120
TGCGTTTGCC GCTTCGCTTC GCGATGAACC GCATCGAGAT GCTGAGGTCA GGATTTTTC	180
TTAACTCGCG TAAGCATTCT GTCATTTTTT TGGTGGCTTT GAACAGCCTG ATGAAAGGTG	240
GTCTCGCCCT TTGAGGCCGA TTCTTGGGCG CTTGGCGGCG TCGAAGCGAT GCTCCACTAC	300
CGATTAAGAT AATTAAATA AGGAAACCGC ATGGTTTCTT ATGTGAATTT GTCTGGCATA	360
CTCCAGCTCA AGGGCAATTT TTGGGCTATT GGCTGAGCAG TTGCCTCTAT ATGGTTATTC	420
AGAATAACAA TTGACTCCTC AGGAGGTCAG CG ATG AGC ATT CTT GGT TTG AAT	473
Met Ser Ile Leu Gly Leu Asn	
1 5	
GGT GCC CCG GTC GGA GCT GAG CAG CTG GGC TCG GCT CTT GAT CGC ATG	521
Gly Ala Pro Val Gly Ala Glu Gln Leu Gly Ser Ala Leu Asp Arg Met	
10 15 20	
AAG AAG GCG CAC CTG GAG CAG GGG CCT GCA AAC TTG GAG CTG CGT CTG	569
Lys Lys Ala His Leu Glu Gln Gly Pro Ala Asn Leu Glu Leu Arg Leu	
25 30 35	
AGT AGG CTG GAT CGT GCG ATT GCA ATG CTT CTG GAA AAT CGT GAA GCA	617
Ser Arg Leu Asp Arg Ala Ile Ala Met Leu Leu Glu Asn Arg Glu Ala	
40 45 50 55	
ATT GCC GAC GCG GTT TCT GCT GAC TTT GGC AAT CGC AGC CGT GAG CAA	665
Ile Ala Asp Ala Val Ser Ala Asp Phe Gly Asn Arg Ser Arg Glu Gln	
60 65 70	
ACA CTG CTT TGC GAC ATT GCT GGC TCG GTG GCA AGC CTG AAG GAT AGC	713
Thr Leu Leu Cys Asp Ile Ala Gly Ser Val Ala Ser Leu Lys Asp Ser	
75 80 85	
CGC GAG CAC GTG GCC AAA TGG ATG GAG CCC GAA CAT CAC AAG GCG ATG	761
Arg Glu His Val Ala Lys Trp Met Glu Pro Glu His His Lys Ala Met	
90 95 100	
TTT CCA GGG GCG GAG GCA CGC GTT GAG TTT CAG CCG CTG GGT GTC GTT	809
Phe Pro Gly Ala Glu Ala Arg Val Glu Phe Gln Pro Leu Gly Val Val	
105 110 115	
GGG GTC ATT AGT CCC TGG AAC TTC CCT ATC GTA CTG GCC TTT GGG CCG	857
Gly Val Ile Ser Pro Trp Asn Phe Pro Ile Val Leu Ala Phe Gly Pro	
120 125 130 135	
CTG GCC GGC ATA TTC GCA GCA GGT AAT CGC GCC ATG CTC AAG CCG TCC	905
Leu Ala Gly Ile Phe Ala Ala Gly Asn Arg Ala Met Leu Lys Pro Ser	
140 145 150	
GAG CTT ACC CCG CGG ACT TCT GCC CTG CTT GCG GAG CTA ATT GCT CGT	953
Glu Leu Thr Pro Arg Thr Ser Ala Leu Leu Ala Glu Leu Ile Ala Arg	
155 160 165	

TAC TTC GAT GAA ACT GAG CTG ACT ACA GTG CTG GGC GAC GCT GAA GTC	1001
Tyr Phe Asp Glu Thr Glu Leu Thr Thr Val Leu Gly Asp Ala Glu Val	
170 175 180	
GGT GCG CTG TTC AGT GCT CAG CCT TTC GAT CAT CTG ATC TTC ACC GGC	1049
Gly Ala Leu Phe Ser Ala Gln Pro Phe Asp His Leu Ile Phe Thr Gly	
185 190 195	
GGC ACT GCC GTG GCC AAG CAC ATC ATG CGT GCC GCG GCG GAT AAC CTA	1097
Gly Thr Ala Val Ala Lys His Ile Met Arg Ala Ala Ala Asp Asn Leu	
200 205 210 215	
GTG CCC GTT ACC CTG GAA TTG GGT GGC AAA TCG CCG GTG ATC GTT TCC	1145
Val Pro Val Thr Leu Glu Leu Gly Gly Lys Ser Pro Val Ile Val Ser	
220 225 230	
CGC AGT GCA GAT ATG GCG GAC GTT GCA CAA CGG GTG TTG ACG GTG AAA	1193
Arg Ser Ala Asp Met Ala Asp Val Ala Gln Arg Val Leu Thr Val Lys	
235 240 245	
ACC TTC AAT GCC GGG CAA ATC TGT CTG GCA CCG GAC TAT GTG CTG CTG	1241
Thr Phe Asn Ala Gly Gln Ile Cys Leu Ala Pro Asp Tyr Val Leu Leu	
250 255 260	
CCG GAA GGGACAGCAA GCGAACCGGA ATTGCCAGCT GGGGCGCCCT CTGGTAAGGT	1297
Pro Glu	
265	
TGGGAAGCCC TGCAAAGTAA ACTGGATGGC TTTCTTGCCG CCAAGGATCT GATGGCGCAG	1357
GGGATCAAGA TCTGATCAAG AGACAGGATG AGGATCGTTT CGC ATG ATT GAA CAA	1412
Met Ile Glu Gln	
1	
GAT GGA TTG CAC GCA GGT TCT CCG GCC GCT TGG GTG GAG AGG CTA TTC	1460
Asp Gly Leu His Ala Gly Ser Pro Ala Ala Trp Val Glu Arg Leu Phe	
5 10 15 20	
GGC TAT GAC TGG GCA CAA CAG ACA ATC GGC TGC TCT GAT GCC GCC GTG	1508
Gly Tyr Asp Trp Ala Gln Gln Thr Ile Gly Cys Ser Asp Ala Ala Val	
25 30 35	
TTC CGG CTG TCA GCG CAG GGG CGC CCG GTT CTT TTT GTC AAG ACC GAC	1556
Phe Arg Leu Ser Ala Gln Gly Arg Pro Val Leu Phe Val Lys Thr Asp	
40 45 50	
CTG TCC GGT GCC CTG AAT GAA CTG CAG GAC GAG GCA GCG CCG CTA TCG	1604
Leu Ser Gly Ala Leu Asn Glu Leu Gln Asp Glu Ala Ala Arg Leu Ser	
55 60 65	
TGG CTG GCC ACG ACG GGC GTT CCT TGC GCA GCT GTG CTC GAC GTT GTC	1652
Trp Leu Ala Thr Thr Gly Val Pro Cys Ala Ala Val Leu Asp Val Val	
70 75 80	

ACT GAA GCG GGA AGG GAC TGG CTG CTA TTG GGC GAA GTG CCG GGG CAG Thr Glu Ala Gly Arg Asp Trp Leu Leu Leu Gly Glu Val Pro Gly Gln 85 90 95 100	1700
GAT CTC CTG TCA TCT CAC CTT GCT CCT GCC GAG AAA GTA TCC ATC ATG Asp Leu Leu Ser Ser His Leu Ala Pro Ala Glu Lys Val Ser Ile Met 105 110 115	1748
GCT GAT GCA ATG CGG CGG CTG CAT ACG CTT GAT CCG GCT ACC TGC CCA Ala Asp Ala Met Arg Arg Leu His Thr Leu Asp Pro Ala Thr Cys Pro 120 125 130	1796
TTC GAC CAC CAA GCG AAA CAT CGC ATC GAG CGA GCA CGT ACT CGG ATG Phe Asp His Gln Ala Lys His Arg Ile Glu Arg Ala Arg Thr Arg Met 135 140 145	1844
GAA GCC GGT CTT GTC GAT CAG GAT GAT CTG GAC GAA GAG CAT CAG GGG Glu Ala Gly Leu Val Asp Gln Asp Asp Leu Asp Glu Glu His Gln Gly 150 155 160	1892
CTC GCG CCA GCC GAA CTG TTC GCC AGG CTC AAG GCG CGC ATG CCC GAC Leu Ala Pro Ala Glu Leu Phe Ala Arg Leu Lys Ala Arg Met Pro Asp 165 170 175 180	1940
GGC GAG GAT CTC GTC GTG ACC CAT GGC GAT GCC TGC TTG CCG AAT ATC Gly Glu Asp Leu Val Val Thr His Gly Asp Ala Cys Leu Pro Asn Ile 185 190 195	1988
ATG GTG GAA AAT GGC CGC TTT TCT GGA TTC ATC GAC TGT GGC CGG CTG Met Val Glu Asn Gly Arg Phe Ser Gly Phe Ile Asp Cys Gly Arg Leu 200 205 210	2036
GGT GTG GCG GAC CGC TAT CAG GAC ATA GCG TTG GCT ACC CGT GAT ATT Gly Val Ala Asp Arg Tyr Gln Asp Ile Ala Leu Ala Thr Arg Asp Ile 215 220 225	2084
GCT GAA GAG CTT GGC GGC GAA TGG GCT GAC CGC TTC CTC GTG CTT TAC Ala Glu Glu Leu Gly Gly Glu Trp Ala Asp Arg Phe Leu Val Leu Tyr 230 235 240	2132
GGT ATC GCC GCT CCC GAT TCG CAG CGC ATC GCC TTC TAT CGC CTT CTT Gly Ile Ala Ala Pro Asp Ser Gln Arg Ile Ala Phe Tyr Arg Leu Leu 245 250 255 260	2180
GAC GAG TTC TTC TGA GCGGGACTCT GGGGTTTCGAA ATGACCGACC AAGCGACGCC Asp Glu Phe Phe 264	2235
CGC CAT GCC AAG CCT GTT CTC GTG CAA AGT CCT GTG GGT GAG TCG AAC His Ala Lys Pro Val Leu Val Gln Ser Pro Val Gly Glu Ser Asn 444 445 450 455	2283
TTG GCG ATG CGC GCA CCC TAC GGA GAA GCG ATC CAC GGA CTG CTC TCT Leu Ala Met Arg Ala Pro Tyr Gly Glu Ala Ile His Gly Leu Leu Ser 460 465 470	2331

GTC CTC CTT TCA ACG GAG TGT TAG AACCGTTGGT AGTGGTTTTG GACGGGCCCA	2385
Val Leu Leu Ser Thr Glu Cys	
475 480 481	
GGAGCATGCG CTTCTGGGCC CGTTTCCTTGA GTATTCATTG GATAGTCACG CGTGGTAGCT	2445
TCGAGCCTGC ACAGCTGATG AGCACCCCTGG AAGGCGCGCT GTACGCGGAC GACTGGGTTC	2505
ATCTTCGCCA TTCATGACGG AACTCCGTTT CCCAGTACCG CGATGACTAT TTTGCCTCTT	2565
CCGATGTCCG ATTCCACGCC GCCTGACGCT AAGCGGGGGC GGGGGCGCCC GCATCCCAGC	2625
CCAGACAGCA ACAAATGAGT AGGCTCTTGG ATGCCGCGGC GGCTGAGATT GGTAACGGCA	2685
ATTTTCGTCAA TGTGACGATG GATTCGATTG CCCGTGCTGC CGGCGTCTCA AAAAAACGC	2745
TGTACGTCTT GGTGGCGAGC AAGGAAGAAC TCATTTCCTG GTTAGTGGCT CGAGACATGT	2805
CCAACCTTGA GGAATTC	2822

2d:

GAATTC	CCGCG	TATCG	CCCCG	TTCTAT	CAGC	GGGCC	GCTTT	CGAAAG	TCAT	GGTGT	TAGCC					60
GGTAGG	GTCT	TTTTCT	TGGC	CATGCT	TGTT	GCCTGA	AACCT	TCGTTG	ACAT	AGGGC	CAGAGG					120
TGCGTT	TGCC	GCTTCG	CCTC	GCGATG	AACC	GCATCG	GAGAT	GCTGAG	GTCA	GGATTT	TTCC					180
TTAACT	CGCG	TAAGCAT	TCT	GTCAT	TTTTTT	TGGTGG	CCTTT	GAACAG	CCCTG	ATGAA	AGGTG					240
GTCTCG	CCCT	TTGAGG	CCGA	TTCTTG	GGCG	CTTGGC	GGCG	TCGAAG	CGAT	GCTCC	ACTAC					300
CGATTA	AGAT	AATTAAA	ATA	AGGAA	ACCGC	ATGGTT	TCTT	ATGTGA	AATTT	GTCTGG	CATA					360
CTCCAG	CTCA	AGGGCA	AATTT	TTGGG	CTATT	GGCTG	AGCAG	TTGCCT	CTAT	ATGGT	TATTC					420
AGAATA	ACAA	TTGACT	CCTC	AGGAG	GTCAG	CG	ATG	AGC	ATT	CTT	GGT	TTG	AAT			473
							Met	Ser	Ile	Leu	Gly	Leu	Asn			
							1				5					
GGT	GCC	CCG	GTC	GGA	GCT	GAG	CAG	CTG	GGC	TCG	GCT	CTT	GAT	CGC	ATG	521
Gly	Ala	Pro	Val	Gly	Ala	Glu	Gln	Leu	Gly	Ser	Ala	Leu	Asp	Arg	Met	
		10					15					20				
AAG	AAG	GCG	CAC	CTG	GAG	CAG	GGG	CCT	GCA	AAC	TTG	GAG	CTG	CGT	CTG	569
Lys	Lys	Ala	His	Leu	Glu	Gln	Gly	Pro	Ala	Asn	Leu	Glu	Leu	Arg	Leu	
		25				30					35					
AGT	AGG	CTG	GAT	CGT	GCG	ATT	GCA	ATG	CTT	CTG	GAA	AAT	CGT	GAA	GCA	617
Ser	Arg	Leu	Asp	Arg	Ala	Ile	Ala	Met	Leu	Leu	Glu	Asn	Arg	Glu	Ala	
		40				45				50					55	
ATT	GCC	GAC	GCG	GTT	TCT	GCT	GAC	TTT	GGC	AAT	CGC	AGC	CGT	GAG	CAA	665
Ile	Ala	Asp	Ala	Val	Ser	Ala	Asp	Phe	Gly	Asn	Arg	Ser	Arg	Glu	Gln	
				60					65					70		
ACA	CTG	CTT	TGC	GAC	ATT	GCT	GGC	TCG	GTG	GCA	AGC	CTG	AAG	GAT	AGC	713
Thr	Leu	Leu	Cys	Asp	Ile	Ala	Gly	Ser	Val	Ala	Ser	Leu	Lys	Asp	Ser	
			75					80					85			
CGC	GAG	CAC	GTG	GCC	AAA	TGG	ATG	GAG	CCC	GAA	CAT	CAC	AAG	GCG	ATG	761
Arg	Glu	His	Val	Ala	Lys	Trp	Met	Glu	Pro	Glu	His	His	Lys	Ala	Met	
		90					95					100				
TTT	CCA	GGG	GCG	GAG	GCA	CGC	GTT	GAG	TTT	CAG	CCG	CTG	GGT	GTC	GTT	809
Phe	Pro	Gly	Ala	Glu	Ala	Arg	Val	Glu	Phe	Gln	Pro	Leu	Gly	Val	Val	
		105				110					115					
GGG	GTC	ATT	AGT	CCC	TGG	AAC	TTC	CCT	ATC	GTA	CTG	GCC	TTT	GGG	CCG	857
Gly	Val	Ile	Ser	Pro	Trp	Asn	Phe	Pro	Ile	Val	Leu	Ala	Phe	Gly	Pro	
120					125					130					135	
CTG	GCC	GGC	ATA	TTC	GCA	GCA	GGT	AAT	CGC	GCC	ATG	CTC	AAG	CCG	TCC	905
Leu	Ala	Gly	Ile	Phe	Ala	Ala	Gly	Asn	Arg	Ala	Met	Leu	Lys	Pro	Ser	
				140					145					150		
GAG	CTT	ACC	CCG	CGG	ACT	TCT	GCC	CTG	CTT	GCG	GAG	CTA	ATT	GCT	CGT	953
Glu	Leu	Thr	Pro</													

TAC TTC GAT GAA ACT GAG CTG ACT ACA GTG CTG GGC GAC GCT GAA GTC	1001
Tyr Phe Asp Glu Thr Glu Leu Thr Thr Val Leu Gly Asp Ala Glu Val	
170 175 180	
GGT GCG CTG TTC AGT GCT CAG CCT TTC GAT CAT CTG ATC TTC ACC GGC	1049
Gly Ala Leu Phe Ser Ala Gln Pro Phe Asp His Leu Ile Phe Thr Gly	
185 190 195	
GGC ACT GCC GTG GCC AAG CAC ATC ATG CGT GCC GCG GCG GAT AAC CTA	1097
Gly Thr Ala Val Ala Lys His Ile Met Arg Ala Ala Ala Asp Asn Leu	
200 205 210 215	
GTG CCC GTT ACC CTG GAA TTG GGT GGC AAA TCG CCG GTG ATC GTT TCC	1145
Val Pro Val Thr Leu Glu Leu Gly Gly Lys Ser Pro Val Ile Val Ser	
220 225 230	
CGC AGT GCA GAT ATG GCG GAC GTT GCA CAA CGG GTG TTG ACG GTG AAA	1193
Arg Ser Ala Asp Met Ala Asp Val Ala Gln Arg Val Leu Thr Val Lys	
235 240 245	
ACC TTC AAT GCC GGG CAA ATC TGT CTG GCA CCG GAC TAT GTG CTG GGG	1241
Thr Phe Asn Ala Gly Gln Ile Cys Leu Ala Pro Asp Tyr Val Leu	
250 255 260 262	
GAGAGGCGGT TTGCGTATTG GGCGCATGCA TAAAACTGT TGTAATTCAT TAAGCATTCT	1301
GGCGACATGG AAGCCATCAC AAACGGCATG ATGAACCTGA ATCGCCAGCG GCATCAGCAC	1361
CTTGTCGCCT TGCCTATAAT ATTTGCCCAT GGACGCACAC CGTGGAAACG GATGAAGGCA	1421
CGAACCCAGT TGACATAAGC CTGTTCCGTT CGTAAACTGT AATGCAAGTA GCGTATGCGC	1481
TCACGCAACT GGTCCAGAAC CTTGACCGAA CGCAGCGGTG GTAACGCCGC AGTGGCGGTT	1541
TTCATGGCTT GTTATGACTG TTTTFTTGTA CAGTCTATGC CTCGGGCATC CAAGCAGCAA	1601
GCGCGTTACG CCGTGGGTCG ATGTTTGATG TTATGGAGCA GCAACG ATG TTA CGC	1656
Met Leu Arg	
1	
AGC AGC AAC GAT GTT ACG CAG CAG GGC AGT CGC CCT AAA ACA AAG TTA	1704
Ser Ser Asn Asp Val Thr Gln Gln Gly Ser Arg Pro Lys Thr Lys Leu	
5 10 15	
GGT GGC TCA AGT ATG GGC ATC ATT CGC ACA TGT AGG CTC GGC CCT GAC	1752
Gly Gly Ser Ser Met Gly Ile Ile Arg Thr Cys Arg Leu Gly Pro Asp	
20 25 30 35	
CAA GTC AAA TCC ATG CCG GCT GCT CTT GAT CTT TTC GGT CGT GAG TTC	1800
Gln Val Lys Ser Met Arg Ala Ala Leu Asp Leu Phe Gly Arg Glu Phe	
40 45 50	
GGA GAC GTA GCC ACC TAC TCC CAA CAT CAG CCG GAC TCC GAT TAC CTC	1848
Gly Asp Val Ala Thr Tyr Ser Gln His Gln Pro Asp Ser Asp Tyr Leu	
55 60 65	
GGG AAC TTG CTC CGT AGT AAG ACA TTC ATC GCG CTT GCT GCC TTC GAC	1896
Gly Asn Leu Leu Arg Ser Lys Thr Phe Ile Ala Leu Ala Ala Phe Asp	
70 75 80	
CAA GAA GCG GTT GTT GGC GCT CTC GCG GCT TAC GTT CTG CCC AGG TTT	1944

Gln	Glu	Ala	Val	Val	Gly	Ala	Leu	Ala	Ala	Tyr	Val	Leu	Pro	Arg	Phe	
85						90					95					
GAG	CAG	CCG	CGT	AGT	GAG	ATC	TAT	ATC	TAT	GAT	CTC	GCA	GTC	TCC	GGC	1992
Glu	Gln	Pro	Arg	Ser	Glu	Ile	Tyr	Ile	Tyr	Asp	Leu	Ala	Val	Ser	Gly	
100					105					110					115	
GAG	CAC	CGG	AGG	CAG	GGC	ATT	GCC	ACC	GCG	CTC	ATC	AAT	CTC	CTC	AAG	2040
Glu	His	Arg	Arg	Gln	Gly	Ile	Ala	Thr	Ala	Leu	Ile	Asn	Leu	Leu	Lys	
				120					125					130		
CAT	GAG	GCC	AAC	GCG	CTT	GGT	GCT	TAT	GTG	ATC	TAC	GTG	CAA	GCA	GAT	2088
His	Glu	Ala	Asn	Ala	Leu	Gly	Ala	Tyr	Val	Ile	Tyr	Val	Gln	Ala	Asp	
			135					140					145			
TAC	GGT	GAC	GAT	CCC	GCA	GTG	GCT	CTC	TAT	ACA	AAG	TTG	GGC	ATA	CGG	2136
Tyr	Gly	Asp	Asp	Pro	Ala	Val	Ala	Leu	Tyr	Thr	Lys	Leu	Gly	Ile	Arg	
	150						155					160				
GAA	GAA	GTG	ATG	CAC	TTT	GAT	ATC	GAC	CCA	AGT	ACC	GCC	ACC	TAA	CAA	2184
Glu	Glu	Val	Met	His	Phe	Asp	Ile	Asp	Pro	Ser	Thr	Ala	Thr			
165						170					175		177			
TTCTGTTCAAG	CCGAGATCGG	CTTCCCTG	CAA	AGT	CCT	GTG	GGT	GAG	TCG	AAC						2236
			Gln	Ser	Pro	Val	Gly	Glu	Ser	Asn						
			451				455									
TTG	GCG	ATG	CGC	GCA	CCC	TAC	GGA	GAA	GCG	ATC	CAC	GGA	CTG	CTC	TCT	2284
Leu	Ala	Met	Arg	Ala	Pro	Tyr	Gly	Glu	Ala	Ile	His	Gly	Leu	Leu	Ser	
460						465					470					
GTC	CTC	CTT	TCA	ACG	GAG	TGT	TAG	AACCGTTGGT	AGTGGTTTTG	GACGGGCCCCA						2338
Val	Leu	Leu	Ser	Thr	Glu	Cys										
475					480	481										
GGAGCATGCG	CTTCTGGGCC	CGTTTCTTGA	GTATTCATTG	GATAGTCACG	CGTGGTAGCT											2398
TCGAGCCTGC	ACAGCTGATG	AGCACCTTGG	AAGGCGCGCT	GTACGCGGAC	GA CTGGGTTC											2458
ATCTTCGCCA	TTCATGACGG	AACTCCGTTC	CCCAGTACCG	CGATGACTAT	TTTGCCTCTT											2518
CCGATGTCCG	ATTCCACGCC	GCCTGACGCT	AAGCGGGGGC	GGGGGCGCCC	GCATCCCAGC											2578
CCAGACAGCA	ACAAATGAGT	AGGCTCTTGG	ATGCCGCGGC	GGCTGAGATT	GGTAACGGCA											2638
ATTTCTGTC	AA	TGTGACGATG	GATTCGATTG	CCCGTGCTGC	CGGCGTCTCA	AAAAAACGC										2698
TGTACGTCTT	GGTGCCGAGC	AAGGAAGAAC	TCATTTCCCG	GTTAGTGGCT	CGAGACATGT											2758
CCAACCTTGA	GGAATTC															2775

图 2e:

GAATTCGCG TATCGCCCGG TTCTATCAGC GGGCCGCTTT CGAAAGTCAT GGTGTTAGCC	60
GGTAGGGTCT TTTCTTGGC CATGCTTGTT GCCTGAACCT TCGTTGACAT AGGGCAGAGG	120
TGCGTTTGCC GCTTCGCTTC GCGATGAACC GCATCGAGAT GCTGAGGTCA GGATTTTCC	180
TTAACTCGCG TAAGCATCTT GTCATTTT TGGTGGCTTT GAACAGCCTG ATGAAAGGTG	240
GTCTCGCCCT TTGAGGCCGA TTCTTGGGCG CTTGGCGGCG TCGAAGCGAT GCTCCACTAC	300
CGATTAAGAT AATTAAAATA AGGAAACCGC ATGGTTTCTT ATGTGAATTT GTCTGGCATA	360
CTCCAGCTCA AGGGCAATTT TTGGGCTATT GGCTGAGCAG TTGCCTCTAT ATGGTTATTC	420
AGAATAACAA TTGACTCCTC AGGAGGTCAG CG ATG AGC ATT CTT GGT TTG AAT	473
Met Ser Ile Leu Gly Leu Asn	
1 5	
GGT GCC CCG GTC GGA GCT GAG CAG CTG GGC TCG GCT CTT GAT CGC ATG	521
Gly Ala Pro Val Gly Ala Glu Gln Leu Gly Ser Ala Leu Asp Arg Met	
10 15 20	
AAG AAG GCG CAC CTG GAG CAG GGG CCT GCA AAC TTG GAG CTG CGT CTG	569
Lys Lys Ala His Leu Glu Gln Gly Pro Ala Asn Leu Glu Leu Arg Leu	
25 30 35	
AGT AGG CTG GAT CGT GCG ATT GCA ATG CTT CTG GAA AAT CGT GAA GCA	617
Ser Arg Leu Asp Arg Ala Ile Ala Met Leu Leu Glu Asn Arg Glu Ala	
40 45 50 55	
ATT GCC GAC GCG GTT TCT GCT GAC TTT GGC AAT CGC AGC CGT GAG CAA	665
Ile Ala Asp Ala Val Ser Ala Asp Phe Gly Asn Arg Ser Arg Glu Gln	
60 65 70	
ACA CTG CTT TGC GAC ATT GCT GGC TCG GTG GCA AGC CTG AAG GAT AGC	713
Thr Leu Leu Cys Asp Ile Ala Gly Ser Val Ala Ser Leu Lys Asp Ser	
75 80 85	
CGC GAG CAC GTG GCC AAA TGG ATG GAG CCC GAA CAT CAC AAG GCG ATG	761
Arg Glu His Val Ala Lys Trp Met Glu Pro Glu His His Lys Ala Met	
90 95 100	
TTT CCA GGG GCG GAG GCA CGC GTT GAG TTT CAG CCG CTG GGT GTC GTT	809
Phe Pro Gly Ala Glu Ala Arg Val Glu Phe Gln Pro Leu Gly Val Val	
105 110 115	
GGG GTC ATT AGT CCC TGG AAC TTC CCT ATC GTA CTG GCC TTT GGG CCG	857
Gly Val Ile Ser Pro Trp Asn Phe Pro Ile Val Leu Ala Phe Gly Pro	
120 125 130 135	
CTG GCC GGC ATA TTC GCA GCA GGT AAT CGC GCC ATG CTC AAG CCG TCC	905
Leu Ala Gly Ile Phe Ala Ala Gly Asn Arg Ala Met Leu Lys Pro Ser	
140 145 150	
GAG CTT ACC CCG CGG ACT TCT GCC CTG CTT GCG GAG CTA ATT GCT CGT	953
Glu Leu Thr Pro Arg Thr Ser Ala Leu Leu Ala Glu Leu Ile Ala Arg	
155 160 165	

TAC TTC GAT GAA ACT GAG CTG ACT ACA GTG CTG GGC GAC GCT GAA GTC Tyr Phe Asp Glu Thr Glu Leu Thr Thr Val Leu Gly Asp Ala Glu Val 170 175 180	1001
GGT GCG CTG TTC AGT GCT CAG CCT TTC GAT CAT CTG ATC TTC ACC GGC Gly Ala Leu Phe Ser Ala Gln Pro Phe Asp His Leu Ile Phe Thr Gly 185 190 195	1049
GGC ACT GCC GTG GCC AAG CAC ATC ATG CGT GCC GCG GCG GAT AAC CTA Gly Thr Ala Val Ala Lys His Ile Met Arg Ala Ala Ala Asp Asn Leu 200 205 210 215	1097
GTG CCC GTT ACC CTG GAA TTG GGT GGC AAA TCG CCG GTG ATC GTT TCC Val Pro Val Thr Leu Glu Leu Gly Gly Lys Ser Pro Val Ile Val Ser 220 225 230	1145
CGC AGT GCA GAT ATG GCG GAC GTT GCA CAA CGG GTG TTG ACG GTG AAA Arg Ser Ala Asp Met Ala Asp Val Ala Gln Arg Val Leu Thr Val Lys 235 240 245	1193
ACC TTC AAT GCC GGG CAA ATC TGT CTG GCA CC GTG GGT GAG TCG AAC Thr Phe Asn Ala Gly Gln Ile Cys Leu Ala Val Gly Glu Ser Asn 250 255 257 454 455	1240
TTG GCG ATG CGC GCA CCC TAC GGA GAA GCG ATC CAC GGA CTG CTC TCT Leu Ala Met Arg Ala Pro Tyr Gly Glu Ala Ile His Gly Leu Leu Ser 460 465 470	1288
GTC CTC CTT TCA ACG GAG TGT TAG AACCGTTGGT AGTGGTTTGT GACGGGCCCA Val Leu Leu Ser Thr Glu Cys 475 480 481	1342
GGAGCATGCG CTTCTGGGCC CGTTTCTTGA GTATTCATTG GATAGTCACG CGTGGTAGCT	1402
TCGAGCCTGC ACAGCTGATG AGCACCCTGG AAGGCGCGCT GTACGCGGAC GACTGGGTTC	1462
ATCTTCGCCA TTCATGACGG AACTCCGTTC CCCAGTACCG CGATGACTAT TTTGCCTCTT	1522
CCGATGTCCG ATTCCACGCC GCCTGACGCT AAGCGGGGGC GGGGGCGCCC GCATCCCAGC	1582
CCAGACAGCA ACAAATGAGT AGGCTCTTGG ATGCCGCGGC GGCTGAGATT GGTAACGGCA	1642
ATTTTCGTCAA TGTGACGATG GATTCGATTG CCCGTGCTGC CGGCGTCTCA AAAAAACGC	1702
TGTACGTCTT GGTGGCGAGC AAGGAAGAAC TCATTTCCCG GTTAGTGGCT CGAGACATGT	1762
CCAACCTTGA GGAATTC	1779

图 2f:

CTGCAGCCGA	GCATCGATTG	AGCACTTTAC	CCAGCTGCGC	TGGCTGACCA	TTCAGAAATGG	60
CCCCGGGCAC	TATCCAATCT	AAATCGATCT	TCGGGCGCCG	CGGGCATCAT	GCCCCGGGCG	120
CTCGCCTCAT	TTCAATCTCT	AACTTGATAA	AAACAGAGCT	GTTCTCCGGT	CTTGGTGGAT	180
CAAGGCCAGT	CGCGGAGAGT	CTCGAAGAGG	AGAGTACAGT	GAACGCCGAG	TCCACATTGC	240
AACCGCAGGC	ATCATCATGC	TCTGCTCAGC	CACGCTACCG	CAGTGTGTCG	ATTGGTCATC	300
CTCCGGTTGA	GGTTACGCAA	GACGCTGGAG	GTATTGTCCG	G ATG CGT TCT CTC GAG		356
				Met Arg Ser Leu Glu		
				1	5	
GCG CTT CTT CCC TTC CCG GGT CGA ATT CTT GAG CGT CTC GAG CAT TGG						404
Ala Leu Leu Pro Phe Pro Gly Arg Ile Leu Glu Arg Leu Glu His Trp						
	10		15		20	
GCT AAG ACC CGT CCA GAA CAA ACC TGC GTT GCT GCC AGG GCG GCA AAT						452
Ala Lys Thr Arg Pro Glu Gln Thr Cys Val Ala Ala Arg Ala Ala Asn						
	25		30		35	
GGG GAA TGG CGT CGT ATC AGC TAC GCG GAA ATG TTC CAC AAC GTC CGC						500
Gly Glu Trp Arg Arg Ile Ser Tyr Ala Glu Met Phe His Asn Val Arg						
	40		45		50	
GCC ATC GCA CAG AGC TTG CTT CCT TAC GGA CTA TCG GCA GAG CGT CCG						548
Ala Ile Ala Gln Ser Leu Leu Pro Tyr Gly Leu Ser Ala Glu Arg Pro						
	55		60		65	
CTG CTT ATC GTC TCT GGA AAT GAC CTG GAA CAT CTT CAG CTG GCA TTT						596
Leu Leu Ile Val Ser Gly Asn Asp Leu Glu His Leu Gln Leu Ala Phe						
	70		75		80	85
GGG GCT ATG TAT GCG GGC ATT CCC TAT TGC CCG GTG TCT CCT GCT TAT						644
Gly Ala Met Tyr Ala Gly Ile Pro Tyr Cys Pro Val Ser Pro Ala Tyr						
	90		95		100	
TCA CTG CTG TCG CAA GAT TTG GCG AAG CTG CGT CAC ATC GTA GGT CTT						692
Ser Leu Leu Ser Gln Asp Leu Ala Lys Leu Arg His Ile Val Gly Leu						
	105		110		115	
CTG CAA CCG GGA CTG GTC TTT GCT GCC GAT GCA GCA CCT TTC CAG GGG						740
Leu Gln Pro Gly Leu Val Phe Ala Ala Asp Ala Ala Pro Phe Gln						
	120		125		130	132
ACAGCAAGCG AACCGGAATT GCCAGCTGGG GCGCCCTCTG GTAAGGTTGG GAAGCCCTGC						800
AAAGTAAACT GGATGGCTTT CTTGCCGCCA AGGATCTGAT GGCGCAGGGG ATCAAGATCT						860
GATCAAGAGA CAGGATGAGG ATCGTTTCGC ATG ATT GAA CAA GAT GGA TTG CAC						914
				Met Ile Glu Gln Asp Gly Leu His		
			1		5	
GCA GGT TCT CCG GCC GCT TGG GTG GAG AGG CTA TTC GGC TAT GAC TGG						962
Ala Gly Ser Pro Ala Ala Trp Val Glu Arg Leu Phe Gly Tyr Asp Trp						
	10		15		20	

GCA CAA CAG ACA ATC GGC TGC TCT GAT GCC GCC GTG TTC CGG CTG TCA	1010
Ala Gln Gln Thr Ile Gly Cys Ser Asp Ala Ala Val Phe Arg Leu Ser	
25 30 35 40	
GCG CAG GGG CGC CCG GTT CTT TTT GTC AAG ACC GAC CTG TCC GGT GCC	1058
Ala Gln Gly Arg Pro Val Leu Phe Val Lys Thr Asp Leu Ser Gly Ala	
45 50 55	
CTG AAT GAA CTG CAG GAC GAG GCA GCG CGG CTA TCG TGG CTG GCC ACG	1106
Leu Asn Glu Leu Gln Asp Glu Ala Ala Arg Leu Ser Trp Leu Ala Thr	
60 65 70	
ACG GGC GTT CCT TGC GCA GCT GTG CTC GAC GTT GTC ACT GAA GCG GGA	1154
Thr Gly Val Pro Cys Ala Ala Val Leu Asp Val Val Thr Glu Ala Gly	
75 80 85	
AGG GAC TGG CTG CTA TTG GGC GAA GTG CCG GGG CAG GAT CTC CTG TCA	1202
Arg Asp Trp Leu Leu Leu Gly Glu Val Pro Gly Gln Asp Leu Leu Ser	
90 95 100	
TCT CAC CTT GCT CCT GCC GAG AAA GTA TCC ATC ATG GCT GAT GCA ATG	1250
Ser His Leu Ala Pro Ala Glu Lys Val Ser Ile Met Ala Asp Ala Met	
105 110 115 120	
CGG CGG CTG CAT ACG CTT GAT CCG GCT ACC TGC CCA TTC GAC CAC CAA	1298
Arg Arg Leu His Thr Leu Asp Pro Ala Thr Cys Pro Phe Asp His Gln	
125 130 135	
GCG AAA CAT CGC ATC GAG CGA GCA CGT ACT CGG ATG GAA GCC GGT CTT	1346
Ala Lys His Arg Ile Glu Arg Ala Arg Thr Arg Met Glu Ala Gly Leu	
140 145 150	
GTC GAT CAG GAT GAT CTG GAC GAA GAG CAT CAG GGG CTC GCG CCA GCC	1394
Val Asp Gln Asp Asp Leu Asp Glu Glu His Gln Gly Leu Ala Pro Ala	
155 160 165	
GAA CTG TTC GCC AGG CTC AAG GCG CGC ATG CCC GAC GGC GAG GAT CTC	1442
Glu Leu Phe Ala Arg Leu Lys Ala Arg Met Pro Asp Gly Glu Asp Leu	
170 175 180	
GTC GTG ACC CAT GGC GAT GCC TGC TTG CCG AAT ATC ATG GTG GAA AAT	1490
Val Val Thr His Gly Asp Ala Cys Leu Pro Asn Ile Met Val Glu Asn	
185 190 195 200	
GGC CGC TTT TCT GGA TTC ATC GAC TGT GGC CGG CTG GGT GTG GCG GAC	1538
Gly Arg Phe Ser Gly Phe Ile Asp Cys Gly Arg Leu Gly Val Ala Asp	
205 210 215	
CGC TAT CAG GAC ATA GCG TTG GCT ACC CGT GAT ATT GCT GAA GAG CTT	1586
Arg Tyr Gln Asp Ile Ala Leu Ala Thr Arg Asp Ile Ala Glu Glu Leu	
220 225 230	
GGC GGC GAA TGG GCT GAC CGC TTC CTC GTG CTT TAC GGT ATC GCC GCT	1634
Gly Gly Glu Trp Ala Asp Arg Phe Leu Val Leu Tyr Gly Ile Ala Ala	
235 240 245	

CCC GAT TCG CAG CGC ATC GCC TTC TAT CGC CTT CTT GAC GAG TTC TTC	1682
Pro Asp Ser Gln Arg Ile Ala Phe Tyr Arg Leu Leu Asp Glu Phe Phe	
250 255 260 264	
TGAGCGGGAC TCTGGGGTTC GAAATGACCG ACCAAGCGAC GCCCCT GTT TTG CAA	1737
Val Leu Gln	
563 565	
TGG CGG TCG GCG AAA GTT GAT GCG CTG TAT CGT GGT GAA GAT CAA TCC	1785
Trp Arg Ser Ala Lys Val Asp Ala Leu Tyr Arg Gly Glu Asp Gln Ser	
570 575 580	
ATG CTG CGT GAC GAG GCC ACA CTG TGA GTTGGTCAGG GGGGGCTTAC	1832
Met Leu Arg Asp Glu Ala Thr Leu	
585 589	
TCGGCGTTTT CCGACACTGC GTTGGTTGCG GCAGTGCACA CCCCCTGGAT TGATTGCGGG	1892
GGTGCCCTGT CGCTGGTGTC GCCTATCGAC TTAGGGGTAA AGGTCCGCTCG CGAAGTTCTG	1952
ATGCGTGCGT CGCTTGAACC ACAAATGGTC GATAGCGTAC TCGCAGGCTC TATGGCTCAA	2012
GCAAGCTTTG ATGCTTACCT GCTCCCGCGG CACATTGGCT TGTACAGCGG TGTMTCCAAG	2072
TCGGTTCGGG CCTTGGGGGT GCAGCGCATT TGCGGCACAG GCTTCGAACT GCTTCGGCAG	2132
GCCGGCGAGC AGATTTCCCA AGGCGCTGAT CACGTGCTGT GTGTCGCGGG CTGCAG	2188

图 2g:

CTGCAGCCGA GCATCGATTG AGCACTTTAC CCAGCTGCGC TGGCTGACCA TTCAGAATGG	60
CCCGCGGCAC TATCCAATCT AAATCGATCT TCGGGCGCCG CGGGCATCAT GCGCGCGGCG	120
CTCGCCTCAT TTCAATCTCT AACTTGATAA AAACAGAGCT GTTCTCCGGT CTTGGTGGAT	180
CAAGGCCAGT CGCGGAGAGT CTCGAAGAGG AGAGTACAGT GAACGCCGAG TCCACATTGC	240
AACCGCAGGC ATCATCATGC TCTGCTCAGC CACGCTACCG CAGTGTGTCG ATTGGTCATC	300
CTCCGGTTGA GGTACGCAA GACGCTGGAG GTATTGTCCG G ATG CGT TCT CTC GAG	356
Met Arg Ser Leu Glu	5
1	
GCG CTT CTT CCC TTC CCG GGT CGA ATT CTT GAG CGT CTC GAG CAT TGG	404
Ala Leu Leu Pro Phe Pro Gly Arg Ile Leu Glu Arg Leu Glu His Trp	
10 15 20	
GCT AAG ACC CGT CCA GAA CAA ACC TGC GTT GCT GCC AGG GCG GCA AAT	452
Ala Lys Thr Arg Pro Glu Gln Thr Cys Val Ala Ala Arg Ala Ala Asn	
25 30 35	
GGG GAA TGG CGT CGT ATC AGC TAC GCG GAA ATG TTC CAC AAC GTC CGC	500
Gly Glu Trp Arg Arg Ile Ser Tyr Ala Glu Met Phe His Asn Val Arg	
40 45 50	
GCC ATC GCA CAG AGC TTG CTT CCT TAC GGA CTA TCG GCA GAG CGT CCG	548
Ala Ile Ala Gln Ser Leu Leu Pro Tyr Gly Leu Ser Ala Glu Arg Pro	
55 60 65	
CTG CTT ATC GTC TCT GGA AAT GAC CTG GAA CAT CTT CAG CTG GCA TTT	596
Leu Leu Ile Val Ser Gly Asn Asp Leu Glu His Leu Gln Leu Ala Phe	
70 75 80 85	
GGG GCT ATG TAT GCG GGC ATT CCC TAT TGC CCG GTG TCT CCT GCT TAT	644
Gly Ala Met Tyr Ala Gly Ile Pro Tyr Cys Pro Val Ser Pro Ala Tyr	
90 95 100	
TCA CTG CTG TCG CAA GAT TTG GCG AAG CTG CGT CAC ATC GTA GGT CTT	692
Ser Leu Leu Ser Gln Asp Leu Ala Lys Leu Arg His Ile Val Gly Leu	
105 110 115	
CTG CAA CCG GGA CTG GTC TTT GCT GCC GAT GCA GCA CCT TTC CAG GGG	740
Leu Gln Pro Gly Leu Val Phe Ala Ala Asp Ala Ala Pro Phe Gln	
120 125 130 132	
GAGAGGCGGT TTGCGTATTG GCGCATGCA TAAAACTGT TGTAATTCAT TAAGCATTCT	800
GCCGACATGG AAGCCATCAC AAACGGCATG ATGAACCTGA ATCGCCAGCG GCATCAGCAC	860
CTTGTCGCCT TGCCTATAAT ATTTGCCCAT GGACGCACAC CGTGGAAACG GATGAAGGCA	920
CGAACCCAGT TGACATAAGC CTGTTCCGGT CGTAACTGT AATGCAAGTA GCGTATGCGC	980
TCACGCAACT GGTCCAGAAC CTTGACCGAA CGCAGCGGTG GTAACGGCGC AGTGGCGGTT	1040
TTTCATGGCTT GTTATGACTG TTTTTTTGTA CAGTCTATGC CTCGGGCATC CAAGCAGCAA	1100

GCGCGTTACG CCGTGGGTCG ATGTTTGATG TTATGGAGCA GCAACG ATG TTA CGC	1155
Met Leu Arg	
1	
AGC AGC AAC GAT GTT ACG CAG CAG GGC AGT CGC CCT AAA ACA AAG TTA	1203
Ser Ser Asn Asp Val Thr Gln Gln Gly Ser Arg Pro Lys Thr Lys Leu	
5 10 15	
GGT GGC TCA AGT ATG GGC ATC ATT CGC ACA TGT AGG CTC GGC CCT GAC	1251
Gly Gly Ser Ser Met Gly Ile Ile Arg Thr Cys Arg Leu Gly Pro Asp	
20 25 30 35	
CAA GTC AAA TCC ATG CGG GCT GCT CTT GAT CTT TTC GGT CGT GAG TTC	1299
Gln Val Lys Ser Met Arg Ala Ala Leu Asp Leu Phe Gly Arg Glu Phe	
40 45 50	
GGA GAC GTA GCC ACC TAC TCC CAA CAT CAG CCG GAC TCC GAT TAC CTC	1347
Gly Asp Val Ala Thr Tyr Ser Gln His Gln Pro Asp Ser Asp Tyr Leu	
55 60 65	
GGG AAC TTG CTC CGT AGT AAG ACA TTC ATC GCG CTT GCT GCC TTC GAC	1395
Gly Asn Leu Leu Arg Ser Lys Thr Phe Ile Ala Leu Ala Ala Phe Asp	
70 75 80	
CAA GAA GCG GTT GTT GGC GCT CTC GCG GCT TAC GTT CTG CCC AGG TTT	1443
Gln Glu Ala Val Val Gly Ala Leu Ala Ala Tyr Val Leu Pro Arg Phe	
85 90 95	
GAG CAG CCG CGT AGT GAG ATC TAT ATC TAT GAT CTC GCA GTC TCC GGC	1491
Glu Gln Pro Arg Ser Glu Ile Tyr Ile Tyr Asp Leu Ala Val Ser Gly	
100 105 110 115	
GAG CAC CCG AGG CAG GGC ATT GCC ACC GCG CTC ATC AAT CTC CTC AAG	1539
Glu His Arg Arg Gln Gly Ile Ala Thr Ala Leu Ile Asn Leu Leu Lys	
120 125 130	
CAT GAG GCC AAC GCG CTT GGT GCT TAT GTG ATC TAC GTG CAA GCA GAT	1587
His Glu Ala Asn Ala Leu Gly Ala Tyr Val Ile Tyr Val Gln Ala Asp	
135 140 145	
TAC GGT GAC GAT CCC GCA GTG GCT CTC TAT ACA AAG TTG GGC ATA CGG	1635
Tyr Gly Asp Asp Pro Ala Val Ala Leu Tyr Thr Lys Leu Gly Ile Arg	
150 155 160	
GAA GAA GTG ATG CAC TTT GAT ATC GAC CCA AGT ACC GCC ACC TAA CAA	1683
Glu Glu Val Met His Phe Asp Ile Asp Pro Ser Thr Ala Thr	
165 170 175 177	
TTCGTTCAAG CCGAGATCGG CTTCCCT GTT TTG CAA TGG CGG TCG GCG AAA	1735
Val Leu Gln Trp Arg Ser Ala Lys	
563 565 570	
GTT GAT GCG CTG TAT CGT GGT GAA GAT CAA TCC ATG CTG CGT GAC GAG	1783
Val Asp Ala Leu Tyr Arg Gly Glu Asp Gln Ser Met Leu Arg Asp Glu	
575 580 585	

GCC ACA CTG TGA GTTGGTCAGG GGGGGCTTAC TCGGCGTTTT CCGACACTGC	1835
Ala Thr Leu	
589	
GTTGGTTGCG GCAGTGCGCA CCCCCTGGAT TGATTGCGGG GGTGCCCTGT CGCTGGTGTC	1895
GCCTATCGAC TTAGGGGTAA AGGTCGCTCG CGAAGTTCTG ATGCGTGCGT CGCTTGAACC	1955
ACAAATGGTC GATAGCGTAC TCGCAGGCTC TATGGCTCAA GCAAGCTTTG ATGCTTACCT	2015
GCTCCCGCGG CACATTGGCT TGTACAGCGG TGTTCCTCAAG TCGGTTCCGG CCTTGGGGGT	2075
GCAGCGCATT TGCGGCACAG GCTTCGAACT GCTTCGGCAG GCCGGCGAGC AGATTTCCTCA	2135
AGGCGCTGAT CACGTGCTGT GTGTCGCGGG CTGCAG	2171

图 2h:

CTGCAGCCGA GCATCGATTG AGCACTTTAC CCAGCTGCGC TGGCTGACCA TTCAGAATGG	60
CCCGCGGCAC TATCCAATCT AAATCGATCT TCGGGCGCCG CGGGCATCAT GCGCGCGGCG	120
CTCGCCTCAT TTCAATCTCT AACTTGATAA AAACAGAGCT GTTCTCCGGT CTTGGTGGAT	180
CAAGGCCAGT CGCGGAGAGT CTCGAAGAGG AGAGTACAGT GAACGCCGAG TCCACATTGC	240
AACCGCAGGC ATCATCATGC TCTGCTCAGC CACGCTACCG CAGTGTGTCG ATTGGTCATC	300
CTCCGGTTGA GGTACGCAA GACGCTGGAG GTATTGTCCG G ATG CGT TCT CTC GAG	356
Met Arg Ser Leu Glu	5
1	
GCG CTT CTT CCC TTC CCG GGT CGA ATT CTT GAG CGT CTC GAG CAT TGG	404
Ala Leu Leu Pro Phe Pro Gly Arg Ile Leu Glu Arg Leu Glu His Trp	
10 15 20	
GCT AAG ACC CGT CCA GAA CAA ACC TGC GTT GCT GCC AGG GCG GCA AAT	452
Ala Lys Thr Arg Pro Glu Gln Thr Cys Val Ala Ala Arg Ala Ala Asn	
25 30 35	
GGG GAA TGG CGT CGT ATC AGC TAC GCG GAA ATG TTC CAC AAC GTC CGC	500
Gly Glu Trp Arg Arg Ile Ser Tyr Ala Glu Met Phe His Asn Val Arg	
40 45 50	
GCC ATC GCA CAG AGC TTG CTT CCT TAC GGA CTA TCG GCA GAG CGT CCG	548
Ala Ile Ala Gln Ser Leu Leu Pro Tyr Gly Leu Ser Ala Glu Arg Pro	
55 60 65	
CTG CTT ATC GTC TCT GGA AAT GAC CTG GAA CAT CTT CAG CTG GCA TTT	596
Leu Leu Ile Val Ser Gly Asn Asp Leu Glu His Leu Gln Leu Ala Phe	
70 75 80 85	
GGG GCT ATG TAT GCG GGC ATT CCC TAT TGC CCG GTG TCT CCT GCT TAT	644
Gly Ala Met Tyr Ala Gly Ile Pro Tyr Cys Pro Val Ser Pro Ala Tyr	
90 95 100	
TCA CTG CTG TCG CAA GAT TTG GCG AAG CTG CGT CAC ATC GTA GGT CTT	692
Ser Leu Leu Ser Gln Asp Leu Ala Lys Leu Arg His Ile Val Gly Leu	
105 110 115	
CTG CAA CCG GGA CTG GTC TTT GCT GCC GAT GCA GCA CCT TTC CAG CGC	740
Leu Gln Pro Gly Leu Val Phe Ala Ala Asp Ala Ala Pro Phe Gln Arg	
120 125 130 133	
GCT GTT TTG CAA TGG CGG TCG GCG AAA GTT GAT GCG CTG TAT CGT GGT	788
Ala Val Leu Gln Trp Arg Ser Ala Lys Val Asp Ala Leu Tyr Arg Gly	
562 565 570 575	
GAA GAT CAA TCC ATG CTG CGT GAC GAG GCC ACA CTG TGA GTTGGTCAGG	837
Glu Asp Gln Ser Met Leu Arg Asp Glu Ala Thr Leu	
580 585 589	
GGGGGCTTAC TCGGCGTTTT CCGACACTGC GTTGGTTGCG GCAGTGGCGCA CCCCCTGGAT	897
TGATTGCGGG GGTGCCCTGT CGCTGGTGTC GCCTATCGAC TTAGGGGTAA AGGTCGCTCG	957

CGAAGTTCTG ATGCGTGCGT CGCTTGAACC ACAAATGGTC GATAGCGTAC TCGCAGGCTC	1017
TATGGCTCAA GCAAGCTTTG ATGCTTACCT GCTCCCGCGG CACATTGGCT TGTACAGCGG	1077
TGTTUCCAAG TCGGTTCCGG CCTTGGGGGT GCAGCGCATT TCGGGCACAG GCTTCGAACT	1137
GCTTCGGCAG GCCGGCGAGC AGATTTCCTA AGGCGCTGAT CACGTGCTGT GTGTCGCGGG	1197
CTGCAG	1203

图 21:

GAATTCCTT GGCACGAAA GGGCGGCAGG CCGCATGGCC ACGGCTGGGC GGTAACGTAT 60
GCTTGCGTTA ATCGTTAACC GTTTGAAATT CCTTGCCAAA TTTCGGCGAG AGAATCATGC 120
GGGTACGCCT TTCCGTGCGC TTTGATCTGC GCTTCCGTGC CTTGAATCAG AAAAATAGTT 180
AATTGACAGA ACTATAGGTT CGCAGTAGCT TTTGCTCACC CACCAAATCC ACAGCACTGG 240
GGTGCACG ATG AAT AGC TAC GAT GGC CGT TGG TCT ACC GTT GAT GTG AAG 290
Met Asn Ser Tyr Asp Gly Arg Trp Ser Thr Val Asp Val Lys
1 5 10
GTT GAA GAA GGT ATC GCT TGG GTC ACG CTG AAC CGC CCG GAG AAG CGC 338
Val Glu Glu Gly Ile Ala Trp Val Thr Leu Asn Arg Pro Glu Lys Arg
15 20 25 30
AAC GCA ATG AGC CCA ACT CTC AAT CGA GAG ATG GTC GAG GTT CTG GAG 386
Asn Ala Met Ser Pro Thr Leu Asn Arg Glu Met Val Glu Val Leu Glu
35 40 45
GTG CTG GAG CAG GAC GCA GAT GCT CGC GTG CTT GTT CTG ACT GGT GCA 434
Val Leu Glu Gln Asp Ala Asp Ala Arg Val Leu Val Leu Thr Gly Ala
50 55 60
GGC GAA TCC TGG ACC GCG GGC ATG GAC CTG AAG GAG TAT TTC CGC GAG 482
Gly Glu Ser Trp Thr Ala Gly Met Asp Leu Lys Glu Tyr Phe Arg Glu
65 70 75
ACC GAT GCT GGC CCC GAA ATT CTG CAA GAG AAG ATT CGT CGGGGACAGC 531
Thr Asp Ala Gly Pro Glu Ile Leu Gln Glu Lys Ile Arg
80 85 90 91
AAGCGAACCG GAATTGCCAG CTGGGGCGCC CTCTGGTAAG GTTGGGAAGC CCTGCAAAGT 591
AAACTGGATG GCTTTCTTGC CGCCAAGGAT CTGATGGCGC AGGGGATCAA GATCTGATCA 651
AGAGACAGGA TGAGGATCGT TTCGC ATG ATT GAA CAA GAT GGA TTG CAC GCA 703
Met Ile Glu Gln Asp Gly Leu His Ala
1 5
GGT TCT CCG GCC GCT TGG GTG GAG AGG CTA TTC GGC TAT GAC TGG GCA 751
Gly Ser Pro Ala Ala Trp Val Glu Arg Leu Phe Gly Tyr Asp Trp Ala
10 15 20 25
CAA CAG ACA ATC GGC TGC TCT GAT GCC GCC GTG TTC CGG CTG TCA GCG 799
Gln Gln Thr Ile Gly Cys Ser Asp Ala Ala Val Phe Arg Leu Ser Ala
30 35 40
CAG GGG CGC CCG GTT CTT TTT GTC AAG ACC GAC CTG TCC GGT GCC CTG 847
Gln Gly Arg Pro Val Leu Phe Val Lys Thr Asp Leu Ser Gly Ala Leu
45 50 55
AAT GAA CTG CAG GAC GAG GCA GCG CGG CTA TCG TGG CTG GCC ACG ACG 895
Asn Glu Leu Gln Asp Glu Ala Ala Arg Leu Ser Trp Leu Ala Thr Thr
60 65 70

GGC GTT CCT TGC GCA GCT GTG CTC GAC GTT GTC ACT GAA GCG GGA AGG Gly Val Pro Cys Ala Ala Val Leu Asp Val Val Thr Glu Ala Gly Arg 75 80 85	943
GAC TGG CTG CTA TTG GGC GAA GTG CCG GGG CAG GAT CTC CTG TCA TCT Asp Trp Leu Leu Leu Gly Glu Val Pro Gly Gln Asp Leu Leu Ser Ser 90 95 100 105	991
CAC CTT GCT CCT GCC GAG AAA GTA TCC ATC ATG GCT GAT GCA ATG CGG His Leu Ala Pro Ala Glu Lys Val Ser Ile Met Ala Asp Ala Met Arg 110 115 120	1039
CGG CTG CAT ACG CTT GAT CCG GCT ACC TGC CCA TTC GAC CAC CAA GCG Arg Leu His Thr Leu Asp Pro Ala Thr Cys Pro Phe Asp His Gln Ala 125 130 135	1087
AAA CAT CGC ATC GAG CGA GCA CGT ACT CGG ATG GAA GCC GGT CTT GTC Lys His Arg Ile Glu Arg Ala Arg Thr Arg Met Glu Ala Gly Leu Val 140 145 150	1135
GAT CAG GAT GAT CTG GAC GAA GAG CAT CAG GGG CTC GCG CCA GCC GAA Asp Gln Asp Asp Leu Asp Glu Glu His Gln Gly Leu Ala Pro Ala Glu 155 160 165	1183
CTG TTC GCC AGG CTC AAG GCG CGC ATG CCC GAC GGC GAG GAT CTC GTC Leu Phe Ala Arg Leu Lys Ala Arg Met Pro Asp Gly Glu Asp Leu Val 170 175 180 185	1231
GTG ACC CAT GGC GAT GCC TGC TTG CCG AAT ATC ATG GTG GAA AAT GGC Val Thr His Gly Asp Ala Cys Leu Pro Asn Ile Met Val Glu Asn Gly 190 195 200	1279
CGC TTT TCT GGA TTC ATC GAC TGT GGC CGG CTG GGT GTG GCG GAC CGC Arg Phe Ser Gly Phe Ile Asp Cys Gly Arg Leu Gly Val Ala Asp Arg 205 210 215	1327
TAT CAG GAC ATA GCG TTG GCT ACC CGT GAT ATT GCT GAA GAG CTT GGC Tyr Gln Asp Ile Ala Leu Ala Thr Arg Asp Ile Ala Glu Glu Leu Gly 220 225 230	1375
GGC GAA TGG GCT GAC CGC TTC CTC GTG CTT TAC GGT ATC GCC GCT CCC Gly Glu Trp Ala Asp Arg Phe Leu Val Leu Tyr Gly Ile Ala Ala Pro 235 240 245	1423
GAT TCG CAG CGC ATC GCC TTC TAT CGC CTT CTT GAC GAG TTC TTC TGA Asp Ser Gln Arg Ile Ala Phe Tyr Arg Leu Leu Asp Glu Phe Phe 250 255 260 264	1471
GCGGGACTCT GGGGTTCGAA ATGACCGACC AAGCGACGCC CC GAG CAG GGC ATG Glu Gln Gly Met 255	1525
AAG CAG TTC CTT GAC GAG AAA AGC ATC AAG CCG GGC TTG CAG ACC TAC Lys Gln Phe Leu Asp Glu Lys Ser Ile Lys Pro Gly Leu Gln Thr Tyr 260 265 270	1573

AAG CGC TGA TAAATGCGCC GGGGCCCTCG CTGCGCCCCC GGCCTTCCAA TAATGACAAT 1632
 Lys Arg
 275 276

 AATGAGGAGT GCCCAATGTT TCACGTGCCC CTGCTTATTG GTGGTAAGCC TTGTTTCAGCA 1692
 TCTGATGAGC GCACCTTCGA GCGTCGTAGC CCGCTGACCG GAGAAGTGGT ATCGCGCGTC 1752
 GCTGCTGCCA GTTTGGAAGA TCGGGACGCC GCAGTGGCCG CTGCACAGGC TCGTTTCTCT 1812
 GAATGGGCGG CGCTTGCTCC GAGCGAACGC CGTGCCCGAC TGCTGCGAGC GCGGGATCTT 1872
 CTAGAGGACC GTTCTTCCGA GTTCACCGCC GCAGCGAGTG AACTGGCGC AGCGGGAAAC 1932
 TGGTATGGGT TTAACGTTTA CCTGGCGGCG GGCATGTTGC GGGGAATTC 1981

图 2j:

GAATTCCCCT GCGGACGAAA GGGCGGCAGG CCGCATGGCC ACGGCTGGGC GGTAACATGAT 60
 GCTTGCGTTA ATCGTTAACC GTTTGAAATT CCTTGCCAAA TTTCGGCGAG AGAATCATGC 120
 GGGTACGCCT TTCCGTGCGC TTTGATCTGC GCTTCCGTGC CTTGAATCAG AAAAATAGTT 180
 AATTGACAGA ACTATAGGTT CGCAGTAGCT TTTGCTCACC CACCAAATCC ACAGCACTGG 240
 GGTGCACG ATG AAT AGC TAC GAT GGC CGT TGG TCT ACC GTT GAT GTG AAG 290
 Met Asn Ser Tyr Asp Gly Arg Trp Ser Thr Val Asp Val Lys
 1 5 10
 GTT GAA GAA GGT ATC GCT TGG GTC ACG CTG AAC CGC CCG GAG AAG CGC 338
 Val Glu Glu Gly Ile Ala Trp Val Thr Leu Asn Arg Pro Glu Lys Arg
 15 20 25 30
 AAC GCA ATG AGC CCA ACT CTC AAT CGA GAG ATG GTC GAG GTT CTG GAG 386
 Asn Ala Met Ser Pro Thr Leu Asn Arg Glu Met Val Glu Val Leu Glu
 35 40 45
 GTG CTG GAG CAG GAC GCA GAT GCT CGC GTG CTT GTT CTG ACT GGT GCA 434
 Val Leu Glu Gln Asp Ala Asp Ala Arg Val Leu Val Leu Thr Gly Ala
 50 55 60
 GGC GAA TCC TGG ACC GCG GGC ATG GAC CTG AAG GAG TAT TTC CGC GAG 482
 Gly Glu Ser Trp Thr Ala Gly Met Asp Leu Lys Glu Tyr Phe Arg Glu
 65 70 75
 ACC GAT GCT GGC CCC GAA ATT CTG CAA GAG AAG ATT CGT CGGGGGAGAG 531
 Thr Asp Ala Gly Pro Glu Ile Leu Gln Glu Lys Ile Arg
 80 85 90 91
 GCGGTTTGCG TATTGGGCGC ATGCATAAAA ACTGTTGTAA TTCATTAAGC ATTCTGCCGA 591
 CATGGAAGCC ATCACAAACG GCATGATGAA CCTGAATCGC CAGCGGCATC AGCACCTTGT 651
 CGCCTTGCGT ATAATATTTG CCCATGGACG CACACCGTGG AAACGGATGA AGGCACGAAC 711
 CCAGTTGACA TAAGCCTGTT CGGTTTCGTAA ACTGTAATGC AAGTAGCGTA TGCGCTCAGC 771
 CAACTGGTCC AGAACCTTGA CCGAACGCAG CGGTGGTAAC GGCGCAGTGG CGGTTTTCAT 831
 GGCTTGTTAT GACTGTTTTT TTGTACAGTC TATGCCTCGG GCATCCAAGC AGCAAGCGCG 891
 TTACGCCGTG GGTCGATGTT TGATGTTATG GAGCAGCAAC G ATG TTA CGC AGC AGC 947
 Met Leu Arg Ser Ser
 1 5
 AAC GAT GTT ACG CAG CAG GGC AGT CGC CCT AAA ACA AAG TTA GGT GGC 995
 Asn Asp Val Thr Gln Gln Gly Ser Arg Pro Lys Thr Lys Leu Gly Gly
 10 15 20
 TCA AGT ATG GGC ATC ATT CGC ACA TGT AGG CTC GGC CCT GAC CAA GTC 1043
 Ser Ser Met Gly Ile Ile Arg Thr Cys Arg Leu Gly Pro Asp Gln Val
 25 30 35

AAA TCC ATG CGG GCT GCT CTT GAT CTT TTC GGT CGT GAG TTC GGA GAC Lys Ser Met Arg Ala Ala Leu Asp Leu Phe Gly Arg Glu Phe Gly Asp 40 45 50	1091
GTA GCC ACC TAC TCC CAA CAT CAG CCG GAC TCC GAT TAC CTC GGG AAC Val Ala Thr Tyr Ser Gln His Gln Pro Asp Ser Asp Tyr Leu Gly Asn 55 60 65	1139
TTG CTC CGT AGT AAG ACA TTC ATC GCG CTT GCT GCC TTC GAC CAA GAA Leu Leu Arg Ser Lys Thr Phe Ile Ala Leu Ala Ala Phe Asp Gln Glu 70 75 80 85	1187
GCG GTT GTT GGC GCT CTC GCG GCT TAC GTT CTG CCC AGG TTT GAG CAG Ala Val Val Gly Ala Leu Ala Ala Tyr Val Leu Pro Arg Phe Glu Gln 90 95 100	1235
CCG CGT AGT GAG ATC TAT ATC TAT GAT CTC GCA GTC TCC GGC GAG CAC Pro Arg Ser Glu Ile Tyr Ile Tyr Asp Leu Ala Val Ser Gly Glu His 105 110 115	1283
CGG AGG CAG GGC ATT GCC ACC GCG CTC ATC AAT CTC CTC AAG CAT GAG Arg Arg Gln Gly Ile Ala Thr Ala Leu Ile Asn Leu Leu Lys His Glu 120 125 130	1331
GCC AAC GCG CTT GGT GCT TAT GTG ATC TAC GTG CAA GCA GAT TAC GGT Ala Asn Ala Leu Gly Ala Tyr Val Ile Tyr Val Gln Ala Asp Tyr Gly 135 140 145	1379
GAC GAT CCC GCA GTG GCT CTC TAT ACA AAG TTG GGC ATA CGG GAA GAA Asp Asp Pro Ala Val Ala Leu Tyr Thr Lys Leu Gly Ile Arg Glu Glu 150 155 160 165	1427
GTG ATG CAC TTT GAT ATC GAC CCA AGT ACC GCC ACC TAA CAATTCGTTT Val Met His Phe Asp Ile Asp Pro Ser Thr Ala Thr 170 175 177	1476
AAGCCGAGAT CGGCTTCCCC GAG CAG GGC ATG AAG CAG TTC CTT GAC GAG Glu Gln Gly Met Lys Gln Phe Leu Asp Glu 255 260	1526
AAA AGC ATC AAG CCG GGC TTG CAG ACC TAC AAG CGC TGA TAAATGCGCC Lys Ser Ile Lys Pro Gly Leu Gln Thr Tyr Lys Arg 265 270 275 276	1575
GGGGCCCTCG CTGCGCCCCC GGCTTCCAA TAATGACAAT AATGAGGAGT GCCCAATGTT	1635
TCACGTGCCC CTGCTTATTG GTGGTAAGCC TTGTTTACGA TCTGATGAGC GCACCTTCGA	1695
GCGTCGTAGC CCGCTGACCG GAGAAGTGGT ATCGCGCGTC GCTGCTGCCA GTTTGGAAGA	1755
TGCGGACGCC GCAGTGGCCG CTGCACAGGC TCGGTTTCCT GAATGGGCGG CGCTTGCTCC	1815
GAGCGAACGC CGTGCCCGAC TGCTGCGAGC GGCGGATCTT CTAGAGGACC GTTCTTCCGA	1875

01.04.30

GTTACCGCC GCAGCGAGTG AACTGGCGC AGCGGGAAC TGGTATGGGT TTAACGTTTA 1935

CCTGGCGGCG GGCATGTTGC GGGGAATTC 1964

Fig 2k:

GAATTCCTT	GGCGACGAAA	GGGCGGCAGG	CCGCATGGCC	ACGGCTGGGC	GGTAACTGAT	60
GCTTGCCTTA	ATCGTTAACC	GTTTGAAATT	CCTTGCCAAA	TTTCGGCGAG	AGAATCATGC	120
GGGTACGCCT	TTCCGTGCGC	TTTGATCTGC	GCTTCCGTGC	CTTGAATCAG	AAAAATAGTT	180
AATTGACAGA	ACTATAGGTT	CGCAGTAGCT	TTTGCTCACC	CACCAAATCC	ACAGCACTGG	240
GGTGCACG	ATG AAT AGC TAC GAT GGC CGT TGG TCT ACC GTT GAT GTG AAG	290				
	Met Asn Ser Tyr Asp Gly Arg Trp Ser Thr Val Asp Val Lys					
	1 5 10					
GTT GAA GAA GGT ATC GCT TGG GTC ACG CTG AAC CGC CCG GAG AAG CGC	338					
Val Glu Glu Gly Ile Ala Trp Val Thr Leu Asn Arg Pro Glu Lys Arg						
15 20 25 30						
AAC GCA ATG AGC CCA ACT CTC AAT CGA GAG ATG GTC GAG GTT CTG GAG	386					
Asn Ala Met Ser Pro Thr Leu Asn Arg Glu Met Val Glu Val Leu Glu						
35 40 45						
GTG CTG GAG CAG GAC GCA GAT GCT CGC GTG CTT GTT CTG ACT GGT GCA	434					
Val Leu Glu Gln Asp Ala Asp Ala Arg Val Leu Val Leu Thr Gly Ala						
50 55 60						
GGC GAA TCC TGG ACC GCG GGC ATG GAC CTG AAG GAG TAT TTC CGC GAG	482					
Gly Glu Ser Trp Thr Ala Gly Met Asp Leu Lys Glu Tyr Phe Arg Glu						
65 70 75						
ACC GAT GCT GGC CCC GAA ATT CTG CAA GAG AAG ATT CGT CGC GAG CAG	530					
Thr Asp Ala Gly Pro Glu Ile Leu Gln Glu Lys Ile Arg Arg Glu Gln						
80 85 90 92 255						
GGC ATG AAG CAG TTC CTT GAC GAG AAA AGC ATC AAG CCG GGC TTG CAG	578					
Gly Met Lys Gln Phe Leu Asp Glu Lys Ser Ile Lys Pro Gly Leu Gln						
260 265 270						
ACC TAC AAG CGC TGA TAAATGCGCC GGGGCCCTCG CTGCGCCCCC GGCCTTCCAA	633					
Thr Tyr Lys Arg						
275 276						
TAATGACAAT AATGAGGAGT GCCCAATGTT TCACGTGCCC CTGCTTATTG GTGGTAAGCC	693					
TTGTTTACGCA TCTGATGAGC GCACCTTCGA GCGTCGTAGC CCGCTGACCG GAGAAGTGGT	753					
ATCGCGCGTC GCTGCTGCCA GTTTGGAAGA TGCGGACGCC GCAGTGGCCG CTGCACAGGC	813					
TGCGTTTCCTT GAATGGGCGG CGCTTGCTCC GAGCGAACGC CGTGCCCGAC TGCTGCGAGC	873					
GGCGGATCTT CTAGAGGACC GTTCTTCCGA GTTCACCGCC GCAGCGAGTG AACTGGCGC	933					
AGCGGGAAAC TGGTATGGGT TTAACGTTTA CCTGGCGGCG GGCATGTTGC GGGGAATTC	992					

图 21:

GAATTCCAAT AATGACAATA ATGAGGAGTG CCCA ATG TTT CAC GTG CCC CTG CTT	55
Met Phe His Val Pro Leu Leu	
1 5	
ATT GGT GGT AAG CCT TGT TCA GCA TCT GAT GAG CGC ACC TTC GAG CGT	103
Ile Gly Gly Lys Pro Cys Ser Ala Ser Asp Glu Arg Thr Phe Glu Arg	
10 15 20	
CGT AGC CCG CTG ACC GGA GAA GTG GTA TCG CGC GTC GCT GCT GCC AGT	151
Arg Ser Pro Leu Thr Gly Glu Val Val Ser Arg Val Ala Ala Ala Ser	
25 30 35	
TTG GAA GAT GCG GAC GCC GCA GTG GCC GCT GCA CAG GCT GCG TTT CCT	199
Leu Glu Asp Ala Asp Ala Ala Val Ala Ala Ala Gln Ala Ala Phe Pro	
40 45 50 55	
GAA TGG GCG GCG CTT GCT CCG AGC GAA CGC CGT GCC CGA CTG CTG CGA	247
Glu Trp Ala Ala Leu Ala Pro Ser Glu Arg Arg Ala Arg Leu Leu Arg	
60 65 70	
GCG GCG GAT CTT CTA GAG GAC CGT TCT TCC GAG TTC ACC GCC GCA GCG	295
Ala Ala Asp Leu Leu Glu Asp Arg Ser Ser Glu Phe Thr Ala Ala Ala	
75 80 85	
AGT GAA ACT GGC GCA GCG GGA AAC TGG TAT GGG TTT AAC GTT TAC CTG	343
Ser Glu Thr Gly Ala Ala Gly Asn Trp Tyr Gly Phe Asn Val Tyr Leu	
90 95 100	
GCG GCG GGC ATG TTG CCG GAA GCC GCG GCC ATG ACC ACA CAG ATT CAG	391
Ala Ala Gly Met Leu Arg Glu Ala Ala Ala Met Thr Thr Gln Ile Gln	
105 110 115	
GGC GAT GTC ATT CCG TCC AAT GTG CCC GGT AGC TTT GCC ATG GCG GTT	439
Gly Asp Val Ile Pro Ser Asn Val Pro Gly Ser Phe Ala Met Ala Val	
120 125 130 135	
CGA CAG CCA TGT GGC GTG GTG CTC GGT ATT GCG CCT TGG AAT GCT CCG	487
Arg Gln Pro Cys Gly Val Val Leu Gly Ile Ala Pro Trp Asn Ala Pro	
140 145 150	
GTA ATC CTT GGC GTA CCG GCT GTT GCG ATG CCG TTG GCA TGC GGC AAT	535
Val Ile Leu Gly Val Arg Ala Val Ala Met Pro Leu Ala Cys Gly Asn	
155 160 165	
ACC GTG GTG TTG AAA AGC TCT GAG CTG AGT CCC TTT ACC CAT CGC CTG	583
Thr Val Val Leu Lys Ser Ser Glu Leu Ser Pro Phe Thr His Arg Leu	
170 175 180	
ATT GGT CAG GTG TTG CAT GAT GCT GGT CTG GGG GAT GGC GTG GTG AAT	631
Ile Gly Gln Val Leu His Asp Ala Gly Leu Gly Asp Gly Val Val Asn	
185 190 195	
GTC ATC AGC AAT GCC CCG CAA GAC GCT CCT GCG GTG GTG GAG CGA CTG	679
Val Ile Ser Asn Ala Pro Gln Asp Ala Pro Ala Val Val Glu Arg Leu	
200 205 210 215	

ATT GCA AAT CCT GCG GTA CGT CGA GTG AAC TTC ACC GGT TCG ACC CAC Ile Ala Asn Pro Ala Val Arg Arg Val Asn Phe Thr Gly Ser Thr His 220 225 230	727
GTT GGA CGG ATC ATT GGT GAG CTG TCT GCG CGT CAT CTG AAG CCT GCT Val Gly Arg Ile Ile Gly Glu Leu Ser Ala Arg His Leu Lys Pro Ala 235 240 245	775
GTG CTG GAA TTA GGT GGT AAG GCT CCG TTC TTG GTC TTG GAC GAT GCC Val Leu Glu Leu Gly Gly Lys Ala Pro Phe Leu Val Leu Asp Asp Ala 250 255 260	823
GAC CTC GAT GCG GCG GTC GAA GCG GCG GCC TTT GGT GCC TAC TTC AAT Asp Leu Asp Ala Ala Val Glu Ala Ala Ala Phe Gly Ala Tyr Phe Asn 265 270 275	871
CAG GGT CAA ATC TGC ATG TCC ACT GAG CGT CTG ATT GTG ACA GCA GTC Gln Gly Gln Ile Cys Met Ser Thr Glu Arg Leu Ile Val Thr Ala Val 280 285 290 295	919
GCA GAC GCC TTT GTT GAA AAG CTG GCG AGG AAG GTC GCC ACA CTG CGT Ala Asp Ala Phe Val Glu Lys Leu Ala Arg Lys Val Ala Thr Leu Arg 300 305 310	967
GCT GGC GAT CCT AAT GAT CCG CAA TCG GTC TTG GGT TCG TTG ATT GAT Ala Gly Asp Pro Asn Asp Pro Gln Ser Val Leu Gly Ser Leu Ile Asp 315 320 325	1015
GCC AAT GCA GGT CAA CGC ATC CAG GTT CTG GTC GAT GAT GCG CTC GGG Ala Asn Ala Gly Gln Arg Ile Gln Val Leu Val Asp Asp Ala Leu 330 335 340 342	1063
GACAGCAAGC GAACCGGAAT TGCCAGCTGG GGCGCCCTCT GGTAAGGTTG GGAAGCCCTG	1123
CAAAGTAAAC TGGATGGCTT TCTTGCCGCC AAGGATCTGA TGGCGCAGGG GATCAAGATC	1183
TGATCAAGAG ACAGGATGAG GATCGTTTCG C ATG ATT GAA CAA GAT GGA TTG Met Ile Glu Gln Asp Gly Leu 1 5	1235
CAC GCA GGT TCT CCG GCC GCT TGG GTG GAG AGG CTA TTC GGC TAT GAC His Ala Gly Ser Pro Ala Ala Trp Val Glu Arg Leu Phe Gly Tyr Asp 10 15 20	1283
TGG GCA CAA CAG ACA ATC GGC TGC TCT GAT GCC GCC GTG TTC CGG CTG Trp Ala Gln Gln Thr Ile Gly Cys Ser Asp Ala Ala Val Phe Arg Leu 25 30 35	1331
TCA GCG CAG GGG CGC CCG GTT CTT TTT GTC AAG ACC GAC CTG TCC GGT Ser Ala Gln Gly Arg Pro Val Leu Phe Val Lys Thr Asp Leu Ser Gly 40 45 50 55	1379
GCC CTG AAT GAA CTG CAG GAC GAG GCA GCG CGG CTA TCG TGG CTG GCC Ala Leu Asn Glu Leu Gln Asp Glu Ala Ala Arg Leu Ser Trp Leu Ala 60 65 70	1427

ACG	ACG	GGC	GTT	CCT	TGC	GCA	GCT	GTG	CTC	GAC	GTT	GTC	ACT	GAA	GCG	1475
Thr	Thr	Gly	Val	Pro	Cys	Ala	Ala	Val	Leu	Asp	Val	Val	Thr	Glu	Ala	
			75					80					85			
GGA	AGG	GAC	TGG	CTG	CTA	TTG	GGC	GAA	GTG	CCG	GGG	CAG	GAT	CTC	CTG	1523
Gly	Arg	Asp	Trp	Leu	Leu	Leu	Gly	Glu	Val	Pro	Gly	Gln	Asp	Leu	Leu	
		90					95					100				
TCA	TCT	CAC	CTT	GCT	CCT	GCC	GAG	AAA	GTA	TCC	ATC	ATG	GCT	GAT	GCA	1571
Ser	Ser	His	Leu	Ala	Pro	Ala	Glu	Lys	Val	Ser	Ile	Met	Ala	Asp	Ala	
	105					110					115					
ATG	CGG	CGG	CTG	CAT	ACG	CTT	GAT	CCG	GCT	ACC	TGC	CCA	TTC	GAC	CAC	1619
Met	Arg	Arg	Leu	His	Thr	Leu	Asp	Pro	Ala	Thr	Cys	Pro	Phe	Asp	His	
120					125					130					135	
CAA	GCG	AAA	CAT	CGC	ATC	GAG	CGA	GCA	CGT	ACT	CGG	ATG	GAA	GCC	GGT	1667
Gln	Ala	Lys	His	Arg	Ile	Glu	Arg	Ala	Arg	Thr	Arg	Met	Glu	Ala	Gly	
			140					145					150			
CTT	GTC	GAT	CAG	GAT	GAT	CTG	GAC	GAA	GAG	CAT	CAG	GGG	CTC	GCG	CCA	1715
Leu	Val	Asp	Gln	Asp	Asp	Leu	Asp	Glu	Glu	His	Gln	Gly	Leu	Ala	Pro	
			155					160					165			
GCC	GAA	CTG	TTC	GCC	AGG	CTC	AAG	GCG	CGC	ATG	CCC	GAC	GGC	GAG	GAT	1763
Ala	Glu	Leu	Phe	Ala	Arg	Leu	Lys	Ala	Arg	Met	Pro	Asp	Gly	Glu	Asp	
	170					175						180				
CTC	GTC	GTG	ACC	CAT	GGC	GAT	GCC	TGC	TTG	CCG	AAT	ATC	ATG	GTG	GAA	1811
Leu	Val	Val	Thr	His	Gly	Asp	Ala	Cys	Leu	Pro	Asn	Ile	Met	Val	Glu	
	185					190					195					
AAT	GGC	CGC	TTT	TCT	GGA	TTC	ATC	GAC	TGT	GGC	CGG	CTG	GGT	GTG	GCG	1859
Asn	Gly	Arg	Phe	Ser	Gly	Phe	Ile	Asp	Cys	Gly	Arg	Leu	Gly	Val	Ala	
200					205					210				215		
GAC	CGC	TAT	CAG	GAC	ATA	GCG	TTG	GCT	ACC	CGT	GAT	ATT	GCT	GAA	GAG	1907
Asp	Arg	Tyr	Gln	Asp	Ile	Ala	Leu	Ala	Thr	Arg	Asp	Ile	Ala	Glu	Glu	
			220					225					230			
CTT	GGC	GGC	GAA	TGG	GCT	GAC	CGC	TTC	CTC	GTG	CTT	TAC	GGT	ATC	GCC	1955
Leu	Gly	Gly	Glu	Trp	Ala	Asp	Arg	Phe	Leu	Val	Leu	Tyr	Gly	Ile	Ala	
			235					240					245			
GCT	CCC	GAT	TCG	CAG	CGC	ATC	GCC	TTC	TAT	CGC	CTT	CTT	GAC	GAG	TTC	2003
Ala	Pro	Asp	Ser	Gln	Arg	Ile	Ala	Phe	Tyr	Arg	Leu	Leu	Asp	Glu	Phe	
	250					255					260					
TTC	TGA	GCGGGACTCT	GGGGTTTCGAA	ATGACCGACC	AAGCGACGCC	CG	GCC	CAG								2057
Phe								Ala	Gln							
264								421								
CGC	GTC	GAT	TCG	GGC	ATT	TGC	CAT	ATC	AAT	GGA	CCG	ACT	GTG	CAT	GAC	2105
Arg	Val	Asp	Ser	Gly	Ile	Cys	His	Ile	Asn	Gly	Pro	Thr	Val	His	Asp	
	425					430					435					

GAG GCT CAG ATG CCA TTC GGT GGG GTG AAG TCC AGC GGC TAC GGC AGC	2153
Glu Ala Gln Met Pro Phe Gly Gly Val Lys Ser Ser Gly Tyr Gly Ser	
440 445 450	
TTC GGC AGT CGA GCA TCG ATT GAG CAC TTT ACC CAG CTG CGC TGG CTG	2201
Phe Gly Ser Arg Ala Ser Ile Glu His Phe Thr Gln Leu Arg Trp Leu	
455 460 465 470	
ACC ATT CAG AAT GGC CCG CGG CAC TAT CCA ATC TAA ATCGATCTTC	2247
Thr Ile Gln Asn Gly Pro Arg His Tyr Pro Ile	
475 480 481	
GGGCGCCGCG GGCATCATGC CCGCGGCGCT CGCCTCATTT CAATCTCTAA CTTGATAAAA	2307
ACAGAGCTGT TCTCCGGTCT TGGTGGATCA AGGCCAGTCG CGGAGAGTCT CGAAGAGGAG	2367
AGTACAGTGA ACGCCGAGTC CACATTGCAA CCGCAGGCAT CATCATGCTC TGCTCAGCCA	2427
CGCTACCGCA GTGTGTCGAT TGGTCATCCT CCGGTTGAGG TTACGCAAGA CGCTGGAGGT	2487
ATTGTCCGGA TGC GTTCTCT CGAGGCGCTT CTTCCCTTCC CGGGTGAAT TC	2539

图 2m:

01.04.30

GAATTC	CAAT	AATGACAATA	ATGAGGAGTG	CCCA	ATG	TTT	CAC	GTG	CCC	CTG	CTT		55			
					Met	Phe	His	Val	Pro	Leu	Leu					
					1				5							
ATT	GGT	GGT	AAG	CCT	TGT	TCA	GCA	TCT	GAT	GAG	CGC	ACC	TTC	GAG	CGT	103
Ile	Gly	Gly	Lys	Pro	Cys	Ser	Ala	Ser	Asp	Glu	Arg	Thr	Phe	Glu	Arg	
	10						15				20					
CGT	AGC	CCG	CTG	ACC	GGA	GAA	GTG	GTA	TCG	CGC	GTC	GCT	GCT	GCC	AGT	151
Arg	Ser	Pro	Leu	Thr	Gly	Glu	Val	Val	Ser	Arg	Val	Ala	Ala	Ala	Ser	
	25					30				35						
TTG	GAA	GAT	GCG	GAC	GCC	GCA	GTG	GCC	GCT	GCA	CAG	GCT	GCG	TTT	CCT	199
Leu	Glu	Asp	Ala	Asp	Ala	Ala	Val	Ala	Ala	Ala	Gln	Ala	Ala	Phe	Pro	
	40				45				50						55	
GAA	TGG	GCG	GCG	CTT	GCT	CCG	AGC	GAA	CGC	CGT	GCC	CGA	CTG	CTG	CGA	247
Glu	Trp	Ala	Ala	Leu	Ala	Pro	Ser	Glu	Arg	Arg	Ala	Arg	Leu	Leu	Arg	
				60					65				70			
GCG	GCG	GAT	CTT	CTA	GAG	GAC	CGT	TCT	TCC	GAG	TTC	ACC	GCC	GCA	GCG	295
Ala	Ala	Asp	Leu	Leu	Glu	Asp	Arg	Ser	Ser	Glu	Phe	Thr	Ala	Ala	Ala	
			75				80						85			
AGT	GAA	ACT	GGC	GCA	GCG	GGA	AAC	TGG	TAT	GGG	TTT	AAC	GTT	TAC	CTG	343
Ser	Glu	Thr	Gly	Ala	Ala	Gly	Asn	Trp	Tyr	Gly	Phe	Asn	Val	Tyr	Leu	
	90						95					100				
GCG	GCG	GGC	ATG	TTG	CGG	GAA	GCC	GCG	GCC	ATG	ACC	ACA	CAG	ATT	CAG	391
Ala	Ala	Gly	Met	Leu	Arg	Glu	Ala	Ala	Ala	Met	Thr	Thr	Gln	Ile	Gln	
	105					110				115						
GGC	GAT	GTC	ATT	CCG	TCC	AAT	GTG	CCC	GGT	AGC	TTT	GCC	ATG	GCG	GTT	439
Gly	Asp	Val	Ile	Pro	Ser	Asn	Val	Pro	Gly	Ser	Phe	Ala	Met	Ala	Val	
	120				125				130					135		
CGA	CAG	CCA	TGT	GGC	GTG	GTG	CTC	GGT	ATT	GCG	CCT	TGG	AAT	GCT	CCG	487
Arg	Gln	Pro	Cys	Gly	Val	Val	Leu	Gly	Ile	Ala	Pro	Trp	Asn	Ala	Pro	
				140					145				150			
GTA	ATC	CTT	GGC	GTA	CGG	GCT	GTT	GCG	ATG	CCG	TTG	GCA	TGC	GGC	AAT	535
Val	Ile	Leu	Gly	Val	Arg	Ala	Val	Ala	Met	Pro	Leu	Ala	Cys	Gly	Asn	
			155				160					165				
ACC	GTG	GTG	TTG	AAA	AGC	TCT	GAG	CTG	AGT	CCC	TTT	ACC	CAT	CGC	CTG	583
Thr	Val	Val	Leu	Lys	Ser	Ser	Glu	Leu	Ser	Pro	Phe	Thr	His	Arg	Leu	
	170						175					180				
ATT	GGT	CAG	GTG	TTG	CAT	GAT	GCT	GGT	CTG	GGG	GAT	GGC	GTG	GTG	AAT	631
Ile	Gly	Gln	Val	Leu	His	Asp	Ala	Gly	Leu	Gly	Asp	Gly	Val	Val	Asn	
	185					190				195						
GTC	ATC	AGC	AAT	GCC	CCG	CAA	GAC	GCT	CCT	GCG	GTG	GTG	GAG	CGA	CTG	679
Val	Ile	Ser	Asn	Ala	Pro	Gln	Asp	Ala	Pro	Ala	Val	Val	Glu	Arg	Leu	
	200				205					210				215		

ATT GCA AAT CCT GCG GTA CGT CGA GTG AAC TTC ACC GGT TCG ACC CAC Ile Ala Asn Pro Ala Val Arg Arg Val Asn Phe Thr Gly Ser Thr His 220 225 230	727
GTT GGA CGG ATC ATT GGT GAG CTG TCT GCG CGT CAT CTG AAG CCT GCT Val Gly Arg Ile Ile Gly Glu Leu Ser Ala Arg His Leu Lys Pro Ala 235 240 245	775
GTG CTG GAA TTA GGT GGT AAG GCT CCG TTC TTG GTC TTG GAC GAT GCC Val Leu Glu Leu Gly Gly Lys Ala Pro Phe Leu Val Leu Asp Asp Ala 250 255 260	823
GAC CTC GAT GCG GCG GTC GAA GCG GCG GCC TTT GGT GCC TAC TTC AAT Asp Leu Asp Ala Ala Val Glu Ala Ala Ala Phe Gly Ala Tyr Phe Asn 265 270 275	871
CAG GGT CAA ATC TGC ATG TCC ACT GAG CGT CTG ATT GTG ACA GCA GTC Gln Gly Gln Ile Cys Met Ser Thr Glu Arg Leu Ile Val Thr Ala Val 280 285 290 295	919
GCA GAC GCC TTT GTT GAA AAG CTG GCG AGG AAG GTC GCC ACA CTG CGT Ala Asp Ala Phe Val Glu Lys Leu Ala Arg Lys Val Ala Thr Leu Arg 300 305 310	967
GCT GGC GAT CCT AAT GAT CCG CAA TCG GTC TTG GGT TCG TTG ATT GAT Ala Gly Asp Pro Asn Asp Pro Gln Ser Val Leu Gly Ser Leu Ile Asp 315 320 325	1015
GCC AAT GCA GGT CAA CGC ATC CAG GTGGGGAGAG GCGGTTTGCG TATTGGGCGC Ala Asn Ala Gly Gln Arg Ile Gln 330 335	1069
ATGCATAAAA ACTGTTGTAA TTCATTAAGC ATTCTGCCGA CATGGAAGCC ATCACAAACG	1129
GCATGATGAA CCTGAATCGC CAGCGGCATC AGCACCTTGT CGCCTTGCGT ATAATATTTG	1189
CCCATGGACG CACACCGTGG AAACGGATGA AGGCACGAAC CCAGTTGACA TAAGCCTGTT	1249
CGGTTTCGTAA ACTGTAATGC AAGTAGCGTA TGCCTCAGC CAACTGGTCC AGAACCTTGA	1309
CCGAACGCAG CGGTGGTAAC GCGCGAGTGG CGGTTTTCAT GGCTTGTTAT GACTGTTTTT	1369
TTGTACAGTC TATGCCTCGG GCATCCAAGC AGCAAGCGCG TTACGCCGTG GGTGATGTT	1429
TGATGTTATG GAGCAGCAAC G ATG TTA CGC AGC AGC AAC GAT GTT ACG CAG Met Leu Arg Ser Ser Asn Asp Val Thr Gln 1 5 10	1480
CAG GGC AGT CGC CCT AAA ACA AAG TTA GGT GGC TCA AGT ATG GGC ATC Gln Gly Ser Arg Pro Lys Thr Lys Leu Gly Gly Ser Ser Met Gly Ile 15 20 25	1528
ATT CGC ACA TGT AGG CTC GGC CCT GAC CAA GTC AAA TCC ATG CGG GCT Ile Arg Thr Cys Arg Leu Gly Pro Asp Gln Val Lys Ser Met Arg Ala 30 35 40	1576

GCT CTT GAT CTT TTC GGT CGT GAG TTC GGA GAC GTA GCC ACC TAC TCC Ala Leu Asp Leu Phe Gly Arg Glu Phe Gly Asp Val Ala Thr Tyr Ser 45 50 55	1624
CAA CAT CAG CCG GAC TCC GAT TAC CTC GGG AAC TTG CTC CGT AGT AAG Gln His Gln Pro Asp Ser Asp Tyr Leu Gly Asn Leu Leu Arg Ser Lys 60 65 70	1672
ACA TTC ATC GCG CTT GCT GCC TTC GAC CAA GAA GCG GTT GTT GGC GCT Thr Phe Ile Ala Leu Ala Ala Phe Asp Gln Glu Ala Val Val Gly Ala 75 80 85 90	1720
CTC GCG GCT TAC GTT CTG CCC AGG TTT GAG CAG CCG CGT AGT GAG ATC Leu Ala Ala Tyr Val Leu Pro Arg Phe Glu Gln Pro Arg Ser Glu Ile 95 100 105	1768
TAT ATC TAT GAT CTC GCA GTC TCC GGC GAG CAC CGG AGG CAG GGC ATT Tyr Ile Tyr Asp Leu Ala Val Ser Gly Glu His Arg Arg Gln Gly Ile 110 115 120	1816
GCC ACC GCG CTC ATC AAT CTC CTC AAG CAT GAG GCC AAC GCG CTT GGT Ala Thr Ala Leu Ile Asn Leu Leu Lys His Glu Ala Asn Ala Leu Gly 125 130 135	1864
GCT TAT GTG ATC TAC GTG CAA GCA GAT TAC GGT GAC GAT CCC GCA GTG Ala Tyr Val Ile Tyr Val Gln Ala Asp Tyr Gly Asp Asp Pro Ala Val 140 145 150	1912
GCT CTC TAT ACA AAG TTG GGC ATA CGG GAA GAA GTG ATG CAC TTT GAT Ala Leu Tyr Thr Lys Leu Gly Ile Arg Glu Glu Val Met His Phe Asp 155 160 165 170	1960
ATC GAC CCA AGT ACC GCC ACC TAA CAATTCGTTC AAGCCGAGAT CGGCTTCCCA Ile Asp Pro Ser Thr Ala Thr 175 177	2014
A TTG GCC CAG CGC GTC GAT TCG GGC ATT TGC CAT ATC AAT GGA CCG ACT Leu Ala Gln Arg Val Asp Ser Gly Ile Cys His Ile Asn Gly Pro Thr 420 425 430 435	2063
GTG CAT GAC GAG GCT CAG ATG CCA TTC GGT GGG GTG AAG TCC AGC GGC Val His Asp Glu Ala Gln Met Pro Phe Gly Gly Val Lys Ser Ser Gly 440 445 450	2111
TAC GGC AGC TTC GGC AGT CGA GCA TCG ATT GAG CAC TTT ACC CAG CTG Tyr Gly Ser Phe Gly Ser Arg Ala Ser Ile Glu His Phe Thr Gln Leu 455 460 465	2159
CGC TGG CTG ACC ATT CAG AAT GGC CCG CGG CAC TAT CCA ATC TAA Arg Trp Leu Thr Ile Gln Asn Gly Pro Arg His Tyr Pro Ile 470 475 480 481	2204
ATCGATCTTC GGGCGCCGCG GGCATCATGC CCGCGGCGCT CGCCTCATT CAATCTCTAA	2264
CTTGATAAAA ACAGAGCTGT TCTCCGGTCT TGGTGGATCA AGGCCAGTCG CGGAGAGTCT	2324

01.04.30

CGAAGAGGAG AGTACAGTGA ACGCCGAGTC CACATTGCAA CCGCAGGCAT CATCATGCTC	2384
TGCTCAGCCA CGCTACCGCA GTGTGTCGAT TGGTCATCCT CCGGTTGAGG TTACGCAAGA	2444
CGCTGGAGGT ATTGTCCGGA TGC GTTCTCT CGAGGCGCTT CTTCCCTTCC CGGGTGGAAT	2504
TC	2506

2n:

01.04.30

GAATTC	CAAT	AATGACAATA	ATGAGGAGTG	CCCA	ATG	TTT	CAC	GTG	CCC	CTG	CTT		55			
					Met	Phe	His	Val	Pro	Leu	Leu					
					1				5							
ATT	GGT	GGT	AAG	CCT	TGT	TCA	GCA	TCT	GAT	GAG	CGC	ACC	TTC	GAG	CGT	103
Ile	Gly	Gly	Lys	Pro	Cys	Ser	Ala	Ser	Asp	Glu	Arg	Thr	Phe	Glu	Arg	
		10					15					20				
CGT	AGC	CCG	CTG	ACC	GGA	GAA	GTG	GTA	TCG	CGC	GTG	GCT	GCT	GCC	AGT	151
Arg	Ser	Pro	Leu	Thr	Gly	Glu	Val	Val	Ser	Arg	Val	Ala	Ala	Ala	Ser	
		25				30					35					
TTG	GAA	GAT	GCG	GAC	GCC	GCA	GTG	GCC	GCT	GCA	CAG	GCT	GCG	TTT	CCT	199
Leu	Glu	Asp	Ala	Asp	Ala	Ala	Val	Ala	Ala	Ala	Gln	Ala	Ala	Phe	Pro	
	40				45				50						55	
GAA	TGG	GCG	GCG	CTT	GCT	CCG	AGC	GAA	CGC	CGT	GCC	CGA	CTG	CTG	CGA	247
Glu	Trp	Ala	Ala	Leu	Ala	Pro	Ser	Glu	Arg	Arg	Ala	Arg	Leu	Leu	Arg	
				60					65					70		
GCG	GCG	GAT	CTT	CTA	GAG	GAC	CGT	TCT	TCC	GAG	TTC	ACC	GCC	GCA	GCG	295
Ala	Ala	Asp	Leu	Leu	Glu	Asp	Arg	Ser	Ser	Glu	Phe	Thr	Ala	Ala	Ala	
			75				80						85			
AGT	GAA	ACT	GGC	GCA	GCG	GGA	AAC	TGG	TAT	GGG	TTT	AAC	GTT	TAC	CTG	343
Ser	Glu	Thr	Gly	Ala	Ala	Gly	Asn	Trp	Tyr	Gly	Phe	Asn	Val	Tyr	Leu	
		90					95					100				
GCG	GCG	GGC	ATG	TTG	CGG	GAA	GCC	GCG	GCC	ATG	ACC	ACA	CAG	ATT	CAG	391
Ala	Ala	Gly	Met	Leu	Arg	Glu	Ala	Ala	Ala	Met	Thr	Thr	Gln	Ile	Gln	
	105					110					115					
GGC	GAT	GTC	ATT	CCG	TCC	AAT	GTG	CCC	GGT	AGC	TTT	GCC	ATG	GCG	GTT	439
Gly	Asp	Val	Ile	Pro	Ser	Asn	Val	Pro	Gly	Ser	Phe	Ala	Met	Ala	Val	
	120				125				130					135		
CGA	CAG	CCA	TGT	GGC	GTG	GTG	CTC	GGT	ATT	GCG	CCT	TGG	AAT	GCT	CCG	487
Arg	Gln	Pro	Cys	Gly	Val	Val	Leu	Gly	Ile	Ala	Pro	Trp	Asn	Ala	Pro	
				140				145					150			
GTA	ATC	CTT	GGC	GTA	CGG	GCT	GTT	GCG	ATG	CCG	TTG	GCA	TGC	GGC	AAT	535
Val	Ile	Leu	Gly	Val	Arg	Ala	Val	Ala	Met	Pro	Leu	Ala	Cys	Gly	Asn	
		155					160						165			
ACC	GTG	GTG	TTG	AAA	AGC	TCT	GAG	CTG	AGT	CCC	TTT	ACC	CAT	CGC	CTG	583
Thr	Val	Val	Leu	Lys	Ser	Ser	Glu	Leu	Ser	Pro	Phe	Thr	His	Arg	Leu	
		170					175					180				
ATT	GGT	CAG	GTG	TTG	CAT	GAT	GCT	GGT	CTG	GGG	GAT	GGC	GTG	GTG	AAT	631
Ile	Gly	Gln	Val	Leu	His	Asp	Ala	Gly	Leu	Gly	Asp	Gly	Val	Val	Asn	
	185					190				195						
GTC	ATC	AGC	AAT	GCC	CCG	CAA	GAC	GCT	CCT	GCG	GTG	GTG	GAG	CGA	CTG	679
Val	Ile	Ser	Asn	Ala	Pro	Gln	Asp	Ala	Pro	Ala	Val	Val	Glu	Arg	Leu	
	200				205					210					215	

ATT GCA AAT CCT GCG GTA CGT CGA GTG AAC TTC ACC GGT TCG ACC CAC Ile Ala Asn Pro Ala Val Arg Arg Val Asn Phe Thr Gly Ser Thr His 220 225 230	727
GTT GGA CGG ATC ATT GGT GAG CTG TCT GCG CGT CAT CTG AAG CCT GCT Val Gly Arg Ile Ile Gly Glu Leu Ser Ala Arg His Leu Lys Pro Ala 235 240 245	775
GTG CTG GAA TTA GGT GGT AAG GCT CCG TTC TTG GTC TTG GAC GAT GCC Val Leu Glu Leu Gly Gly Lys Ala Pro Phe Leu Val Leu Asp Asp Ala 250 255 260	823
GAC CTC GAT GCG GCG GTC GAA GCG GCG GCC TTT GGT GCC TAC TTC AAT Asp Leu Asp Ala Ala Val Glu Ala Ala Ala Phe Gly Ala Tyr Phe Asn 265 270 275	871
CAG GGT CAA ATC TGC ATG TCC ACT GAG CGT CTG ATT GTG ACA GCA GTC Gln Gly Gln Ile Cys Met Ser Thr Glu Arg Leu Ile Val Thr Ala Val 280 285 290 295	919
GCA GAC GCC TTT GTT GAA AAG CTG GCG AGG AAG GTC GCC ACA CTG CGT Ala Asp Ala Phe Val Glu Lys Leu Ala Arg Lys Val Ala Thr Leu Arg 300 305 310	967
GCT GGC GAT CCT AAT GAT CCG CAA TCG GTC TTG GGT TCG TTG ATT GAT Ala Gly Asp Pro Asn Asp Pro Gln Ser Val Leu Gly Ser Leu Ile Asp 315 320 325	1015
GCC AAT GCA GGT CAA CGC ATC CAG GTT CTG GTC GAT GAT GCG CTC GCA Ala Asn Ala Gly Gln Arg Ile Gln Val Leu Val Asp Asp Ala Leu Ala 330 335 340	1063
AAA GGC GCG CAATGGAA TTG GCC CAG CGC GTC GAT TCG GGC ATT TGC CAT Lys Gly Ala Leu Ala Gln Arg Val Asp Ser Gly Ile Cys His 345 346 420 425 430	1113
ATC AAT GGA CCG ACT GTG CAT GAC GAG GCT CAG ATG CCA TTC GGT GGG Ile Asn Gly Pro Thr Val His Asp Glu Ala Gln Met Pro Phe Gly Gly 435 440 445	1161
GTG AAG TCC AGC GGC TAC GGC AGC TTC GGC AGT CGA GCA TCG ATT GAG Val Lys Ser Ser Gly Tyr Gly Ser Phe Gly Ser Arg Ala Ser Ile Glu 450 455 460	1209
CAC TTT ACC CAG CTG CGC TGG CTG ACC ATT CAG AAT GGC CCG CGG CAC His Phe Thr Gln Leu Arg Trp Leu Thr Ile Gln Asn Gly Pro Arg His 465 470 475	1257
TAT CCA ATC TAA ATCGATCTTC GGGCGCCGCG GGCATCATGC CCGCGGCGCT Tyr Pro Ile 480 481	1309
CGCCTCATT CAATCTCTAA CTTGATAAAA ACAGAGCTGT TCTCCGGTCT TGGTGGATCA	1369
AGGCCAGTCG CGGAGACTCT CGAAGAGGAG AGTACASTGA ACGCCGAGTC CACATTGCAA	1429

010430

CCGCAGGCAT CATCATGCTC TGCTCAGCCA CGCTACCGCA GTGTGTGGAT TGGTCATCCT	1489
CCGGTTGAGG TTACGCAAGA CGCTGGAGGT ATTGTCCGGA TCGTTCTCT CGAGGCGCTT	1549
CTTCCCTTCC CGGGTGAAT TC	1571

图 20:

GAATTCGCG GTCGGCGAAA GTTGATGCGC TGTATCGTGG TGAAGATCAA TCCATGCTGC	60
GTGACGAGGC CACACT GTG AGT TGG TCA GGG GGG GCT TAC TCG GCG TTT TCC	112
Met Ser Trp Ser Gly Gly Ala Tyr Ser Ala Phe Ser	
1 5 10	
GAC ACT GCG TTG GTT GCG GCA GTG CGC ACC CCC TGG ATT GAT TGC GGG	160
Asp Thr Ala Leu Val Ala Ala Val Arg Thr Pro Trp Ile Asp Cys Gly	
15 20 25	
GGT GCC CTG TCG CTG GTG TCG CCT ATC GAC TTA GGG GTA AAG GTC GCT	208
Gly Ala Leu Ser Leu Val Ser Pro Ile Asp Leu Gly Val Lys Val Ala	
30 35 40	
CGC GAA GTT CTG ATG CGT GCG TCG CTT GAA CCA CAA ATG GTC GAT AGC	256
Arg Glu Val Leu Met Arg Ala Ser Leu Glu Pro Gln Met Val Asp Ser	
45 50 55 60	
GTA CTC GCA GGC TCT ATG GCT CAA GCA AGC TTT GAT GCT TAC CTG CTC	304
Val Leu Ala Gly Ser Met Ala Gln Ala Ser Phe Asp Ala Tyr Leu Leu	
65 70 75	
CCG CGG CAC ATT GGC TTG TAC AGC GGT GTT CCC AAG TCG GTT CCG GCC	352
Pro Arg His Ile Gly Leu Tyr Ser Gly Val Pro Lys Ser Val Pro Ala	
80 85 90	
TTG GGG GTG CAG CGC ATT TGC GGC ACA GGC TTC GAA CTG CTT CGG CAG	400
Leu Gly Val Gln Arg Ile Cys Gly Thr Gly Phe Glu Leu Leu Arg Gln	
95 100 105	
GCC GGC GAG CAG ATT TCC CAA GGC GCT GAT CAC GTG CTG TGT GTC GCG	448
Ala Gly Glu Gln Ile Ser Gln Gly Ala Asp His Val Leu Cys Val Ala	
110 115 120	
GCA GAG TCC ATG TCG CGT AAC CCC ATC GCG TCG TAT ACA CAC CGG GGC	496
Ala Glu Ser Met Ser Arg Asn Pro Ile Ala Ser Tyr Thr His Arg Gly	
125 130 135 140	
GGG TTC CGC CTC GGT GCG CCC GTT GAG TTC AAG GAT TTT TTG TGG GAG	544
Gly Phe Arg Leu Gly Ala Pro Val Glu Phe Lys Asp Phe Leu Trp Glu	
145 150 155	
GCA TTG TTT GAT CCT GCT CCA GGA CTC GAC ATG ATC GCT ACC GCA GAA	592
Ala Leu Phe Asp Pro Ala Pro Gly Leu Asp Met Ile Ala Thr Ala Glu	
160 165 170	
AAC CTG GGGACAGCAA GCGAACCGGA ATTGCCAGCT GGGGCGCCCT CTGGTAAGGT	648
Asn Leu	
174	
TGGGAAGCCC TGCAAAGTAA ACTGGATGGC TTTCTTGCCG CCAAGGATCT GATGGCGCAG	708
GGGATCAAGA TCTGATCAAG AGACAGGATG AGGATCGTTT CGC ATG ATT GAA CAA	763
Met Ile Glu Gln	
1	

GAT	GGA	TTG	CAC	GCA	GGT	TCT	CCG	GCC	GCT	TGG	GTG	GAG	AGG	CTA	TTC	811
Asp	Gly	Leu	His	Ala	Gly	Ser	Pro	Ala	Ala	Trp	Val	Glu	Arg	Leu	Phe	
5					10					15					20	
GGC	TAT	GAC	TGG	GCA	CAA	CAG	ACA	ATC	GGC	TGC	TCT	GAT	GCC	GCC	GTG	859
Gly	Tyr	Asp	Trp	Ala	Gln	Gln	Thr	Ile	Gly	Cys	Ser	Asp	Ala	Ala	Val	
				25					30					35		
TTC	CGG	CTG	TCA	GCG	CAG	GGG	CGC	CCG	GTT	CTT	TTT	GTC	AAG	ACC	GAC	907
Phe	Arg	Leu	Ser	Ala	Gln	Gly	Arg	Pro	Val	Leu	Phe	Val	Lys	Thr	Asp	
			40					45					50			
CTG	TCC	GGT	GCC	CTG	AAT	GAA	CTG	CAG	GAC	GAG	GCA	GCG	CGG	CTA	TCG	955
Leu	Ser	Gly	Ala	Leu	Asn	Glu	Leu	Gln	Asp	Glu	Ala	Ala	Arg	Leu	Ser	
		55				60						65				
TGG	CTG	GCC	ACG	ACG	GGC	GTT	CCT	TGC	GCA	GCT	GTG	CTC	GAC	GTT	GTC	1003
Trp	Leu	Ala	Thr	Thr	Gly	Val	Pro	Cys	Ala	Ala	Val	Leu	Asp	Val	Val	
	70					75					80					
ACT	GAA	GCG	GGA	AGG	GAC	TGG	CTG	CTA	TTG	GGC	GAA	GTG	CCG	GGG	CAG	1051
Thr	Glu	Ala	Gly	Arg	Asp	Trp	Leu	Leu	Leu	Gly	Glu	Val	Pro	Gly	Gln	
85				90						95					100	
GAT	CTC	CTG	TCA	TCT	CAC	CTT	GCT	CCT	GCC	GAG	AAA	GTA	TCC	ATC	ATG	1099
Asp	Leu	Leu	Ser	Ser	His	Leu	Ala	Pro	Ala	Glu	Lys	Val	Ser	Ile	Met	
				105				110						115		
GCT	GAT	GCA	ATG	CGG	CGG	CTG	CAT	ACG	CTT	GAT	CCG	GCT	ACC	TGC	CCA	1147
Ala	Asp	Ala	Met	Arg	Arg	Leu	His	Thr	Leu	Asp	Pro	Ala	Thr	Cys	Pro	
			120					125					130			
TTC	GAC	CAC	CAA	GCG	AAA	CAT	CGC	ATC	GAG	CGA	GCA	CGT	ACT	CGG	ATG	1195
Phe	Asp	His	Gln	Ala	Lys	His	Arg	Ile	Glu	Arg	Ala	Arg	Thr	Arg	Met	
	135					140						145				
GAA	GCC	GGT	CTT	GTC	GAT	CAG	GAT	GAT	CTG	GAC	GAA	GAG	CAT	CAG	GGG	1243
Glu	Ala	Gly	Leu	Val	Asp	Gln	Asp	Asp	Leu	Asp	Glu	Glu	His	Gln	Gly	
	150					155					160					
CTC	GCG	CCA	GCC	GAA	CTG	TTC	GCC	AGG	CTC	AAG	GCG	CGC	ATG	CCC	GAC	1291
Leu	Ala	Pro	Ala	Glu	Leu	Phe	Ala	Arg	Leu	Lys	Ala	Arg	Met	Pro	Asp	
165					170					175					180	
GGC	GAG	GAT	CTC	GTC	GTG	ACC	CAT	GGC	GAT	GCC	TGC	TTG	CCG	AAT	ATC	1339
Gly	Glu	Asp	Leu	Val	Val	Thr	His	Gly	Asp	Ala	Cys	Leu	Pro	Asn	Ile	
				185				190						195		
ATG	GTG	GAA	AAT	GGC	CGC	TTT	TCT	GGA	TTC	ATC	GAC	TGT	GGC	CGG	CTG	1387
Met	Val	Glu	Asn	Gly	Arg	Phe	Ser	Gly	Phe	Ile	Asp	Cys	Gly	Arg	Leu	
			200					205					210			
GGT	GTG	GCG	GAC	CGC	TAT	CAG	GAC	ATA	GCG	TTG	GCT	ACC	CGT	GAT	ATT	1435
Gly	Val	Ala	Asp	Arg	Tyr	Gln	Asp	Ile	Ala	Leu	Ala	Thr	Arg	Asp	Ile	
	215					220						225				

GCT GAA GAG CTT GGC GGC GAA TGG GCT GAC CGC TTC CTC GTG CTT TAC Ala Glu Glu Leu Gly Gly Glu Trp Ala Asp Arg Phe Leu Val Leu Tyr 230 235 240	1483
GGT ATC GCC GCT CCC GAT TCG CAG CGC ATC GCC TTC TAT CGC CTT CTT Gly Ile Ala Ala Pro Asp Ser Gln Arg Ile Ala Phe Tyr Arg Leu Leu 245 250 255 260	1531
GAC GAG TTC TTC TGA GCGGGACTCT GGGGTTCGAA ATGACCGACC AAGCGACGCC Asp Glu Phe Phe 264	1586
CA TTG AGG GCG CAA GAG GAG AAA TGG ATT GAC CAA GAG ATC GTG GCT Leu Arg Ala Gln Glu Glu Lys Trp Ile Asp Gln Glu Ile Val Ala 197 200 205 210	1633
GTT ACG GAT GAA CAG TTC GAT TTA GAG GGC TAC AAC AGT CGA GCA ATT Val Thr Asp Glu Gln Phe Asp Leu Glu Gly Tyr Asn Ser Arg Ala Ile 215 220 225	1681
GAA CTG CCT CGG AAG GCA AAA TTG TTG ATC GTG ACA GTC ATC CGC GGC Glu Leu Pro Arg Lys Ala Lys Leu Leu Ile Val Thr Val Ile Arg Gly 230 235 240	1729
CTA GCA GTC TTT GAA GCC CTT TCC CGA TTG AAG CCT GTT CAT TCT GGC Leu Ala Val Phe Glu Ala Leu Ser Arg Leu Lys Pro Val His Ser Gly 245 250 255	1777
GGG GTG CAG ACT GCG GGC AAC AGC TGT GCC GTA GTG GAC GGC GCC GCG Gly Val Gln Thr Ala Gly Asn Ser Cys Ala Val Val Asp Gly Ala Ala 260 265 270 275	1825
GCG GCT TTG GTG GCT CGA GAG TCG TCT GCG ACA CAG CCG GTC TTG GCT Ala Ala Leu Val Ala Arg Glu Ser Ser Ala Thr Gln Pro Val Leu Ala 280 285 290	1873
AGG ATA CTG GCT ACC TCC GTA GTC GGG ATC GAG CCC GAG CAT ATG GGG Arg Ile Leu Ala Thr Ser Val Val Gly Ile Glu Pro Glu His Met Gly 295 300 305	1921
CTC GGC CCT GCG CCC GCG ATT CGC CTG CTG CTT GCG CGT AGT GAT CTT Leu Gly Pro Ala Pro Ala Ile Arg Leu Leu Leu Ala Arg Ser Asp Leu 310 315 320	1969
AGT TTG AGG GAT ATC GAC CTC TTT GAG ATA AAC GAG GCG CAG GCC GCC Ser Leu Arg Asp Ile Asp Leu Phe Glu Ile Asn Glu Ala Gln Ala Ala 325 330 335	2017
CAA GTT CTA GCG GTA CAG CAT GAA TTG GGT ATT GAG CAC TCA AAA CTT Gln Val Leu Ala Val Gln His Glu Leu Gly Ile Glu His Ser Lys Leu 340 345 350 355	2065
AAT ATT TGG GGC GGG GCC ATT GCA CTT GGA CAC CCG CTT GCC GCG ACC Asn Ile Trp Gly Gly Ala Ile Ala Leu Gly His Pro Leu Ala Ala Thr 360 365 370	2113

GGA TTG CGT CTC TGC ATG ACC CTC GCT CAC CAA TTG CAA GCT AAT AAC	2161
Gly Leu Arg Leu Cys Met Thr Leu Ala His Gln Leu Gln Ala Asn Asn	
375 380 385	
TTT CGA TAT GGA ATT GCC TCG GCA TGC ATT GGT GGG GGA CAG GGG ATG	2209
Phe Arg Tyr Gly Ile Ala Ser Ala Cys Ile Gly Gly Gly Gln Gly Met	
390 395 400	
GCG GTT CTT TTA GAG AAT CCC CAC TTC GGT TCG TCC TCT GCA CGA AGT	2257
Ala Val Leu Leu Glu Asn Pro His Phe Gly Ser Ser Ser Ala Arg Ser	
405 410 415	
TCG ATG ATT AAC AGA GTT GAC CAC TAT CCA CTG AGC TAA CGGGCATCTC	2306
Ser Met Ile Asn Arg Val Asp His Tyr Pro Leu Ser	
420 425 430 431	
CTTTGTTGCT TTGAGGTGGC GCACGAAGGA GGGCTCGAAA ATCTCTGCTA AAAACAAGAA	2366
GAAGGAACAG GGAACATGAT TAGTTTCGCT CGTATGGCAG AAAGTTTAGG AGTCCAGGCT	2426
AAACTTGCCC TTGCCTTCGC ACTCGTATTA TGTGTCGGGC TGATTGTTAC CGGCACGGGT	2486
TTCTACAGTG TACATACCTT GTCAGGGTTG GTGGGAATTC	2526

图 2p:

GAATTCGCG GTCGGCGAAA GTTGATGCGC TGTATCGTGG TGAAGATCAA TCCATGCTGC 60
 GTGACGAGGC CACACT GTG AGT TGG TCA GGG GGG GCT TAC TCG GCG TTT TCC 112
 Met Ser Trp Ser Gly Gly Ala Tyr Ser Ala Phe Ser
 1 5 10

GAC ACT GCG TTG GTT GCG GCA GTG CGC ACC CCC TGG ATT GAT TGC GGG 160
 Asp Thr Ala Leu Val Ala Ala Val Arg Thr Pro Trp Ile Asp Cys Gly
 15 20 25

GGT GCC CTG TCG CTG GTG TCG CCT ATC GAC TTA GGG GTA AAG GTC GCT 208
 Gly Ala Leu Ser Leu Val Ser Pro Ile Asp Leu Gly Val Lys Val Ala
 30 35 40

CGC GAA GTT CTG ATG CGT GCG TCG CTT GAA CCA CAA ATG GTC GAT AGC 256
 Arg Glu Val Leu Met Arg Ala Ser Leu Glu Pro Gln Met Val Asp Ser
 45 50 55 60

GTA CTC GCA GGC TCT ATG GCT CAA GCA AGC TTT GAT GCT TAC CTG CTC 304
 Val Leu Ala Gly Ser Met Ala Gln Ala Ser Phe Asp Ala Tyr Leu Leu
 65 70 75

CCG CGG CAC ATT GGC TTG TAC AGC GGT GTT CCC AAG TCG GTT CCG GCC 352
 Pro Arg His Ile Gly Leu Tyr Ser Gly Val Pro Lys Ser Val Pro Ala
 80 85 90

TTG GGG GTG CAG CGC ATT TGC GGC ACA GGC TTC GAA CTG CTT CGG CAG 400
 Leu Gly Val Gln Arg Ile Cys Gly Thr Gly Phe Glu Leu Leu Arg Gln
 95 100 105

GCC GGC GAG CAG ATT TCC CAA GGC GCT GAT CAC GTG CTG TGT GTC GCG 448
 Ala Gly Glu Gln Ile Ser Gln Gly Ala Asp His Val Leu Cys Val Ala-
 110 115 120

GCA GAG TCC ATG TCG CGT AAC CCC ATC GCG TCG TAT ACA CAC CGG GGC 496
 Ala Glu Ser Met Ser Arg Asn Pro Ile Ala Ser Tyr Thr His Arg Gly
 125 130 135 140

GGG TTC CGC CTC GGT GCG CCC GTT GAG TTC AAG GAT TTT TTG TGG GAG 544
 Gly Phe Arg Leu Gly Ala Pro Val Glu Phe Lys Asp Phe Leu Trp Glu
 145 150 155

GCA TTG TTT GAT CCT GCT CCA GGA CTC GAC ATG ATC GCT ACC GCA GAA 592
 Ala Leu Phe Asp Pro Ala Pro Gly Leu Asp Met Ile Ala Thr Ala Glu
 160 165 170

AAC CTG GGGGAGAGGC GGTTTGCGTA TTGGGCGCAT GCATAAAAAC TGTTGTAATT 648
 Asn Leu
 174

CATTAAGCAT TCTGCCGACA TGGAAGCCAT CACAAACGGC ATGATGAACC TGAATCGCCA 708

GCGGCATCAG CACCTTGTCG CCTTGCGTAT AATATTTGCC CATGGACGCA CACCGTGGAA 768

ACGGATGAAG GCACGAACCC AGTTGACATA AGCCTGTTTCG GTTCGTAAAC TGTAATGCAA 828

GTAGCGTATG CGCTCACGCA ACTGGTCCAG AACCTTGACC GAACGCAGCG GTGGTAACGG 888

CGCAGTGGCG GTTTTCATCG CTTGTTATGA CTGTTTTTTT GTACAGTCTA TGCCTCGGGC	948
ATCCAAGC AGCAAGCGCG TTACGCCGTG GGTGATGTTTG ATGTTATGGA GCAGCAACG	1007
ATG TTA CGC AGC AGC AAC GAT GTT ACG CAG CAG GGC AGT CGC CCT AAA Met Leu Arg Ser Ser Asn Asp Val Thr Gln Gln Gly Ser Arg Pro Lys 1 5 10 15	1055
ACA AAG TTA GGT GGC TCA AGT ATG GGC ATC ATT CGC ACA TGT AGG CTC Thr Lys Leu Gly Ser Ser Met Gly Ile Ile Arg Thr Cys Arg Leu 20 25 30	1103
GGC CCT GAC CAA GTC AAA TCC ATG CGG GCT GCT CTT GAT CTT TTC GGT Gly Pro Asp Gln Val Lys Ser Met Arg Ala Ala Leu Asp Leu Phe Gly 35 40 45	1151
CGT GAG TTC GGA GAC GTA GCC ACC TAC TCC CAA CAT CAG CCG GAC TCC Arg Glu Phe Gly Asp Val Ala Thr Tyr Ser Gln His Gln Pro Asp Ser 50 55 60	1199
GAT TAC CTC GGG AAC TTG CTC CGT AGT AAG ACA TTC ATC GCG CTT GCT Asp Tyr Leu Gly Asn Leu Leu Arg Ser Lys Thr Phe Ile Ala Leu Ala 65 70 75 80	1247
GCC TTC GAC CAA GAA GCG GTT GTT GGC GCT CTC GCG GCT TAC GTT CTG Ala Phe Asp Gln Glu Ala Val Val Gly Ala Leu Ala Ala Tyr Val Leu 85 90 95	1295
CCC AGG TTT GAG CAG CCG CGT AGT GAG ATC TAT ATC TAT GAT CTC GCA Pro Arg Phe Glu Gln Pro Arg Ser Glu Ile Tyr Ile Tyr Asp Leu Ala 100 105 110	1343
GTC TCC GGC GAG CAC CGG AGG CAG GGC ATT GCC ACC GCG CTC ATC AAT Val Ser Gly Glu His Arg Arg Gln Gly Ile Ala Thr Ala Leu Ile Asn 115 120 125	1391
CTC CTC AAG CAT GAG GCC AAC GCG CTT GGT GCT TAT GTG ATC TAC GTG Leu Leu Lys His Glu Ala Asn Ala Leu Gly Ala Tyr Val Ile Tyr Val 130 135 140	1439
CAA GCA GAT TAC GGT GAC GAT CCC GCA GTG GCT CTC TAT ACA AAG TTG Gln Ala Asp Tyr Gly Asp Asp Pro Ala Val Ala Leu Tyr Thr Lys Leu 145 150 155 160	1487
GGC ATA CGG GAA GAA GTG ATG CAC TTT GAT ATC GAC CCA AGT ACC GCC Gly Ile Arg Glu Glu Val Met His Phe Asp Ile Asp Pro Ser Thr Ala 165 170 175	1535
ACC TAA CAATTCGTTC AAGCCGAGAT CGGCTTCCCA TTG AGG GCG CAA GAG GAG Thr 177 Leu Arg Ala Gln Glu Glu 197 200	1589
AAA TGG ATT GAC CAA GAG ATC GTG GCT GTT ACG GAT GAA CAG TTC GAT Lys Trp Ile Asp Gln Glu Ile Val Ala Val Thr Asp Glu Gln Phe Asp 205 210 215	1637

TTA GAG GGC TAC AAC AGT CGA GCA ATT GAA CTG CCT CGG AAG GCA AAA Leu Glu Gly Tyr Asn Ser Arg Ala Ile Glu Leu Pro Arg Lys Ala Lys 220 225 230	1685
TTG TTG ATC GTG ACA GTC ATC CGC GGC CTA GCA GTC TTT GAA GCC CTT Leu Leu Ile Val Thr Val Ile Arg Gly Leu Ala Val Phe Glu Ala Leu 235 240 245 250	1733
TCC CGA TTG AAG CCT GTT CAT TCT GGC GGG GTG CAG ACT GCG GGC AAC Ser Arg Leu Lys Pro Val His Ser Gly Gly Val Gln Thr Ala Gly Asn 255 260 265	1781
AGC TGT GCC GTA GTG GAC GGC GCC GCG GCG GCT TTG GTG GCT CGA GAG Ser Cys Ala Val Val Asp Gly Ala Ala Ala Leu Val Ala Arg Glu 270 275 280	1829
TCG TCT GCG ACA CAG CCG GTC TTG GCT AGG ATA CTG GCT ACC TCC GTA Ser Ser Ala Thr Gln Pro Val Leu Ala Arg Ile Leu Ala Thr Ser Val 285 290 295	1877
GTC GGG ATC GAG CCC GAG CAT ATG GGG CTC GGC CCT GCG CCC GCG ATT Val Gly Ile Glu Pro Glu His Met Gly Leu Gly Pro Ala Pro Ala Ile 300 305 310	1925
CGC CTG CTG CTT GCG CGT AGT GAT CTT AGT TTG AGG GAT ATC GAC CTC Arg Leu Leu Leu Ala Arg Ser Asp Leu Ser Leu Arg Asp Ile Asp Leu 315 320 325 330	1973
TTT GAG ATA AAC GAG GCG CAG GCC GCC CAA GTT CTA GCG GTA CAG CAT Phe Glu Ile Asn Glu Ala Gln Ala Ala Gln Val Leu Ala Val Gln His 335 340 345	2021
GAA TTG GGT ATT GAG CAC TCA AAA CTT AAT ATT TGG GGC GGG GCC ATT Glu Leu Gly Ile Glu His Ser Lys Leu Asn Ile Trp Gly Gly Ala Ile 350 355 360	2069
GCA CTT GGA CAC CCG CTT GCC GCG ACC GGA TTG CGT CTC TGC ATG ACC Ala Leu Gly His Pro Leu Ala Ala Thr Gly Leu Arg Leu Cys Met Thr 365 370 375	2117
CTC GCT CAC CAA TTG CAA GCT AAT AAC TTT CGA TAT GGA ATT GCC TCG Leu Ala His Gln Leu Gln Ala Asn Asn Phe Arg Tyr Gly Ile Ala Ser 380 385 390	2165
GCA TGC ATT GGT GGG GGA CAG GGG ATG GCG GTT CTT TTA GAG AAT CCC Ala Cys Ile Gly Gly Gln Gln Gly Met Ala Val Leu Leu Glu Asn Pro 395 400 405 410	2213
CAC TTC GGT TCG TCC TCT GCA CGA AGT TCG ATG ATT AAC AGA GTT GAC His Phe Gly Ser Ser Ser Ala Arg Ser Ser Met Ile Asn Arg Val Asp 415 420 425	2261
CAC TAT CCA CTG AGC TAA CGGGCATCTC CTTTGTGCT TTGAGGTGGC His Tyr Pro Leu Ser 430 431	2309

GCACGAAGGA GGGCTCGAAA ATCTCTGCTA AAAACAAGAA GAAGGAACAG GGAACATGAT	2369
TAGTTTCGCT CGTATGGCAG AAAGTTTAGG AGTCCAGGCT AAAGTTGCCC TTGCCTTCGC	2429
ACTCGTATTA TGTGTCGGGC TGATTGTTAC CGGCACGGGT TTCTACAGTG TACATACCTT	2489
GTCAGGGTTG GTGGGAATTC	2509

图 2q:

GAATTCGCG GTCGGCGAAA GTTGATGCGC TGTATCGTGG TGAAGATCAA TCCATGCTGC	60
GTGACGAGGC CACACT GTG AGT TGG TCA GGG GGG GCT TAC TCG GCG TTT TCC	112
Met Ser Trp Ser Gly Gly Ala Tyr Ser Ala Phe Ser	
1 5 10	
GAC ACT GCG TTG GTT GCG GCA GTG CGC ACC CCC TGG ATT GAT TGC GGG	160
Asp Thr Ala Leu Val Ala Ala Val Arg Thr Pro Trp Ile Asp Cys Gly	
15 20 25	
GGT GCC CTG TCG CTG GTG TCG CCT ATC GAC TTA GGG GTA AAG GTC GCT	208
Gly Ala Leu Ser Leu Val Ser Pro Ile Asp Leu Gly Val Lys Val Ala	
30 35 40	
CGC GAA GTT CTG ATG CGT GCG TCG CTT GAA CCA CAA ATG GTC GAT AGC	256
Arg Glu Val Leu Met Arg Ala Ser Leu Glu Pro Gln Met Val Asp Ser	
45 50 55 60	
GTA CTC GCA GGC TCT ATG GCT CAA GCA AGC TTT GAT GCT TAC CTG CTC	304
Val Leu Ala Gly Ser Met Ala Gln Ala Ser Phe Asp Ala Tyr Leu Leu	
65 70 75	
CCG CGG CAC ATT GGC TTG TAC AGC GGT GTT CCC AAG TCG GTT CCG GCC	352
Pro Arg His Ile Gly Leu Tyr Ser Gly Val Pro Lys Ser Val Pro Ala	
80 85 90	
TTG GGG GTG CAG CGC ATT TGC GGC ACA GGC TTC GAA CTG CTT CGG CAG	400
Leu Gly Val Gln Arg Ile Cys Gly Thr Gly Phe Glu Leu Leu Arg Gln	
95 100 105	
GCC GGC GAG CAG ATT TCC CAA GGC GCT GAT CAC GTG CTG TGT GTC GCG	448
Ala Gly Glu Gln Ile Ser Gln Gly Ala Asp His Val Leu Cys Val Ala	
110 115 120	
GCA GAG TCC ATG TCG CGT AAC CCC ATC GCG TCG TAT ACA CAC CGG GGC	496
Ala Glu Ser Met Ser Arg Asn Pro Ile Ala Ser Tyr Thr His Arg Gly	
125 130 135 140	
GGG TTC CGC CTC GGT GCG CCC GTT GAG TTC AAG GAT TTT TTG TGG GAG	544
Gly Phe Arg Leu Gly Ala Pro Val Glu Phe Lys Asp Phe Leu Trp Glu	
145 150 155	
GCA TTG TTT GAT CCT GCT CCA GGA CTC GAC ATG ATC GCT ACC GCA GAA	592
Ala Leu Phe Asp Pro Ala Pro Gly Leu Asp Met Ile Ala Thr Ala Glu	
160 165 170	
AAC CTG GCG CGC A TTG AGG GCG CAA GAG GAG AAA TGG ATT GAC CAA GAG	641
Asn Leu Ala Arg Leu Arg Ala Gln Glu Glu Lys Trp Ile Asp Gln Glu	
175 176 197 200 205	
ATC GTG GCT GTT ACG GAT GAA CAG TTC GAT TTA GAG GGC TAC AAC AGT	689
Ile Val Ala Val Thr Asp Glu Gln Phe Asp Leu Glu Gly Tyr Asn Ser	
210 215 220	
CGA GCA ATT GAA CTG CCT CGG AAG GCA AAA TTG TTG ATC GTG ACA GTC	737
Arg Ala Ile Glu Leu Pro Arg Lys Ala Lys Leu Leu Ile Val Thr Val	
225 230 235 240	

ATC CGC GGC CTA GCA GTC TTT GAA GCC CTT TCC CGA TTG AAG CCT GTT Ile Arg Gly Leu Ala Val Phe Glu Ala Leu Ser Arg Leu Lys Pro Val 245 250 255	785
CAT TCT GGC GGG GTG CAG ACT GCG GGC AAC AGC TGT GCC GTA GTG GAC His Ser Gly Gly Val Gln Thr Ala Gly Asn Ser Cys Ala Val Val Asp 260 265 270	833
GGC GCC GCG GCG GCT TTG GTG GCT CGA GAG TCG TCT GCG ACA CAG CCG Gly Ala Ala Ala Leu Val Ala Arg Glu Ser Ser Ala Thr Gln Pro 275 280 285	881
GTC TTG GCT AGG ATA CTG GCT ACC TCC GTA GTC GGG ATC GAG CCC GAG Val Leu Ala Arg Ile Leu Ala Thr Ser Val Val Gly Ile Glu Pro Glu 290 295 300	929
CAT ATG GGC CTC GGC CCT GCG CCC GCG ATT CGC CTG CTG CTT GCG CGT His Met Gly Leu Gly Pro Ala Pro Ala Ile Arg Leu Leu Leu Ala Arg 305 310 315 320	977
AGT GAT CTT AGT TTG AGG GAT ATC GAC CTC TTT GAG ATA AAC GAG GCG Ser Asp Leu Ser Leu Arg Asp Ile Asp Leu Phe Glu Ile Asn Glu Ala 325 330 335	1025
CAG GCC GCC CAA GTT CTA GCG GTA CAG CAT GAA TTG GGT ATT GAG CAC Gln Ala Ala Gln Val Leu Ala Val Gln His Glu Leu Gly Ile Glu His 340 345 350	1073
TCA AAA CTT AAT ATT TGG GGC GGG GCC ATT GCA CTT GGA CAC CCG CTT Ser Lys Leu Asn Ile Trp Gly Gly Ala Ile Ala Leu Gly His Pro Leu 355 360 365	1121
GCC GCG ACC GGA TTG CGT CTC TGC ATG ACC CTC GCT CAC CAA TTG CAA Ala Ala Thr Gly Leu Arg Leu Cys Met Thr Leu Ala His Gln Leu Gln 370 375 380	1169
GCT AAT AAC TTT CGA TAT GGA ATT GCC TCG GCA TGC ATT GGT GGG GGA Ala Asn Asn Phe Arg Tyr Gly Ile Ala Ser Ala Cys Ile Gly Gly Gly 385 390 395 400	1217
CAG GGG ATG GCG GTT CTT TTA GAG AAT CCC CAC TTC GGT TCG TCC TCT Gln Gly Met Ala Val Leu Leu Glu Asn Pro His Phe Gly Ser Ser Ser 405 410 415	1265
GCA CGA AGT TCG ATG ATT AAC AGA GTT GAC CAC TAT CCA CTG AGC TAA Ala Arg Ser Ser Met Ile Asn Arg Val Asp His Tyr Pro Leu Ser 420 425 430 431	1313
CGGGCATCTC CTTTGTGCT TTGAGGTGGC GCACGAAGGA GGGCTCGAAA ATCTCTGCTA	1373
AAAACAAGAA GAAGGAACAG GGAACATGAT TAGTTTCGCT CGTATGGCAG AAAGTTTAGG	1433
AGTCCAGGCT AAACCTGCCC TTGCCTTCGC ACTCGTATTA TGTGTCGGGC TGATTGTTAC	1493
CGGCACGGGT TTCTACAGTG TACATACCTT GTCAGGGTTG GTGGGAATTC	1543

图 2r:

序列 1

CTGCAGCCAG	GGCTGAAAAG	GAGGGATTCA	GTGAGGTCAT	GAAGGGAGGG	GACGGCGCCT	60
GGCTCCAATT	GCTCGATGGC	GCCGCGATTG	AGTGTCTTGG	GCGCGGTCTT	GGAGAGTTCTG	120
GCTAGGGAGA	TAAATTTGCT	GGCCATGGTG	GCGGCCCTTG	ATGGGTTGGA	TGATTTTCTG	180
CATTCTGCAT	CATGAAATTC	ATGAAATCAT	CACTTTTCGG	GGGGTGGGTG	CACGGGATTG	240
AAGGTTGCTA	GGAGAGTGCA	TTGCTCGTAA	GCCCAGGAAG	CACGCGGGTT	TCAGGATGGT	300
GCATGGAAAT	GGCATGAGCT	TTGCTGGATA	TGATTAGAGA	CATTAACTAT	TTTGGCGGAA	360
TGGAAGCAGC	ATTCCTCGCC	CGGTAGAGCG	GTAACCGCGA	CATTCAGGAC	CGTAAAAAGG	420
AAAGAGCATG	CAACTGACCA	ACAAGAAAAT	CGTCGTCACC	GGAGTGTCTT	CCGGTATCGG	480
TGCCGAAACT	GCCCGCGTTC	TGCGCTCTCA	CGGCGCCACA	GTGATTGGCG	TAGATCGCAA	540
CATGCCGAGC	CTGACTCTGG	ATGCTTTCGT	TCAGGCTGAC	CTGAGCCATC	CTGAAGGCAT	600
CGATAAGGCC	ATCGGGACAG	CAAGCGAACC	GGAATTGCCA	GCTGGGGCGC	CCTCTGGTAA	660
GGTTGGGAAG	CCCTGCAAAAG	TAAACTGGAT	GGCTTTCCTT	CCGCCAAGGA	TCTGATGGCG	720
CAGGGGATCA	AGATCTGATC	AAGAGACAGG	ATGAGGATCG	TTTCGCATGA	TTGAACAAGA	780
TGGATTGCAC	GCAGGTTCTC	CGGCCGCTTG	GGTGGAGAGG	CTATTCCGGCT	ATGACTGGGC	840
ACAACAGACA	ATCGGCTGCT	CTGATGCCGC	CGTGTTCGGG	CTGTCAGCGC	AGGGGCGCCC	900
GGTCTTTTTT	GTCAAGACCG	ACCTGTCCGG	TGCCCTGAAT	GAAGTGCAGG	ACGAGGCAGC	960
GCGGCTATCG	TGGCTGGCCA	CGACGGGCGT	TCCTTGCGCA	GCTGTGCTCG	ACGTTGTCAC	1020
TGAAGCGGGA	AGGGACTGGC	TGCTATTGGG	CGAAGTGCCG	GGGCAGGATC	TCCTGTCATC	1080
TCACCTTGCT	CCTGCCGAGA	AAGTATCCAT	CATGGCTGAT	GCAATGCGGC	GGCTGCATAC	1140
GCTTGATCCG	GCTACCTGCC	CATTCGACCA	CCAAGCGAAA	CATCGCATCG	AGCGAGCAGC	1200
TACTCGGATG	GAAGCCGGTC	TTGTCGATCA	GGATGATCTG	GACGAAGAGC	ATCAGGGGCT	1260
CGCGCCAGCC	GAAGTGTTCG	CCAGGCTCAA	GGCGCGCATG	CCCGACGGCG	AGGATCTCGT	1320
CGTGACCCAT	GGCGATGCCT	GCTTGCCGAA	TATCATGGTG	GAAAATGGCC	GCTTTTCTGG	1380
ATTCATCGAC	TGTGGCCGGC	TGGGTGTGGC	GGACCGCTAT	CAGGACATAG	CGTTGGCTAC	1440
CCGTGATATT	GCTGAAGAGC	TTGGCGGCGA	ATGGGCTGAC	CGCTTCCTCG	TGCTTTACGG	1500
TATCGCCGCT	CCCGATTTCG	AGCGCATCGC	CTTCTATCGC	CTTCTTGACG	AGTTCTTCTG	1560
AGCGGGACTC	TGGGGTTTGA	AATGACCGAC	CAAGCGACGC	CCTGGCCGCG	GTGATTGCAT	1620
TCATGTGTGC	TGAGGAGTCA	CGTTGGATCA	ACGGCATAAA	TATTCCAGTG	GACGGAGGTT	1680
TGGCATCGAC	CTACGTGTAA	GTTTCGTGGAC	GCCCTTTGCA	CGCGCACTAT	ATCTCTATGC	1740
AGCAGCTGAA	AGCAGCTTTG	GTTTTGATCG	GAGGTAGCGG	GCGGAAAGGT	GCAGAATGTC	1800
TAAATAATAA	AGGATTCTTG	TGAAGCTTTA	GTTGTCCGTA	AACGAAAATA	AAAATAAAGA	1860
GGAATGATAT	GAAAGCAAGT	AGATCAGTCT	GCACTTTCAA	AATAGCTACC	CTGGCAGGCG	1920
CCATTTATGC	AGCGCTGCCA	ATGTCAGCTG	CAAACTCGAT	GCAGCTGGAT	GTAGGTAGCT	1980
CGGATTGGAC	GGTGCCTTGG	GGACAACACC	CTCAAGTATA	GCCTTGCCCTC	TCGCCTGAAT	2040
GAGCAAGACT	CAAGTCTGAC	AAATGCGCCG	ACTGTCAATG	GTTATATCCG	GATATTCAAA	2100
GTCAGGGTGA	TCGTAACTTT	GACCGGGGGC	TTGGTATCCA	ATCGTCTCGA	TATTCTGGCT	2160
GCAG						2164

序列 2

CTGCAGCCAG	GGCTGAAAAG	GAGGGATTCA	GTGAGGTCAT	GAAGGGAGGG	GACGGCGCCT	60
GGCTCCAATT	GCTCGATGGC	GCCGCGATTG	AGTGTCTTGG	GCGCGGTCTT	GGAGAGTTCG	120
GCTAGGGAGA	TAAATTTGCT	GGCCATGGTG	GCGGCCCTTG	ATGGGTTGGA	TGATTTTCTG	180
CATTCTGCAT	CATGAAATTC	ATGAAATCAT	CACTTTTCGG	GGGGTGGGTG	CACGGGATTG	240
AAGGTTGCTA	GGAGAGTGCA	TTGCTCGTAA	GCCCAGGAAG	CACGCGGGTT	TCAGGATGGT	300
GCATGGAAAT	GGCATGAGCT	TTGCTGGATA	TGATTAGAGA	CATTAACAT	TTTGGCGGAA	360
TGGAAGCACG	ATTCCTCGCC	CGGTAGAGCG	GTAACCGCGA	CATTTCAGGAC	CGTAAAAAGG	420
AAAGAGCATG	CAACTGACCA	ACAAGAAAAT	CGTCGTCACC	GGAGTGTCTT	CCGGTATCGG	480
TGCCGAAACT	GCCCGCGTTC	TGCGCTCTCA	CGGCGCCACA	GTGATTGGCG	TAGATCGCAA	540
CATGCCGAGC	CTGACTCTGG	ATGCTTTCGT	TCAGGCTGAC	CTGAGCCATC	CTGAGGGGAG	600
AGGCGGTTTG	CGTATTGGGC	GCATGCATAA	AAACTGTTGT	AATTCATTAA	GCATTCTGCC	660
GACATGGAAG	CCATCACAAA	CGGCATGATG	AACCTGAATC	GCCAGCGGCA	TCAGCACCTT	720
GTCGCCTTGC	GTATAATATT	TGCCCATGGA	CGCACACCGT	GGAACCGGAT	GAAGGCACGA	780
ACCCAGTTGA	CATAAGCCTG	TTCGGTTCGT	AAACTGTAAT	GCAAGTAGCG	TATGCGCTCA	840
CGCAACTGGT	CCAGAACCTT	GACCGAACGC	AGCGGTGGTA	ACGGCGCAGT	GGCGGTTTTT	900
ATGGCTTGTT	ATGACTGTTT	TTTTGTACAG	TCTATGCCCT	GGGCATCCAA	GCAGCAAGCG	960
CGTTACGCCG	TGGGTCGATG	TTTGATGTTA	TGGAGCAGCA	ACGATGTTAC	GCAGCAGCAA	1020
CGATGTTACG	CAGCAGGGCA	GTGCCCCTAA	AACAAAGTTA	GGTGGCTCAA	GTATGGGCAT	1080
CATTTCGCACA	TGTAGGCTCG	GCCCTGACCA	AGTCAAATCC	ATGCGGGCTG	CTCTTGATCT	1140
TTTCGGTTCGT	GAGTTCGGAG	ACGTAGCCAC	CTACTCCCAA	CATCAGCCGG	ACTCCGATTA	1200
CCTCGGGAAC	TTGCTCCGTA	GTAAGACATT	CATCGCGCTT	GCTGCCTTCG	ACCAAGAAGC	1260
GGTTGTTGGC	GCTCTCGCGG	CTTACGTTCT	GCCCAGGTTT	GAGCAGCCGC	GTAGTGAGAT	1320
CTATATCTAT	GATCTCGCAG	TCTCCGGCGA	GCACCGGAGG	CAGGGCATTG	CCACCGCGCT	1380
CATCAATCTC	CTCAAGCATG	AGGCCAACGC	GCTTGGTGCT	TATGTGATCT	ACGTGCAAGC	1440
AGATTACGGT	GACGATCCCG	CAGTGGCTCT	CTATACAAAG	TTGGGCATAC	GGGAAGAAGT	1500
GATGCACTTT	GATATCGACC	CAAGTACCGC	CACCTAACAA	TTCGTTCAAG	CCGAGATCGG	1560
CTTCCCTGAT	TGCATTTCATG	TGTGCTGAGG	AGTCACGTTG	GATCAACGGC	ATAAATATTC	1620
CAGTGGACGG	AGGTTTGGCA	TCGACCTACG	TGTAAGTTCG	TGGACGCCCT	TTGCACGCCG	1680
ACTATATCTC	TATGCAGCAG	CTGAAAGCAG	CTTTGGTTTT	GATCGGAGGT	AGCGGGCGGA	1740
AAGGTGCAGA	ATGTCTAAAT	AATAAAGGAT	TCTTGTGAAG	CTTTAGTTGT	CCGTAAACGA	1800
AAATAAAAAT	AAAGAGGAAT	GATATGAAAG	CAAGTAGATC	AGCTGCAACT	TTCAAAATAG	1860
CTACCCTGGC	AGGCGCCATT	TATGCAGCGC	TGCCAATGTC	AGCTGCAAAC	TCGATGCAGC	1920
TGGATGTAGG	TAGCTCGGAT	TGGACGGTGC	GTTGGGGACA	ACACCCCTCA	GTATAGCCCT	1980
GCCTCTCGCC	TGAATGAGCA	AGACTCAAGT	CTGACAAATG	CGCCGACTGT	CAATGGTTAT	2040
ATCCGGATAT	TCAAAGTCAG	GGTGATCGTA	ACTTTGACCG	GGGGCTTGGT	ATCCAATCGT	2100
CTCGATATTC	TGGCTGCAG					2119

序列 3

CTGCAGCCAG	GGCTGAAAAG	GAGGGATTCA	GTGAGGTCAT	GAAGGGAGGG	GACGGCGCCT	60
GGCTCCAATT	GCTCGATGGC	GCCGCGATTG	AGTGTCTTGG	GCGCGGTCTT	GGAGAGTTCG	120
GCTAGGGAGA	TAAATTTGCT	GGCCATGGTG	GCGGCCCCCTG	ATGGGTGGGA	TGATTTTCTG	180
CATTCTGCAT	CATGAAATTC	ATGAAATCAT	CACTTTTCGG	GGGGTGGGTG	CACGGGATTG	240
AAGGTTGCTA	GGAGAGTGCA	TTGCTCGTAA	GCCCAGGAAG	CACGCGGGTT	TCAGGATGGT	300
GCATGGAAAT	GGCATGAGCT	TTGCTGGATA	TGATTAGAGA	CATTAACTAT	TTTGGCGGAA	360
TGGAAGCACG	ATTCCCTCGCC	CGGTAGAGCG	GTAACCGCGA	CATTCAGGAC	CGTAAAAAGG	420
AAAGAGCATG	CAACTGACCA	ACAAGAAAAT	CGTCGTCACC	GGAGTGTCCCT	CCGGTATCGG	480
TGCCGAAACT	GCCCGCGTTC	TGCGCTCTCA	CGGCGCCACA	GTGATTGGCG	TAGATCGCAA	540
CATGCCGAGC	CTGACTCTGG	ATGCTTTTCGT	TCAGGCTGAC	CTGAGCCATC	CTGAAGGCAT	600
CGATCAACGG	CATAAATATT	CCAGTGGACG	GAGGTTTGGC	ATCGACCTAC	GTGTAAGTTC	660
GTGGACGCCC	TTTGCACGCG	CACATATATCT	CTATGCAGCA	GCTGAAAGCA	GCTTTGGTTT	720
TGATCGGAGG	TAGCGGGCGG	AAAGGTGCAG	AATGTCTAAA	TAATAAAGGA	TTCTTGTGAA	780
GCTTTAGTTG	TCCGTAAACG	AAAATAAAAA	TAAAGAGGAA	TGATATGAAA	GCAAGTAGAT	840
CAGTCTGCAC	TTTCAAATA	GCTACCCTGG	CAGGCGCCAT	TTATGCAGCG	CTGCCAATGT	900
CAGCTGCAAA	CTCGATGCAG	CTGGATGTAG	GTAGCTCGGA	TTGGACGGTG	CGTTGGGGAC	960
AACACCCTCA	AGTATAGCCT	TGCCTCTCGC	CTGAATGAGC	AAGACTCAAG	TCTGACAAAT	1020
GCGCCGACTG	TCAATGGTTA	TATCCGGATA	TTCAAAGTCA	GGGTGATCGT	AACTTTGACC	1080
GGGGGCTTGG	TATCCAATCG	TCTCGATATT	CTGGCTGCAG			1120

序列 4

GAATTCGCGG	TATCGCCCGG	TTCTATCAGC	GGGCCGCTTT	CGAAAGTCAT	GGTGTTAGCC	60
GGTAGGGTCT	TTTTCTTGGC	CATGCTTGTT	GCCTGAACCT	TCGTTGACAT	AGGGCAGAGG	120
TGCGTTTGCC	GCTTCGCTTC	GCGATGAACC	GCATCGAGAT	GCTGAGGTCA	GGATTTTTCC	180
TTAACTCGCG	TAAGCATTCT	GTCATTTTTT	TGGTGGCTTT	GAACAGCCTG	ATGAAAGGTG	240
GTCTCGCCCT	TTGAGGCCGA	TTCTTGGGCG	CTTGGCGGCG	TCGAAGCGAT	GCTCCACTAC	300
CGATTAAGAT	AATTAAAATA	AGGAAACCGC	ATGGTTTCTT	ATGTGAATTT	GTCTGGCATA	360
CTCCAGCTCA	AGGGCAATTT	TTGGGCTATT	GGCTGAGCAG	TTGCCCTCTAT	ATGGTTATTTC	420
AGAATAACAA	TTGACTCCTC	AGGAGGTCAG	CGATGAGCAT	TCCTGGTTTG	AATGGTGCCC	480
CGGTGCGGAG	TGAGCAGCTG	GGCTCGGCTC	TTGATCGCAT	GAAGAAGCCG	CACCTGGAGC	540
AGGGGCCCTGC	AAACTTGGAG	CTGCGTCTGA	GTAGGCTGGA	TCGTGCGATT	GCAATGCTTC	600
TGGAAAATCG	TGAAGCAATT	GCCGACGCGG	TTTCTGCTGA	CTTTGGCAAT	CGCAGCCGTG	660
AGCAAACACT	GCTTTGCGAC	ATTGCTGGCT	CGGTGGCAAG	CCTGAAGGAT	AGCCGCGAGC	720
ACGTGGCCAA	ATGGATGGAG	CCCGAACATC	ACAAGGCGAT	GTTTCCAGGG	GCGGAGGCAC	780
GCGTTGAGTT	TCAGCCGCTG	GGTGTCGTTG	GGGTCAATTAG	TCCCTGGAAC	TTCCCTATCG	840
TACTGGCCTT	TGGGCCGCTG	GCCGGCATAT	TCGCAGCAGG	TAATCGCGCC	ATGCTCAAGC	900
CGTCCGAGCT	TACCCCGCGG	ACTTCTGCCC	TGCTTGCGGA	GCTAATTGCT	CGTFACTTTC	960
ATGAACTGA	GCTGACTACA	GTGCTGGGCG	ACGCTGAAGT	CGGTGCGCTG	TTCACTGCTC	1020
AGCCTTTCGA	TCATCTGATC	TTACCCGGCG	GCACTGCCGT	GGCCAAGCAC	ATCATGCGTG	1080
CCGCGGCGGA	TAACCTAGTG	CCCGTTACCC	TGGAATTGGG	TGGCAAATCG	CCCGTGATCG	1140
TTTCCCGCAG	TGCAGATATG	GCGGACGTTG	CACAACGGGT	GTTGACGGTG	AAAACCTTCA	1200
ATGCCGGGCA	AATCTGTCTG	GCACCGGACT	ATGTGCTGCT	GCCGGAAGGG	ACAGCAAGCG	1260
AACCGGAATT	GCCAGCTGGG	GCGCCCTCTG	GTAAGGTTGG	GAAGCCCTGC	AAAGTAAACT	1320
GGATGGCTTT	CTTGCCGCCA	AGGATCTGAT	GGCGCAGGGG	ATCAAGATCT	GATCAAGAGA	1380
CAGGATGAGG	ATCGTTTCGC	ATGATTGAAC	AAGATGGATT	GCACGCAGGT	TCTCCGGCCG	1440
CTTGGGTGGA	GAGGCTATTG	GGCTATGACT	GGGCACAACA	GACAATCGGC	TGCTCTGATG	1500
CCGCCGTGTT	CCGGCTGTCA	GCGCAGGGGC	GCCCGGTTCT	TTTGTCAAG	ACCGACCTGT	1560
CCGGTGCCCT	GAATGAACTG	CAGGACGAGG	CAGCGCGGCT	ATCGTGGCTG	GCCACGACGG	1620
GCGTTCCCTG	GCGAGCTGTG	CTCGACGTTG	TCACTGAAGC	GGGAAGGGAC	TGGCTGCTAT	1680
TGGGCGAAGT	GCCGGGGCAG	GATCTCCTGT	CATCTCACCT	TGCTCCTGCC	GAGAAAGTAT	1740
CCATCATGGC	TGATGCAATG	CGGCGGCTGC	ATACGCTTGA	TCCGGCTACC	TGCCCATTCT	1800
ACCACCAAGC	GAAACATCGC	ATCGAGCGAG	CACGTACTCG	GATGGAAGCC	GGTCTTGTCG	1860
ATCAGGATGA	TCTGGACGAA	GAGCATCAGG	GGCTCGCGCC	AGCCGAAGTG	TTCGCCAGGC	1920
TCAAGGCGCG	CATGCCCCGAC	GGCGAGGATC	TCGTGCTGAC	CCATGGCGAT	GCCTGCTTGC	1980
CGAATATCAT	GGTGGAAAAT	GGCCGCTTTT	CTGGATTTCAT	CGACTGTGGC	CGGCTGGGTG	2040
TGGCGGACCG	CTATCAGGAC	ATAGCGTTGG	CTACCCGTGA	TATTGCTGAA	GAGCTTGGCG	2100
GCGAATGGGC	TGACCGCTTC	CTCGTGCTTT	ACGGTATCGC	CGCTCCCGAT	TCGCAGCGCA	2160

TCGCCTTCTA	TCGCCTTCTT	GACGAGTTCT	TCTGAGCGGG	ACTCTGGGGT	TCGAAATGAC	2220
CGACCAAGCG	ACGCCCCGCA	TGCCAAGCCT	GTTCTCGTGC	AAAGTCCTGT	GGGTGAGTCG	2280
AACTTGGCGA	TGCGCGCACC	CTACGGAGAA	GCGATCCACG	GACTGCTCTC	TGTCTCCTTT	2340
TCAACGGAGT	GTTAGAACCG	TTGGTAGTGG	TTTTGGACGG	GCCCAGGAGC	ATGCGCTTCT	2400
GGGCCCCGTT	CTTGAGTATT	CATTGGATAG	TCACCGGTGG	TAGCTTCGAG	CCTGCACAGC	2460
TGATGAGCAC	CCTGGAAGGC	GCGCTGTACG	CGGACGACTG	GGTTCATCTT	CGCCATTCAT	2520
GACGGAACCT	CGTTCCCCAG	TACCGCGATG	ACTATTTTGC	CTCTTCCGAT	GTCCGATTCC	2580
ACGCCGCCCT	ACGCTAAGCG	GGGGCGGGGG	CGCCCGCATC	CCAGCCCAGA	CAGCAACAAA	2640
TGAGTAGGCT	CTTGGATGCC	GCGGCGGCTG	AGATTGGTAA	CGGCAATTTC	GTCAATGTGA	2700
CGATGGATTG	GATTGCCCCG	GCTGCCGGCG	TCTCAAAAAA	AACGCTGTAC	GTCTTGGTGG	2760
CGAGCAAGGA	AGAACTCATT	TCCCGGTTAG	TGGCTCGAGA	CATGTCCAAC	CTTGAGGAAT	2820
TC						2822

序列 5

GAATTCCGCG	TATCGCCCGG	TTCTATCAGC	GGGCCGCTTT	CGAAAGTCAT	GGTGTAGCC	60
GGTAGGGTCT	TTTTCTTGGC	CATGCTTGTT	GCCTGAACCT	TCGTTGACAT	AGGGCAGAGG	120
TGCGTTTGCC	GCTTCGCTTC	GCGATGAACC	GCATCGAGAT	GCTGAGGTCA	GGATTTTTCC	180
TTAACTCGCG	TAAAGCATCT	GTCATTTTTT	TGGTGGCTTT	GAACAGCCTG	ATGAAAGGTG	240
GTCTCGCCCT	TTGAGGCCGA	TTCTTGGGCG	CTTGGCGGCG	TCGAAGCGAT	GCTCCACTAC	300
CGATTAAGAT	AATTAAAATA	AGGAAACCGC	ATGGTTTCTT	ATGTGAATTT	GTCTGGCATA	360
CTCCAGCTCA	AGGGCAATTT	TTGGGCTATT	GGCTGAGCAG	TTGCCTCTAT	ATGGTTATTC	420
AGAATAACAA	TTGACTCCTC	AGGAGGTCAG	CGATGAGCAT	TCTTGGTTTG	AATGGTGCCC	480
CGGTCCGAGC	TGAGCAGCTG	GGCTCGGCTC	TTGATCGCAT	GAAGAAGGCG	CACCTGGAGC	540
AGGGGCCTGC	AAACTTGGAG	CTGCGTCTGA	GTAGGCTGGA	TCGTGCGATT	GCAATGCTTC	600
TGGAAAATCG	TGAAGCAATT	GCCGACGCGG	TTTCTGCTGA	CTTTGGCAAT	CGCAGCCGTG	660
AGCAAACACT	GCTTTGCGAC	ATTGCTGGCT	CGGTGGCAAG	CCTGAAGGAT	AGCCGCGAGC	720
ACGTGGCCAA	ATGGATGGAG	CCCGAACATC	ACAAGGCGAT	GTTTCCAGGG	GCGGAGGCAC	780
GCGTTGAGTT	TCAGCCGCTG	GGTGTCTGTT	GGGTCATTAG	TCCCTGGAAC	TTCCCTATCG	840
TACTGGCCTT	TGGGCCGCTG	GCCGGCATAT	TCGCAGCAGG	TAATCGCGCC	ATGCTCAAGC	900
CGTCCGAGCT	TACCCCGCGG	ACTTCTGCCC	TGCTTGCGGA	GCTAATTGCT	CGTTACTTCG	960
ATGAAACTGA	GCTGACTACA	GTGCTGGGCG	ACGCTGAAGT	CGGTGCGCTG	TTCAGTGCTC	1020
AGCCTTTTCA	TCATCTGATC	TTACCCGGCG	GCACTGCCGT	GGCCAAGCAC	ATCATGCGTG	1080
CCGCGGCGGA	TAACCTAGTG	CCCGTTACCC	TGGAATTGGG	TGGCAAATCG	CCGGTGATCG	1140
TTTCCCGCAG	TGCAGATATG	GCGGACGTTG	CACAACGGGT	GTTGACGGTG	AAAACCTTCA	1200
ATGCCGGGCA	AATCTGTCTG	GCACCGGACT	ATGTGCTGGG	GGAGAGGCGG	TTTGCGTATT	1260
GGGCGCATGC	ATAAAAACTG	TTGTAATTCA	TTAAGCATTC	TGCCGACATG	GAAGCCATCA	1320
CAAACGGCAT	GATGAACCTG	AATCGCCAGC	GGCATCAGCA	CCTTGTCGCC	TTGCGTATAA	1380
TATTTGCCCA	TGGACGCACA	CCGTGGAAAC	GGATGAAGGC	ACGAACCCAG	TTGACATAAG	1440
CCTGTTCCGT	TCGTAAACTG	TAATGCAAGT	AGCGTATGCG	CTCACGCAAC	TGGTCCAGAA	1500
CCTTGACCGA	ACGCAGCGGT	GGTAACGGCG	CAGTGGCGGT	TTTCATGGCT	TGTTATGACT	1560
GTTTTTTTGT	ACAGTCTATG	CCTCGGGCAT	CCAAGCAGCA	AGCGCGTTAC	GCCGTGGGTC	1620
GATGTTTGAT	GTTATGGAGC	AGCAACGATG	TTACGCAGCA	GCAACGATGT	TACGCAGCAG	1680
GGCAGTCGCC	CTAAAACAAA	GTTAGGTGGC	TCAAGTATGG	GCATCATTCG	CACATGTAGG	1740
CTCGGCCCTG	ACCAAGTCAA	ATCCATGCCG	GCTGCTCTTG	ATCTTTTCGG	TCGTGAGTTC	1800
GGAGACGTAG	CCACCTACTC	CCAACATCAG	CCGGACTCCG	ATTACCTCGG	GAACCTTGCTC	1860
CGTAGTAAGA	CATTCATCGC	GCTTGCTGCC	TTGACCAAG	AAGCGGTTGT	TGGCGCTCTC	1920
GCGGCTTACG	TTCTGCCCAG	GTTTGAGCAG	CCGCGTAGTG	AGATCTATAT	CTATGATCTC	1980
GCAGTCTCCG	GCGAGCACCG	GAGGCAGGGC	ATTGCCACCG	CGCTCATCAA	TCTCCTCAAG	2040
CATGAGGCCA	ACGCGCTTGG	TGCTTATGTG	ATCTACGTGC	AAGCAGATTA	CGGTGACGAT	2100
CCCGCAGTGG	CTCTCTATAC	AAAGTTGGGC	ATACGGGAAG	AAGTGATGCA	CTTTGATATC	2160

GACCCAAGTA	CCGCCACCTA	ACAATTGCTT	CAAGCCGAGA	TCGGCTTCCC	TGCAAAGTCC	2220
TGTGGGTGAG	TCGAACTTGG	CGATGCGCGC	ACCCTACGGA	GAAGCGATCC	ACGGACTGCT	2280
CTCTGTCCTC	CTTTCAACGG	AGTGTTAGAA	CCGTTGGTAG	TGGTTTTGGA	CGGGCCCAGG	2340
AGCATGCGCT	TCTGGGCCCC	TTTCTTGAGT	ATTCAATTGGA	TAGTCACGCG	TGGTAGCTTC	2400
GAGCCTGCAC	AGCTGATGAG	CACCCGTGAA	GGCGCGCTGT	ACGCGGACGA	CTGGGTTCAT	2460
CTTCGCCATT	CATGACGGAA	CTCCGTTCCC	CAGTACCGCG	ATGACTATTT	TGCCTCTTCC	2520
GATGTCCGAT	TCCACGCCGC	CTGACGCTAA	GCGGGGGCGG	GGGCGCCCGC	ATCCCAGCCC	2580
AGACAGCAAC	AAATGAGTAG	GCTCTTGGAT	GCCGCGGCGG	CTGAGATTGG	TAACGGCAAT	2640
TTCGTCAATG	TGACGATGGA	TTCGATTGCC	CGTGCTGCCG	GCGTCTCAAA	AAAAACGCTG	2700
TACGTCTTGG	TGGCGAGCAA	GGAAGAACTC	ATTTCCCGGT	TAGTGGCTCG	AGACATGTCC	2760
AACCTTGAGG	AATTC					2775

序列 6

GAATCCGCG	TATCGCCCGG	TTCTATCAGC	GGGCCGCTTT	CGAAAGTCAT	GGTGTTAGCC	60
GGTAGGGTCT	TMTTCTTGGC	CATGCTTGTT	GCCTGAACCT	TCGTTGACAT	AGGGCAGAGG	120
TGCGTTTGCC	GCTTCGCTTC	GCGATGAACC	GCATCGAGAT	GCTGAGGTCA	GGATTTTTC	180
TTAACTCGCG	TAAGCATTCT	GTCATTTTTT	TGGTGGCTTT	GAACAGCCTG	ATGAAAGGTG	240
GTCTCGCCCT	TTGAGGCCGA	TTCTTGGGCG	CTTGGCGGCG	TCGAAGCGAT	GCTCCACTAC	300
CGATTAAGAT	AATTAATAA	AGGAAACCGC	ATGGTTTCTT	ATGTGAATTT	GTCTGGCATA	360
CTCCAGCTCA	AGGGCAATTT	TTGGGCTATT	GGCTGAGCAG	TTGCCTCTAT	ATGGTTATTC	420
AGAATAACAA	TTGACTCCTC	AGGAGGTCAG	CGATGAGCAT	TCTTGGTTTG	AATGGTGCCC	480
CGGTCCGAGC	TGAGCAGCTG	GGCTCGGCTC	TTGATCGCAT	GAAGAAGGCG	CACCTGGAGC	540
AGGGGCCCTGC	AAACTTGGAG	CTGCGTCTGA	GTAGGCTGGA	TCGTGCGATT	GCAATGCTTC	600
TGGAATAATCG	TGAAGCAATT	GCCGACGCGG	TTTCTGCTGA	CTTTGGCAAT	CGCAGCCGTG	660
AGCAAACACT	GCTTTGCGAC	ATTGCTGGCT	CGGTGGCAAG	CCTGAAGGAT	AGCCGCGAGC	720
ACGTGGCCAA	ATGGATGGAG	CCCGAACATC	ACAAGGCGAT	GTTTCCAGGG	GCGGAGGCAC	780
GCCTTGAGTT	TCAGCCGCTG	GGTGTCGTTG	GGGTCAATTAG	TCCCTGGAAC	TTCCCTATCG	840
TACTGGCCTT	TGGGCCGCTG	GCCGGCATAT	TCGCAGCAGG	TAATCGCGCC	ATGCTCAAGC	900
CGTCCGAGCT	TACCCCGCGG	ACTTCTGCCC	TGCTTGCGGA	GCTAATTGCT	CGTTACTTCG	960
ATGAAACTGA	GCTGACTACA	GTGCTGGGCG	ACGCTGAAGT	CGGTGCGCTG	TTCAGTGCTC	1020
AGCCTTTCGA	TCATCTGATC	TTCAACCGCG	GCACTGCCGT	GGCCAAGCAC	ATCATGCGTG	1080
CCGCGGCGGA	TAACCTAGTG	CCCGTTACCC	TGGAATTGGG	TGGCAAATCG	CCGGTGATCG	1140
TTTCCCGCAG	TGCAGATATG	GCGGACGTTG	CACAACGGGT	GTTGACGGTG	AAAACCTTCA	1200
ATGCCGGGCA	AATCTGTCTG	GCACCGTGGG	TGAGTCGAAC	TTGGCGATGC	GCGCACCTTA	1260
CGGAGAAGCG	ATCCACGGAC	TGCTCTCTGT	CCTCCTTTCA	ACGGAGTGTT	AGAACCGTTG	1320
GTAGTGGMPT	TGGACGGGCC	CAGGAGCATG	CGCTTCTGGG	CCCGTTTCTT	GAGTATTCAT	1380
TGGATAGTCA	CGCGTGCTAG	CTTCGAGCCT	GCACAGCTGA	TGAGCACCCCT	GGAAGGCGCG	1440
CTGTACGCGG	ACGACTGGGT	TCATCTTCGC	CATTCATGAC	GGAACTCCGT	TCCCCAGTAC	1500
CGCGATGACT	ATTTTGCTTC	TTCCGATGTC	CGATTCCACG	CCGCCTGACG	CTAAGCGGGG	1560
GCGGGGGCGC	CCGCATCCCA	GCCCAGACAG	CAACAAATGA	GTAGGCTCTT	GGATGCCGCG	1620
GCGGCTGAGA	TTGGTAACGG	CAATTTCTGC	AATGTGACGA	TGGATTTCGAT	TGCCCCGTGCT	1680
GCCGGCGTCT	CAAAAAAAC	GCTGTACGTC	TTGGTGGCGA	GCAAGGAAGA	ACTCATTTCC	1740
CGGTTAGTGG	CTCGAGACAT	GTCCAACCTT	GAGGAATTC			1779

序列 7

CTGCAGCCGA	GCATCGATTG	AGCACTTTAC	CCAGCTGCGC	TGGCTGACCA	TTCAGAATGG	60
CCCCGCGCAC	TATCCAATCT	AAATCGATCT	TCGGGCGCCG	CGGGCATCAT	GCCCCGCGCG	120
CTCGCCTCAT	TTCAATCTCT	AACTTGATAA	AAACAGAGCT	GTTCTCCGGT	CTTGGTGGAT	180
CAAGGCCAGT	CGCGGAGAGT	CTCGAAGAGG	AGAGTACAGT	GAACGCCGAG	TCCACATTGC	240
AACCGCAGGC	ATCATCATGC	TCTGCTCAGC	CACGCTACCG	CAGTGTGTCG	ATTGGTCATC	300
CTCCGGTTGA	GGTTACGCAA	GACGCTGGAG	GTATTGTCCG	GATGCGTTCT	CTCGAGGCGC	360
TTCTTCCCTT	CCCCGGTCTGA	ATTCTTGAGC	GTCTCGAGCA	TTGGGCTAAG	ACCCGTCCAG	420
AACAAACCTG	CGTTGCTGCC	AGGGCGGCCA	ATGGGGAATG	GCGTCGTATC	AGCTACGCGG	480
AAATGTTCCA	CAACGTCCGC	GCCATCGCAC	AGAGCTTGCT	TCCTTACGGA	CTATCGGCAG	540
AGCGTCCGCT	GCTTATCGTC	TCTGGAATG	ACCTGGAACA	TCTTCAGCTG	GCATTTGGGG	600
CTATGTATGC	GGGCATPCCC	TATGCCCCGG	TGTCCTCTGC	TTATTCACTG	CTGTGCGAAG	660
ATTTGGCGAA	GCTGCGTCAC	ATCGTAGGTC	TTCTGCAACC	GGGACTGGTC	TTTGCTGCCG	720
ATGCAGCACC	TTTCCAGGGG	ACAGCAAGCG	AACCGGAATT	GCCAGCTGGG	GCGCCCTCTG	780
GTAAGGTTGG	GAAGCCCTGC	AAAGTAAACT	GGATGGCTTT	CTTGCCGCCA	AGGATCTGAT	840
GGCGCAGGGG	ATCAAGATCT	GATCAAGAGA	CAGGATGAGG	ATCGTTTCGC	ATGATTGAAC	900
AAGATGGATT	GCACGCAGGT	TCTCCGGCCG	CTTGGGTGGA	GAGGCTATTG	GGCTATGACT	960
GGGCACAACA	GACAAATCGC	TGCTCTGATG	CCGCCGTGTT	CCGGCTGTCA	GCGCAGGGGC	1020
GCCCGGTTCT	TTTTGTCAAG	ACCGACCTGT	CCGGTGCCCT	GAATGAACTG	CAGGACGAGG	1080
CAGCGCGGCT	ATCGTGGCTG	GCCACGACGG	GCGTTCCTTG	CQCAGCTGTG	CTCGACGTTG	1140
TCACTGAAGC	GGGAAGGGAC	TGGCTGCTAT	TGGGCGAAGT	GCCGGGGCAG	GATCTCCTGT	1200
CATCTCACCT	TGCTCCTGCC	GAGAAAGTAT	CCATCATGGC	TGATGCAATG	CGGCGGCTGC	1260
ATACGCTTGA	TCCGGGTACC	TGCCCATTCG	ACCACCAAGC	GAAACATCGC	ATCGAGCGAG	1320
CACGTACTCG	GATGGAAGCC	GGTCTTGTCG	ATCAGGATGA	TCTGGACGAA	GAGCATCAGG	1380
GGCTCGCGCC	AGCCGAACTG	TTGCGCCAGG	TCAAGGCGCG	CATGCCCCGAC	GGCGAGGATC	1440
TCGTCTGTGAC	CCATGGCGAT	GCCTGCTTGC	CGAATATCAT	GGTGGAATAAT	GGCCGCTTTT	1500
CTGGATTTCAT	CGACTGTGGC	CGGCTGGGTG	TGGCGGACCG	CTATCAGGAC	ATAGCGTTGG	1560
CTACCCGTGA	TATTGCTGAA	GAGCTTGCGC	GCGAATGGGC	TGACCGCTTC	CTCGTGCTTT	1620
ACGGTATCGC	CGCTCCCGAT	TGCGAGCGCA	TCGCCTTCTA	TCGCCTTCTT	GACGAGTTCT	1680
TCTGAGCGGG	ACTCTGGGGT	TCGAAATGAC	CGACCAAGCG	ACGCCCCCTGT	TTTGCAATGG	1740
CGGTGCGCGA	AAGTTGATGC	GCTGTATCGT	GGTGAAGATC	AATCCATGCT	GCGTGACGAG	1800
GCCACACTGT	GAGTTGGTCA	GGGGGGGCTT	ACTCGGCGTT	TTCCGACACT	GCGTTGGTTG	1860
CGGCAGTGCG	CACCCCTTGG	ATTGATTGCG	GGGGTGCCCT	GTCGCTGGTG	TCGCCTATCG	1920
ACTTAGGGGT	AAAGGTCGCT	CGCGAAGTTC	TGATGCGTGC	GTCGCTTGAA	CCACAAATGG	1980
TCGATAGCGT	ACTCGCAGGC	TCTATGGCTC	AAGCAAGCTT	TGATGCTTAC	CTGCTCCCGC	2040
GGCACATTGG	CTTGATACAGC	GGTGTTCCTA	AGTCGGTTCC	GGCCTTGGGG	GTGCAGCGCA	2100
TTTGCGGCAC	AGGCTTCGAA	CTGCTTCGGC	AGGCCGGCGA	GCAGATTTC	CAAGGCGCTG	2160
ATCACGTGCT	GTGTGTCCGG	GGCTGCAG				2188

序列 8

CTGCAGCCGA	GCATCGATTG	AGCACTTTAC	CCAGCTGCGC	TGGCTGACCA	TTCAGAATGG	60
CCCGCGGCAC	TATCCAATCT	AAATCGATCT	TCGGGCGCCG	CGGGCATCAT	GCCC GCGGCG	120
CTCGCCTCAT	TTCAATCTCT	AACTTGATAA	AAACAGAGCT	GTTCTCCGGT	CTTGGTGGAT	180
CAAGGCCAGT	CGCGGAGAGT	CTCGAAGAGG	AGAGTACAGT	GAACGCCGAG	TCCACATTCG	240
AACCGCAGGC	ATCATCATGC	TCTGCTCAGC	CACGCTACCG	CAGTGTGTCT	ATTGGTCATC	300
CTCCGGTTGA	GGTTACGCAA	GACGCTGGAG	GTATTGTCCG	GATGCGTTCT	CTCGAGGCGC	360
TTCTTCCCTT	CCCGGGTCGA	ATTCTTGAGC	GTCTCGAGCA	TTGGGCTAAG	ACCCGTCCAG	420
AACAAACCTG	CGTTGCTGCC	AGGGCGGCAA	ATGGGGAATG	GCGTCGTATC	AGCTACGCGG	480
AAATGTTCCA	CAACGTCCGC	GCCATCGCAC	AGAGCTTGCT	TCCTTACGGA	CTATCGGCAG	540
AGCGTCCGCT	GCTTATCGTC	TCTGGAATG	ACCTGGAACA	TCTTCAGCTG	GCATTTGGGG	600
CTATGTATGC	GGGCATTCCC	TATTGCCCCG	TGTCTCCTGC	TTATTCACTG	CTGTGCGAAG	660
ATTTGGCGAA	GCTGCGTCAC	ATCGTAGGTC	TTCTGCAACC	GGGACTGGTC	TTTGCTGCCG	720
ATGCAGCACC	TTTCCAGGGG	GAGAGGCGGT	TTGCGTATTG	GGCGCATGCA	TAAAAACTGT	780
TGTAATTCAT	TAAGCATTCCT	GCCGACATGG	AAGCCATCAC	AAACGGCATG	ATGAACCTGA	840
ATCGCCAGCG	GCATCAGCAC	CTTGTCGCCT	TGCGTATAAT	ATTTGCCCAT	GGACGCACAC	900
CGTGGAAACG	GATGAAGGCA	CGAACCAGT	TGACATAAGC	CTGTTCCGGT	CGTAAACTGT	960
AATGCAAGTA	GCGTATGCGC	TCACGCAACT	GGTCCAGAAC	CTTGACCGAA	CGCAGCGGTG	1020
GTAACGGCGC	AGTGGCGGTT	TTCATGGCTT	GTTATGACTG	TTTMTTGTGA	CAGTCTATGC	1080
CTCGGGCATC	CAAGCAGCAA	GCGCGTTACG	CCGTGGGTCT	ATGTTTGATG	TTATGGAGCA	1140
GCAACGATGT	TACGCAGCAG	CAACGATGTT	ACGCAGCAGG	GCAGTCGCCC	TAAAAACAAAG	1200
TTAGGTGGCT	CAAGTATGGG	CATCATTCGC	ACATGTAGGC	TCGGCCCTGA	CCAAGTCAAA	1260
TCCATGCGGG	CTGCTCTTGA	TCTTTTCGGT	CGTGAGTTCT	GAGACGTAGC	CACCTACTCC	1320
CAACATCAGC	CGGACTCCGA	TTACCTCGGG	AACTTGCTCC	GTAGTAAGAC	ATTCATCGCG	1380
CTTGCTGCCT	TCGACCAAGA	AGCGGTTGTT	GGCGCTCTCG	CGGCTTACGT	TCTGCCCAGG	1440
TTTGAGCAGC	CGCGTAGTGA	GATCTATATC	TATGATCTCG	CAGTCTCCGG	CGAGCACCGG	1500
AGGCAGGGCA	TTGCCACCGC	GCTCATCAAT	CTCCTCAAGC	ATGAGGCCAA	CGCGCTTGCT	1560
GCTTATGTGA	TCTACGTGCA	AGCAGATTAC	GGTGACGATC	CCGCAGTGGC	TCTCTATACA	1620
AAGTTGGGCA	TACGGGAAGA	AGTGATGCAC	TTTGATATCG	ACCCAAGTAC	CGCCACCTAA	1680
CAATTCGTTT	AAGCCGAGAT	CGGCTTCCCC	TGTTTTGCAA	TGGCGGTCTG	CGAAAGTTGA	1740
TGCGCTGTAT	CGTGGTGAAG	ATCAATCCAT	GCTGCGTGAC	GAGGCCACAC	TGTGAGTTGG	1800
TCAGGGGGGG	CTTACTCGGC	GTTTTCCGAC	ACTGCGTTGG	TTGCGGCAGT	GCGCACCCCC	1860
TGGATTGATT	GCGGGGGTGC	CCTGTCGCTG	GTGTCGCCTA	TCGACTTAGG	GGTAAAGGTC	1920
GCTCGCGAAG	TTCTGATGCG	TGCGTCGCTT	GAACCACAAA	TGGTCGATAG	CGTACTCGCA	1980
GGCTCTATGG	CTCAAGCAAG	CTTTGATGCT	TACCTGCTCC	CGCGGCACAT	TGGCTTGTAC	2040
AGCGGTGTTT	CCAAGTCGGT	TCCGGCCTTG	GGGGTGCAGC	GCATTTGCGG	CACAGGCTTC	2100
GAAGTGCTTC	GGCAGGCCGG	CGAGCAGATT	TCCCAAGGCG	CTGATCACGT	GCTGTGTGTC	2160
GCGGGCTGCA	G					2171

序列 9

CTGCAGCCGA	GCATCGATTG	AGCACTTTAC	CCAGCTGCGC	TGGCTGACCA	TTCAGAATGG	60
CCCGCGGCAC	TATCCAATCT	AAATCGATCT	TCGGGCGCCG	CGGGCATCAT	GCCCGCGGCG	120
CTCGCCTCAT	TTCAATCTCT	AAC TTGATAA	AAACAGAGCT	GTTCTCCGGT	CTTGGTGGAT	180
CAAGGCCAGT	CGCGGAGAGT	CTCGAAGAGG	AGAGTACAGT	GAACGCCGAG	TCCACATTGC	240
AACCGCAGGC	ATCATCATGC	TCTGCTCAGC	CACGCTACCG	CAGTGTGTCT	ATTGGTCATC	300
CTCCGGTTGA	GGTTACGCAA	GACGCTGGAG	GTATTGTCCG	GATGCGTTCT	CTCGAGGCGC	360
TTCTTCCCTT	CCCGGGTCGA	ATTCTTGAGC	GTCTCGAGCA	TTGGGCTAAG	ACCCGTCCAG	420
AACAAACCTG	CGTTGCTGCC	AGGGCGGCAA	ATGGGGAATG	GCGTCGTATC	AGCTACGCGG	480
AAATGTTCCA	CAACGTCCGC	GCCATCGCAC	AGAGCTTGCT	TCCTTACCGA	CTATCGGCAG	540
AGCGTCCGCT	GCTTATCGTC	TCTGGAAATG	ACCTGGAACA	TCTTCAGCTG	GCATTTGGGG	600
CTATGTATGC	GGGCATTCCC	TATTGCCCGG	TGTCTCCTGC	TTATTCACTG	CTGTCGCAAG	660
ATTTGGCGAA	GCTGCGTCAC	ATCGTAGGTC	TTCTGCAACC	GGGACTGGTC	TTTGCTGCCG	720
ATGCAGCACC	TTTCCAGCGC	GCTGTTTTGC	AATGGCGGTC	GGCGAAAGTT	GATGCGCTGT	780
ATCGTGGTGA	AGATCAATCC	ATGCTGCGTG	ACGAGGCCAC	ACTGTGAGTT	GGTCAGGGGG	840
GGCTTACTCG	GCGTTTTCCG	ACACTGCGTT	GGTTGCGGCA	GTGCGCACCC	CCTGGATTGA	900
TTGCGGGGGT	GCCCTGTTCG	TGGTGTTCGC	TATCGACTTA	GGGGTAAAGG	TCGCTCGCGA	960
AGTTCTGATG	CGTGCGTCGC	TTGAACCACA	AATGGTTCGAT	AGCGTACTCG	CAGGCTCTAT	1020
GGCTCAAGCA	AGCTTTGATG	CTTACCTGCT	CCCGCGGCAC	ATTGGCTTGT	ACAGCGGTGT	1080
TCCCAAGTCG	GTTCCGGCCT	TGGGGGTGCA	GCGCATTTCG	GGCACAGGCT	TCGAACTGCT	1140
TCGGCAGGCC	GGCGAGCAGA	TTTCCCAAGG	CGCTGATCAC	GTGCTGTGTG	TCGCGGGCTG	1200
CAG						1203

序列 10

GAATTCCTCT	GGCGACGAAA	GGGCGGCAGG	CCGCATGGCC	ACGGCTGGGC	GGTAACTGAT	60
GCTTGCGTTA	ATCGTTAACC	GTTTGAAATT	CCTTGCCAAA	TTTCGGCGAG	AGAATCATGC	120
GGGTACGCCCT	TTCCGTGCGC	TTTGATCTGC	GCTTCCGTGC	CTTGAATCAG	AAAAATAGTT	180
AATTGACAGA	ACTATAGGTT	CGCAGTAGCT	TTTGCTCACC	CACCAAATCC	ACAGCACTGG	240
GGTGACGAT	GAATAGCTAC	GATGGCCGTT	GGTCTACCGT	TGATGTGAAG	GTTGAAGAAG	300
GTATCGCTTG	GGTCACGCTG	AACCGCCCGG	AGAAGCGCAA	CGCAATGAGC	CCAACTCTCA	360
ATCGAGAGAT	GGTCGAGGTT	CTGGAGGTGC	TGGAGCAGGA	CGCAGATGCT	CGCGTGCTTG	420
TTCTGACTGG	TGCAGGCGAA	TCCTGGACCG	CGGGCATGGA	CCTGAAGGAG	TATTTCCGCG	480
AGACCGATGC	TGGCCCCGAA	ATTCTGCAAG	AGAAGATTCTG	TCGGGGACAG	CAAGCGAACC	540
GGAATTGCCA	GCTGGGGCGC	CCTCTGGTAA	GGTTGGGAAG	CCCTGCAAAG	TAAACTGGAT	600
GGCTTTCTTG	CCGCCAAGGA	TCTGATGGCG	CAGGGGATCA	AGATCTGATC	AAGAGACAGG	660
ATGAGGATCG	TTTCGCATGA	TTGAACAAGA	TGGATTGCAC	GCAGGTTCTC	CGGCCGCTTG	720
GGTGGAGAGG	CTATTGGGCT	ATGACTGGGC	ACAACAGACA	ATCGGCTGCT	CTGATGCCGC	780
CGTGTTCGCG	CTGTACGCG	AGGGGCGCCC	GGTTCCTTTT	GTCAAGACCG	ACCTGTCCGG	840
TGCCCCGACT	GAACTGCAGG	ACGAGGACAG	CGGGCTATCG	TGGCTGGCCA	CGACGGGCGT	900
TCCTTGCGCA	GCTGTGCTCG	ACGTTGTCAC	TGAAGCGGGA	AGGGACTGGC	TGCTATTGGG	960
CGAAGTGCCG	GGGCAGGATC	TCCTGTCATC	TCACCTTGCT	CCTGCCGAGA	AAGTATCCAT	1020
CATGGCTGAT	GCAATGCGGC	GGCTGCATAC	GCTTGATCCG	GCTACCTGCC	CATTCGACCA	1080
CCAAGCGAAA	CATCGCATCG	AGCGAGCACG	TACTCGGATG	GAAGCCGGTC	TTGTCCGATCA	1140
GGATGATCTG	GACGAAGAGC	ATCAGGGGCT	CGCGCCAGCC	GAAGCTGTTG	CCAGGCTCAA	1200
GGCGCGCATG	CCCGACGGCG	AGGATCTCGT	CGTGACCCAT	GGCGATGCCT	GCTTGCCGAA	1260
TATCATGGTG	GAAAATGGCC	GCTTTTCTGG	ATTTCATCGAC	TGTGGCCGGC	TGGGTGTGGC	1320
GGACCGCTAT	CAGGACATAG	CGTTGGCTAC	CCGTGATATT	GCTGAAGAGC	TTGGCGGCGA	1380
ATGGGCTGAC	CGCTTCCTCG	TGCTTTACGG	TATCGCCGCT	CCCGATTGCG	AGCGCATCGC	1440
CTTCTATCGC	CTTCTTGACG	AGTTCTTCTG	AGCGGGACTC	TGGGGTTCGA	AATGACCGAC	1500
CAAGCGACGC	CCCGAGCAGG	GCATGAAGCA	GTTCTTTGAC	GAGAAAAGCA	TCAAGCCGGG	1560
CTTGACAGAC	TACAAGCGCT	GATAAATGCG	CCGGGGCCCT	CGCTGCGCCC	CCGGCCTTCC	1620
AATAATGACA	ATAATGAGGA	GTGCCCAATG	TTTCACGTGC	CCCTGCTTAT	TGGTGGTAAG	1680
CCTTGTTTCA	CATCTGATGA	GCGCACCTTC	GAGCGTCGTA	GCCCGCTGAC	CGGAGAAGTG	1740
GTATCGCGCG	TCGCTGCTGC	CAGTTTGGAA	GATGCGGACG	CCGCAGTGGC	CGCTGCACAG	1800
GCTGCGTTTC	CTGAATGGGC	GGCGCTTGCT	CCGAGCGAAC	GCCGTGCCCG	ACTGCTGCCA	1860
GCGGCGGATC	TTCTAGAGGA	CCGTTCTTCC	GAGTTCACCG	CCGCAGCGAG	TGAAACTGGC	1920
GCAGCGGGAA	ACTGGTATGG	GTTTAACGTT	TACCTGGCGG	CGGGCATGTT	GCGGGGAATT	1980
C						1981

序列 11

GAATCCCCCT	GGCGACGAAA	GGGCGGCAGG	CCGCATGGCC	ACGGCTGGGC	GGTAACTGAT	60
GCTTGCCTTA	ATCGTTAACC	GTITGAAATT	CCTTGCCAAA	TTTCGGCGAG	AGAATCATGC	120
GGGTACGCCCT	TTCCGTGCGC	TTTGATCTGC	GCTTCCGTGC	CTTGAATCAG	AAAAATAGTT	180
AATTGACAGA	ACTATAGGTT	CGCAGTAGCT	TTTGCTCACC	CACCAAATCC	ACAGCACTGG	240
GGTGCACGAT	GAATAGCTAC	GATGGCCGTT	GGTCTACCGT	TGATGTGAAG	GTTGAAGAAG	300
GTATCGCTTG	GGTCACGCTG	AACCGCCCGG	AGAAGCGCAA	CGCAATGAGC	CCAACTCTCA	360
ATCGAGAGAT	GGTCGAGGTT	CTGGAGGTGC	TGGAGCAGGA	CGCAGATGCT	CGCGTGCTTG	420
TTCTGACTGG	TGCAGGCGAA	TCCTGGACCG	CGGGCATGGA	CCTGAAGGAG	TATTTCCGCG	480
AGACCGATGC	TGGCCCCGAA	ATTCTGCAAG	AGAAGATTCT	TGGGGGAGA	GGCGGTTTGC	540
GTATTGGGCG	CATGCATAAA	AACTGTTGTA	ATTCATTAAAG	CATTCTGCCG	ACATGGAAGC	600
CATCACAAAC	GGCATGATGA	ACCTGAATCG	CCAGCGGCAT	CAGCACCTTG	TCGCCTTGCG	660
TATAATATTT	GCCCATGGAC	GCACACCGTG	GAAACGGATG	AAGGCACGAA	CCCAGTTGAC	720
ATAAGCCTGT	TCGGTTCGTA	AACTGTAATG	CAAGTAGCGT	ATGCGCTCAC	GCAACTGGTC	780
CAGAACCCTG	ACCGAACGCA	GCGGTGGTAA	CGGCGCAGTG	GCGGTTTTCA	TGGCTTGTTA	840
TGACTGTTTT	TTTGATACAGT	CTATGCCTCG	GGCATCCAAG	CAGCAAGCGC	GTTACGCCGT	900
GGGTCGATGT	TTGATGTTAT	GGAGCAGCAA	CGATGTTACG	CAGCAGCAAC	GATGTTACGC	960
AGCAGGGCAG	TCGCCCTAAA	ACAAAGTTAG	GTGGCTCAAG	TATGGGCATC	ATTCGCACAT	1020
GTAGGCTCGG	CCCTGACCAA	GTCAAATCCA	TGCGGGCTGC	TCTTGATCTT	TTTCGGTCGTG	1080
AGTTCGGAGA	CGTAGCCACC	TACTCCCAAC	ATCAGCCGGA	CTCCGATTAC	CTCGGGAACT	1140
TGCTCCGCTAG	TAAGACATTC	ATCGCGCTTG	CTGCCTTCGA	CCAAGAAGCG	GTTGTTGGCG	1200
CTCTCGCGGC	TTACGTTCTG	CCCAGGTTTG	AGCAGCCGCG	TAGTGAGATC	TATATCTATG	1260
ATCTCGCAGT	CTCCGGCGAG	CACCGGAGGC	AGGGCATTGC	CACCGCGCTC	ATCAATCTCC	1320
TCAAGCATGA	GGCCAACGCG	CTTGGTGCTT	ATGTGATCTA	CGTGCAAGCA	GATTACGGTG	1380
ACGATCCCGC	AGTGGCTCTC	TATACAAAAGT	TGGGCATACG	GGAAGAAGTG	ATGCACTTTG	1440
ATATCGACCC	AAGTACCGCC	ACCTAACAAAT	TCGTTCAAGC	CGAGATCGGC	TTCCCCGAGC	1500
AGGGCATGAA	GCAGTTCCTT	GACGAGAAAA	GCATCAAGCC	GGGCTTGACG	ACCTACAAGC	1560
GCTGATAAAT	GCGCCGGGGC	CCTCGCTGCG	CCCCCGGCCCT	TCCAATAATG	ACAATAATGA	1620
GGAGTGCCCA	ATGTTTCACG	TGCCCCCTGCT	TATTGGTGGT	AAGCCTTGTT	CAGCATCTGA	1680
TGAGCGCACC	TTGAGCGGTC	GTAGCCCGCT	GACCGGAGAA	GTGGTATCGC	GCGTCGCTGC	1740
TGCCAGTTTG	GAAGATGCGG	ACGCCGCGAGT	GGCCGCTGCA	CAGGCTGCGT	TTCTTGAATG	1800
GGCGGCGCTT	GCTCCGAGCG	AACGCCGTGC	CCGACTGCTG	CGAGCGGCGG	ATCTTCTAGA	1860
GGACCGTTCT	TCCGAGTTCA	CCGCCGCAGC	GAGTGAAACT	GGCGCAGCGG	GAAACTGGTA	1920
TGGGTTTAAC	GTTTACCTGG	CGGCGGGCAT	GTTGCGGGGA	ATTTC		1964

序列 12

GAATCCCCCT	GGCGACGAAA	GGGCGGCAGG	CCGCATGGCC	ACGGCTGGGC	GCTAACTGAT	60
GCTTGCGTTA	ATCGTTAACC	GTGTGAAATT	CCTTGCCAAA	TTTCGGCGAG	AGAATCATGC	120
GGGTACGCCT	TTCCGTGCGC	TTTGATCTGC	GCTTCCGTGC	CTTGAATCAG	AAAAATAGTT	180
AATTGACAGA	ACTATAGGTT	CGCAGTAGCT	TTTGCTCACC	CACCAAATCC	ACAGCACTGG	240
GGTGCACGAT	GAATAGCTAC	GATGGCCGTT	GGTCTACCGT	TGATGTGAAG	GTTGAAGAAG	300
GTATCGCTTG	GGTCACGCTG	AACCGCCCGG	AGAAGCGCAA	CGCAATGAGC	CCAACTCTCA	360
ATCGAGAGAT	GGTCGAGGTT	CTGGAGGTGC	TGGAGCAGGA	CGCAGATGCT	CGCGTGCTTG	420
TTCTGACTGG	TGCAGGCGAA	TCCTGGACCG	CGGGCATGGA	CCTGAAGGAG	TATTTCCGCG	480
AGACCGATGC	TGGCCCCGAA	ATTCTGCAAG	AGAAGATTCTG	TCGCGAGCAG	GGCATGAAGC	540
AGTTCCCTGA	CGAGAAAAGC	ATCAAGCCGG	GCTTGCAGAC	CTACAAGCGC	TGATAAATGC	600
GCCGGGGCCC	TCGCTGCGCC	CCCGGCCCTC	CAATAATGAC	AATAATGAGG	AGTGCCCAAT	660
GTTTCACGTG	CCCCTGCTTA	TTGGTGGTAA	GCCTTGTTCA	GCATCTGATG	AGCGCACCTT	720
CGAGCGTCGT	AGCCCGCTGA	CCGGAGAAGT	GGTATCGCGC	GTCGCTGCTG	CCAGTTTGGA	780
AGATGCGGAC	GCCGCAGTGG	CCGCTGCACA	GGCTGCGTTT	CCTGAATGGG	CGGCGCTTGC	840
TCCGAGCGAA	CGCCGTGCCC	GACTGCTGCG	AGCGGCGGAT	CTTCTAGAGG	ACCGTTCTTC	900
CGAGTTCACC	GCCGCAGCGA	GTGAAACTGG	CGCAGCGGGA	AACTGGTATG	GGTTTAACGT	960
TTACCTGGCG	GCGGGCATGT	TGCGGGGAAT	TC			992

序列 13

GAATTCCAAT	AATGACAATA	ATGAGGAGTG	CCCAATGTTT	CACGTGCCCC	TGCTTATTGG	60
TGGTAAGCCT	TGTTACAGCAT	CTGATGAGCG	CACCTTCGAG	CGTCGTAGCC	CGCTGACCGG	120
AGAAGTGGTA	TGCGCGCTCG	CTGCTGCCAG	TTTGGAAGAT	GCGGACGCCG	CAGTGGCCCG	180
TGCACAGGCT	GCGTTTCCTG	AATGGGCGGC	GCTTGCTCCG	AGCGAACGCC	GTGCCCGACT	240
GCTGCGAGCG	GCGGATCTTC	TAGAGGACCG	TTCTTCCGAG	TTCACCGCCG	CAGCGAGTGA	300
AACTGGCGCA	GCGGGAAACT	GGTATGGGTT	TAACGTTTAC	CTGGCGGCGG	GCATGTTGCG	360
GGAAGCCGCG	GCCATGACCA	CACAGATTCA	GGGCGATGTC	ATTCCGTCCA	ATGTGCCCGG	420
TAGCTTTGCC	ATGGCGGTTT	GACAGCCATG	TGGCGTGGTG	CTCGGTATTG	CGCCTTGGA	480
TGCTCCGGTA	ATCCTTGCGG	TACGGGCTGT	TGCGATGCCG	TTGGCATGCG	GCAATACCGT	540
GGTGTGAAA	AGCTCTGAGC	TGAGTCCCTT	TACCCATCGC	CTGATTGGTC	AGGTGTTGCA	600
TGATGCTGGT	CTGGGGGATG	GCGTGGTGAA	TGTCATCAGC	AATGCCCCGC	AAGACGCTCC	660
TGCGGTGGTG	GAGCGACTGA	TTGCAAATCC	TGCGGTACGT	CGAGTGAAC	TCACCGGTTT	720
GACCCACGTT	GGACGGATCA	TTGGTGAGCT	GTCTGCGCGT	CATCTGAAGC	CTGCTGTGCT	780
GGAATTAGGT	GGTAAGGCTC	CGTTCTTGGT	CTTGGACGAT	GCCGACCTCG	ATGCGGCGGT	840
CGAAGCGGCG	GCCTTTGGTG	CCTACTTCAA	TCAGGGTCAA	ATCTGCATGT	CCACTGAGCG	900
TCTGATTGTG	ACAGCAGTCG	CAGACGCCTT	TGTTGAAAAG	CTGGCGAGGA	AGGTCCGCAC	960
ACTGCGTGCT	GGCGATCCTA	ATGATCCGCA	ATCGGTCTTG	GGTTCGTTGA	TTGATGCCAA	1020
TGCAGGTCAA	CGCATCCAGG	TTCTGGTCGA	TGATGCGCTC	GGGGACAGCA	AGCGAACCGG	1080
AATTGCCAGC	TGGGGCGCCC	TCTGGTAAGG	TTGGGAAGCC	CTGCAAAGTA	AACTGGATGG	1140
CTTTCTTGCC	GCCAAGGATC	TGATGGCGCA	GGGGATCAAG	ATCTGATCAA	GAGACAGGAT	1200
GAGGATCGTT	TGCGATGATT	GAACAAGATG	GATTGCACGC	AGGTTCCTCC	GCCGCTTGGG	1260
TGGAGAGGCT	ATTCGGCTAT	GAATGGGCAC	AACAGACAAT	CGGCTGCTCT	GATGCCGCCG	1320
TGTTCCGGCT	GTGAGCGCAG	GGGCGCCCGG	TTCTTTTGT	CAAGACCGAC	CTGTCCGGTG	1380
CCCTGAATGA	ACTGCAGGAC	GAGGCAGCGC	GGCTATCGTG	GCTGGCCACG	ACGGGCGTTC	1440
CTTGCGCAGC	TGTGCTCGAC	GTGTCTACTG	AAGCGGGAAG	GGACTGGCTG	CTATTGGGCG	1500
AAGTGCCGGG	GCAGGATCTC	CTGTCTCTC	ACCTTGCTCC	TGCCGAGAAA	GTATCCATCA	1560
TGGCTGATGC	AATGCGGCGG	CTGCATACGC	TTGATCCGGC	TACCTGCCCC	TTGACACCAC	1620
AAGCGAAACA	TGCGATCGAG	CGAGCACGTA	CTCGGATGGA	AGCCGGTCTT	GTCGATCAGG	1680
ATGATCTGGA	CGAAGAGCAT	CAGGGGCTCG	CGCCAGCCGA	ACTGTTCCGC	AGGCTCAAGG	1740
CGCGCATGCC	CGACGGCGAG	GATCTCGTCG	TGACCCATGG	CGATGCCTGC	TTGCCGAATA	1800
TCATGGTGGA	AAATGGCCGC	TTTCTTGAT	TCATCGACTG	TGGCCGGCTG	GGTGTGGCGG	1860
ACCGTATCA	GGACATAGCG	TTGGCTACCC	GTGATATTGC	TGAAGAGCTT	GGCGGCGAAT	1920
GGGCTGACCG	CTTCTCTGTG	CTTTACGGTA	TCGCCGCTCC	CGATTGCGAG	CGCATCGCCT	1980
TCTATCGCCT	TCTTGACGAG	TTCTTCTGAG	CGGGACTCTG	GGGTTGAAA	TGACCGACCA	2040
AGCGACGCCC	GGCCAGCGC	GTGATTCGG	GCATTTGCCA	TATCAATGGA	CCGACTGTGC	2100
ATGACGAGGC	TCAGATGCCA	TTGGGTGGG	TGAAGTCCAG	CGGCTACGGC	AGCTTCGGCA	2160

01.04.30

GTCGAGCATC	GATTGAGCAC	TTTACCCAGC	TGCGCTGGCT	GACCAATCAG	AATGGCCCCG	2220
GGCACTATCC	AATCTAAATC	GATCTTCGGG	CGCCGCGGGC	ATCATGCCCC	CGGCGCTCGC	2280
CTCATTTCAA	TCTCTAACTT	GATAAAAACA	GAGCTGTTCT	CCGGTCTTGG	TGGATCAAGG	2340
CCAGTCGCGG	AGAGTCTCGA	AGAGGAGAGT	ACAGTGAACG	CCGAGTCCAC	ATTGCAACCG	2400
CAGGCATCAT	CATGCTCTGC	TCAGCCACGC	TACCGCAGTG	TGTCGATTGG	TCATCCTCCG	2460
GTTGAGGTTA	CGCAAGACGC	TGGAGGTATT	GTCCGGATGC	GTTCTCTCGA	GGCGCTTCTT	2520
CCCTTCCCGG	GTGGAATTC					2539

序列 14

GAATTCCAAT	AATGACAATA	ATGAGGAGTG	CCCAATGTTT	CACGTGCCCC	TGCTTATTGG	60
TGGTAAGCCT	TGTTACAGCAT	CTGATGAGCG	CACCTTCGAG	CGTCGTAGCC	CGCTGACCGG	120
AGAAGTGGTA	TCGCGCGTCG	CTGCTGCCAG	TTTGGGAAGAT	GCGGACGCGG	CAGTGCCCGC	180
TGCACAGGCT	GCGTTTCCTG	AATGGGCGGC	GCTTGCTCCG	AGCGAACGCC	GTGCCCAGCT	240
GCTGCGAGCG	GCGGATCTTC	TAGAGGACCG	TTCTTCCGAG	TTACCCGCCG	CAGCGAGTGA	300
AACTGGCGCA	GCGGGAAACT	GGTATGGGTT	TAACGTTTAC	CTGGCGGCGG	GCATGTTGCG	360
GGAAGCCGCG	GCCATGACCA	CACAGATTCA	GGGCGATGTC	ATTCCGTCCA	ATGTGCCCCG	420
TAGCTTTGCC	ATGGCGGTTT	GACAGCCATG	TGGCGTGGTG	CTCGGTATTG	CGCCTTGGA	480
TGCTCCGGTA	ATCCTTGCGG	TACGGGCTGT	TGCGATGCCG	TTGGCATGCG	GCAATACCGT	540
GGTGTGAAA	AGCTCTGAGC	TGAGTCCCTT	TACCCATCGC	CTGATTGGTC	AGGTGTTGCA	600
TGATGCTGGT	CTGGGGGATG	GCGTGGTGAA	TGTCATCAGC	AATGCCCCGC	AAGACGCTCC	660
TGCGGTGGTG	GAGCGACTGA	TTGCAAATCC	TGCGGTACGT	CGAGTGAACT	TCACCGGTTT	720
GACCCACGTT	GGACGGATCA	TTGGTGAGCT	GTCTGCGCGT	CATCTGAAGC	CTGCTGTGCT	780
GGAATTAGGT	GGTAAGGCTC	CGTTCTTGGT	CTTGGACGAT	GCCGACCTCG	ATGCGGCGGT	840
CGAAGCGGCG	GCTTTTGGTG	CCTACTTCAA	TCAGGGTCAA	ATCTGCATGT	CCACTGACCG	900
CTGATTGTG	ACAGCAGTCT	CAGACGCTTT	TGTTGAAAAG	CTGGCGAGGA	AGGTCGCCAC	960
ACTGCGTGCT	GCGCATCCTA	ATGATCCGCA	ATCGGTCTTG	GGTTCGTTGA	TTGATGCCAA	1020
TGCAGGTCAA	CGCATCCAGG	TGGGGAGAGG	CGGTTTGCCT	ATTGGGCGCA	TGCATAAAAA	1080
CTGTTGTAAT	TCATTAAGCA	TTCTGCCGAC	ATGGAAGCCA	TCACAAACGG	CATGATGAAC	1140
CTGAATCGCC	AGCGGCATCA	GCACCTTGTC	GCCTTGCGTA	TAATATTTGC	CCATGGACGC	1200
ACACCGTGGA	AACGGATGAA	GGCACGAACC	CAGTTGACAT	AAGCCTGTTC	GGTTTCGTAA	1260
CTGTAATGCA	AGTAGCGTAT	GCGCTCACGC	AACTGGTCCA	GAACCTTGAC	CGAACGCAGC	1320
GGTGGTAACG	GCGCAGTGCC	GGTTTTCATG	GCTTGTTATG	ACTGTTTTTT	TGTACAGTCT	1380
ATGCCCTCGG	CATCCAAGCA	GCAAGCGCGT	TACGCCGTGG	GTCGATGTTT	GATGTTATGG	1440
AGCAGCAACG	ATGTTACGCA	GCAGCAACGA	TGTTACGCAG	CAGGGCAGTC	GCCCTAAAC	1500
AAAGTTAGGT	GGCTCAAGTA	TGGGCATCAT	TCGCACATGT	AGGCTCGGCC	CTGACCAAGT	1560
CAAATCCATG	CGGGCTGCTC	TTGATCTTTT	CGGTGCTGAG	TTCCGGAGACG	TAGCCACCTA	1620
CTCCCAACAT	CAGCCGGACT	CCGATTACCT	CGGGAACCTG	CTCCGTAGTA	AGACATTTCAT	1680
CGCGCTTGCT	GCCTTCGACC	AAGAAGCGGT	TGTTGGCGCT	CTCGCGGCTT	ACGTTCTGCC	1740
CAGGTTTGAG	CAGCCGCGTA	GTGAGATCTA	TATCTATGAT	CTCGCAGTCT	CCGGCGAGCA	1800
CCGGAGGCAG	GGCATTGCCA	CCGCGCTCAT	CAATCTCCTC	AAGCATGAGG	CCAACGCGCT	1860
TGGTGCTTAT	GTGATCTACG	TGCAAGCAGA	TTACGGTGAC	GATCCCGCAG	TGGCTCTCTA	1920
TACAAAGTTG	GGCATACGGG	AAGAAGTGAT	GCACTTTGAT	ATCGACCCAA	GTACCGCCAC	1980
CTAACAATTC	GTTCAAGCCG	AGATCGGCTT	CCCAATTGGC	CCAGCGCGTC	GATTGCGGCA	2040
TTTGCCATAT	CAATGGACCG	ACTGTGCATG	ACGAGGCTCA	GATGCCATTG	GGTGGGGTGA	2100
AGTCCAGCGG	CTACGGCAGC	TTCCGGCAGTC	GAGCATCGAT	TGAGCACTTT	ACCCAGCTGC	2160

010430

GCTGGCTGAC	CATTCAGAAT	GGCCCGCGGC	ACTATCCAAT	CTAAATCGAT	CTTCGGGCGC	2220
CGCGGGCATC	ATGCCCCGCG	CGCTCGCCTC	ATTTCAATCT	CTAACTTGAT	AAAAACAGAG	2280
CTGTTCTCCG	GTCTTGGTGG	ATCAAGGCCA	GTCGCGGAGA	GTCTCGAAGA	GGAGAGTACA	2340
GTGAACGCCG	AGTCCACATT	GCAACCGCAG	GCATCATCAT	GCTCTGCTCA	GCCACGCTAC	2400
CGCAGTGTGT	CGATTGGTCA	TCCTCCGGTT	GAGGTTACGC	AAGACGCTGG	AGGTATTGTC	2460
CGGATGCGTT	CTCTCGAGGC	GCTTCTTCCC	TCCCCGGGTG	GAATTC		2506

序列 15

GAATTC	CAAT	AATGACAATA	ATGAGGAGTG	CCCAATGTTT	CACGTGCCCC	TGCTTATTGG	60
TGCTAAGCCT	TGTT	CAGCAT	CTGATGAGCG	CACCTTCGAG	CGTCGTAGCC	CGCTGACCGG	120
AGAAGTGGTA	TCGCGCGTCG	CTGCTGCCAG	TTTGGAAGAT	GCGGACGCCG	CAGTGGCCGC	180	
TGCACAGGCT	GCGTTTCCTG	AATGGGCGGC	GCTTGCTCCG	AGCGAACGCC	GTGCCCAGCT	240	
GCTGCGAGCG	GCGGATCTTC	TAGAGGACCG	TTCTTCCGAG	TTACCCGCCG	CAGCGAGTGA	300	
AACTGGCGCA	GCGGGAAACT	GGTATGGGTT	TAACGTTTAC	CTGGCGGCGG	GCATGTTGCG	360	
GGAAGCCGCG	GCCATGACCA	CACAGATTCA	GGGCGATGTC	ATTCCGTCCA	ATGTGCCCGG	420	
TAGCTTTGCC	ATGGCGGTTT	GACAGCCATG	TGGCGTGGTG	CTCGGTATTG	CGCCTTGCAA	480	
TGCTCCGGTA	ATCCTTGCGG	TACGGGCTGT	TGCGATGCCG	TTGGCATGCG	GCAATACCGT	540	
GGTGTTGAAA	AGCTCTGAGC	TGAGTCCCTT	TACCCATCGC	CTGATTGGTC	AGGTGTTGCA	600	
TGATGCTGGT	CTGGGGGATG	GCGTGGTGAA	TGTCATCAGC	AATGCCCCGC	AAGACGCTCC	660	
TGCGGTGGTG	GAGCGACTGA	TTGCAAATCC	TGCGGTACGT	CGAGTGAAC	TCACCGGTTT	720	
GACCCACGTT	GGACGGATCA	TTGGTGAGCT	GTCTGCGCGT	CATCTGAAGC	CTGCTGTGCT	780	
GGAAATTAGGT	GGTAAGGCTC	CGTTCTTGGT	CTTGGACGAT	GCCGACCTCG	ATGCGGCGGT	840	
CGAAGCGGCG	GCCTTTGGTG	CCTACTTCAA	TCAGGGTCAA	ATCTGCATGT	CCACTGAGCG	900	
TCTGATTGTG	ACAGCAGTCG	CAGACGCCTT	TGTTGAAAAG	CTGGCGAGGA	AGGTGCCCCAC	960	
ACTGCGTGCT	GGCGATCCTA	ATGATCCGCA	ATCGGTCTTG	GGTTCGTTGA	TTGATGCCAA	1020	
TGCAGGTCAA	CGCATCCAGG	TTCTGGTCTGA	TGATGCGCTC	GCAAAAGGCG	CGCAATGGAA	1080	
TTGGCCCAGC	GCGTCGATTC	GGGCATTTGC	CATATCAATG	GACCGACTGT	GCATGACGAG	1140	
GCTCAGATGC	CATTTCGGTG	GGTGAAGTCC	AGCGGCTACG	GCAGCTTCGG	CAGTCGAGCA	1200	
TCGATTGAGC	ACTTTACCCA	GCTGCGCTGG	CTGACCATTG	AGAATGGCCC	GCGGCACTAT	1260	
CCAATCTAAA	TCGATCTTCG	GGCGCCGCGG	GCATCATGCC	CGCGGCGCTC	GCCTCATTTT	1320	
AATCTCTAAC	TTGATAAAAA	CAGAGCTGTT	CTCCGGTCTT	GGTGGATCAA	GGCCAGTCGC	1380	
GGAGAGTCTC	GAAGAGGAGA	GTACAGTGAA	CGCCGAGTCC	ACATTGCAAC	CGCAGGCATC	1440	
ATCATGCTCT	GCTCAGCCAC	GCTACCGCAG	TGTGTCGATT	GGTCATCCTC	CGGTTGAGGT	1500	
TACGCAAGAC	GCTGGAGGTA	TTGTCCGGAT	GCGTTCTCTC	GAGGCGCTTC	TTCCCTTCCC	1560	
GGGTGGAATT	C					1571	

序列 16

GAATTCCGCG	GTCGGCGAAA	GTTGATGCGC	TGTATCGTGG	TGAAGATCAA	TCCATGCTGC	60
GTGACGAGGC	CACACTGTGA	GTTGGTCAGG	GGGGGCTTAC	TCGGCGTTTT	CCGACACTGC	120
GTTGGTTGCG	GCAGTGCGCA	CCCCCTGGAT	TGATTGCGGG	GGTGCCCTGT	CGCTGGTGTG	180
GCCTATCGAC	TTAGGGGTAA	AGGTCGCTCG	CGAAGTTCTG	ATGCGTGCGT	CGCTTGAACC	240
ACAAATGGTC	GATAGCGTAC	TCGCAGGCTC	TATGGCTCAA	GCAAGCTTTG	ATGCTTACCT	300
GCTCCC CGCG	CACATTGGCT	TGTACAGCGG	TGTTCCCAAG	TCGGTTCCGG	CCTTGGGGGT	360
GCAGCGCATT	TGCGGCACAG	GCTTCGAACT	GCTTCGGCAG	GCCGGCGAGC	AGATTTCCCA	420
AGGCGCTGAT	CACGTGCTGT	GTGTGCGGGC	AGAGTCCATG	TCGCGTAACC	CCATCGCGTC	480
GTATACACAC	CGGGGCGGGT	TCCGCCTCGG	TGCGCCCGTT	GAGTTCAAGG	ATTTTTTGTG	540
GGAGGCATTG	TTTGATCCTG	CTCCAGGACT	CGACATGATC	GCTACCGCAG	AAAACCTGGG	600
GACAGCAAGC	GAACCGGAAT	TGCCAGCTGG	GGCGCCCTCT	GGTAAGGTTG	GGAAGCCCTG	660
CAAAGTAAAC	TGGATTGGCTT	TCTTGCCGCC	AAGGATCTGA	TGGCGCAGGG	GATCAAGATC	720
TGATCAAGAG	ACAGGATGAG	GATCGTTTCG	CATGATTGAA	CAAGATGGAT	TGCACGCAGG	780
TTCTCCGGCC	GCTTGGGTGG	AGAGGCTATT	CGGCTATGAC	TGGGCACAAC	AGACAATCCG	840
CTGCTCTGAT	GCCGCCGTGT	TCCGGCTGTC	AGCGCAGGGG	CGCCCGGTTT	TTTTTGTCAA	900
GACCGACCTG	TCCGGTGCCC	TGAATGAACT	GCAGGACGAG	GCAGCGCGGC	TATCGTGGCT	960
GGCCACGACG	GGCGTTCCCTT	GCGCAGCTGT	GCTCGACGTT	GTCAC TGAAG	CGGGAAGGGA	1020
CTGGCTGCTA	TTGGGCGAAG	TGCCGGGGCA	GGATCTCCTG	TCATCTCACC	TTGCTCCTGC	1080
CGAGAAAGTA	TCCATCATGG	CTGATGCAAT	GCGGCGGCTG	CATACGCTTG	ATCCGGCTAC	1140
CTGCCCATTC	GACCACCAAG	CGAAACATCG	CATCGAGCGA	GCACGTACTC	GGATGGAAGC	1200
CGGTCTTTGTC	GATCAGGATG	ATCTGGACGA	AGAGCATCAG	GGGCTCGCGC	CAGCCGAACT	1260
GTTCGCCAGG	CTCAAGGCGC	GCATGCCCGA	CGGCGAGGAT	CTCGTCGTGA	CCCATGGCGA	1320
TGCCTGCTTG	CCGAATATCA	TGGTGGA AAA	TGGCCGCTTT	TCTGGATTCA	TCGACTGTGG	1380
CCGGCTGGGT	GTGGCGGACC	GCTATCAGGA	CATAGCGTTG	GCTACCCGTG	ATATTGCTGA	1440
AGAGCTTGGC	GGCGAATGGG	CTGACCGCTT	CCTCGTGCTT	TACGGTATCG	CCGCTCCCGA	1500
TTGCGAGCGC	ATCGCCTTCT	ATCGCCTTCT	TGACGAGTTC	TTCTGAGCGG	GACTCTGGGG	1560
TTCGAAATGA	CCGACCAAGC	GACGCCCATT	GAGGGCGCAA	GAGGAGAAAT	GGATTGACCA	1620
AGAGATCGTG	GCTGTTACGG	ATGAACAGTT	CGATTTAGAG	GGCTACAACA	GTCGAGCAAT	1680
TGAAGTGCCT	CGGAAGGCAA	AATTGTTGAT	CGTGACAGTC	ATCCGCGGCC	TAGCAGTCTT	1740
TGAAGCCCTT	TCCCGATTGA	AGCCTGTTCA	TTCTGGCGGG	GTGCAGACTG	CGGGCAACAG	1800
CTGTGCCGTA	GTGGACGGCG	CCGCGGCGGC	TTTGGTGGCT	CGAGAGTCGT	CTGCGACACA	1860
GCCGGTCTTG	GCTAGGATAC	TGGCTACCTC	CGTAGTCGGG	ATCGAGCCCG	AGCATATGGG	1920
GCTCGGCCCT	GCGCCCGCGA	TTCGCCTGCT	GCTTGCGCGT	AGTGATCTTA	GTTTGAGGGA	1980
TATCGACCTC	TTTGAGATAA	ACGAGGCGCA	GGCCGCCCAA	GTTCTAGCGG	TACAGCATGA	2040
ATTGGGTATT	GAGCACTCAA	AAC'TTAATAT	TTGGGGCGGG	GCCATTGCAC	TTGGACACCC	2100
GCTTGCCGCG	ACGGGATTGC	GTCTCTGCAT	GACCCTCGCT	CACCAATTGC	AAGCTAATAA	2160

CTTTCGATAT	GGAATTGCCT	CGGCATGCAT	TGGTGGGGGA	CAGGGGATGG	CGGTTCTTTT	2220
AGAGAATCCC	CAC TTCGGTT	CGTCCTCTGC	ACGAAGTTCG	ATGATTAACA	GAGTTGACCA	2280
CTATCCACTG	AGCTAACGGG	CATCTCCTTT	GTTGCTTTGA	GGTGGCGCAC	GAAGGAGGGC	2340
TCGAAAATCT	CTGCTAAAAA	CAAGAAGAAG	GAACAGGGAA	CATGATTAGT	TTCGCTCGTA	2400
TGGCAGAAAG	TTTAGGAGTC	CAGGCTAAAC	TTGCCCTTGC	CTTCGCACTC	GTATTATGTG	2460
TCGGGCTGAT	TGTTACCGGC	ACGGGTTTCT	ACAGTGTACA	TACCTTGTCA	GGGTTGGTGG	2520
GAATTC						2526

序列 17

GAATTCCGCG	GTCGGCGAAA	GTTGATGCGC	TGTATCGTGG	TGAAGATCAA	TCCATGCTGC	60
GTGACGAGGC	CACACTGTGA	GTTGGTCAGG	GGGGGCTTAC	TCGGCGTTTT	CCGACACTGC	120
GTTGGTTGCG	GCAGTGCGCA	CCCCCTGGAT	TGATTGCGGG	GGTGCCCTGT	CGCTGGTGTC	180
GCCTATCGAC	TTAGGGGTAA	AGGTCGCTCG	CGAAGTTCCTG	ATGCGTGCGT	CGCTTGAACC	240
ACAAATGGTC	GATAGCGTAC	TCGCAGGCTC	TATGGCTCAA	GCAAGCTTTG	ATGCTTACCT	300
GCTCCCGCGG	CACATTGGCT	TGTACAGCGG	TGTTCCCAAG	TCGGTTCCGG	CCTTGGGGGT	360
GCAGCGCATT	TGCGGCACAG	GCTTCGAAC	GCTTCGGCAG	GCCGGCGAGC	AGATTGCCCA	420
AGGCGCTGAT	CACGTGCTGT	GTGTCGCGGC	AGAGTCCATG	TCGCGTAACC	CCATCGCGTC	480
GTATACACAC	CGGGGCGGGT	TCCGCTCCGG	TGCGCCCGTT	GAGTTCAAGG	ATTTTTTGTG	540
GGAGGCATTG	TTTGATCCTG	CTCCAGGACT	CGACATGATC	GCTACCGCAG	AAAACCTGGG	600
GGAGAGGCGG	TTTGCGTATT	GGGCGCATGC	ATAAAACTG	TTGTAATTCA	TTAAGCATTC	660
TGCCGACATG	GAAGCCATCA	CAAACGGCAT	GATGAACCTG	AATCGCCAGC	GGCATCAGCA	720
CCTTGTCGCC	TTGCGTATAA	TATTTGCCCA	TGGACGCACA	CCGTGGAAC	GGATGAAGGC	780
ACGAACCCAG	TTGACATAAG	CCTGTTCCGT	TCGTAAACTG	TAATGCAAGT	AGCGTATGCG	840
CTCACGCAAC	TGGTCCAGAA	CCTTGACCGA	ACGCAGCGGT	GGTAACGGCG	CAGTGGCGGT	900
TTTCATGGCT	TGTTATGACT	GTTTTTTTGT	ACAGTCTATG	CCTCGGGCAT	CCAAGCAGCA	960
AGCGCGTTAC	GCCGTGGGTC	GATGTTTGAT	GTTATGGAGC	AGCAACGATG	TTACGCAGCA	1020
GCAACGATGT	TACGCAGCAG	GGCAGTCGCC	CTAAAACAAA	GTTAGGTGGC	TCAAGTATGG	1080
GCATCATTCG	CACATGTAGG	CTCGGCCCTG	ACCAAGTCAA	ATCCATGCGG	GCTGCTCTTG	1140
ATCTTTTCGG	TCGTGAGTTC	GGAGACGTAG	CCACCTACTC	CCAACATCAG	CCGGAATCCG	1200
ATTACCTCGG	GAACCTGCTC	CGTAGTAAGA	CATTCATCGC	GCTTGCTGCC	TTCGACCAAG	1260
AAGCGGTTGT	TGGCGCTCTC	GCGGCTTACG	TTCTGCCAG	GTTTGAGCAG	CCGCGTAGTG	1320
AGATCTATAT	CTATGATCTC	GCAGTCTCCG	GCGAGCACCG	GAGGCAGGGC	ATTGCCACCG	1380
CGCTCATCAA	TCTCCTCAAG	CATGAGGCCA	ACGCGCTTGG	TGCTTATGTG	ATCTACGTGC	1440
AAGCAGATTA	CGGTGACGAT	CCCGCAGTGG	CTCTCTATAC	AAAGTTGGGC	ATACGGGAAG	1500
AAGTGATGCA	CTTTGATATC	GACCCAAGTA	CCGCCACCTA	ACAATTCGTT	CAAGCCGAGA	1560
TCGGCTTCCC	ATTGAGGGCG	CAAGAGGAGA	AATGGATTGA	CCAAGAGATC	GTGGCTGTTA	1620
CGGATGAACA	GTTCGATTTA	GAGGGCTACA	ACAGTCGAGC	AATTGAACTG	CCTCGGAAGG	1680
CAAAATTGTT	GATCGTGACA	GTCATCCGCG	GCCTAGCAGT	CTTTGAAGCC	CTTTCCCGAT	1740
TGAAGCCTGT	TCATTCTGGC	GGGGTGACAG	CTGCGGGCAA	CAGCTGTGCC	GTAGTGACG	1800
GCGCCGCGGC	GGCTTTGGTG	GCTCGAGAGT	CGTCTGCGAC	ACAGCCGGTC	TTGGCTAGGA	1860
TACTGGCTAC	CTCCGTAGTC	GGGATCGAGC	CCGAGCATAT	GGGGCTCGGC	CCTGCGCCCG	1920
CGATTGCGCT	GCTGCTTGCG	CGTAGTGATC	TTAGTTTGAG	GGATATCGAC	CTCTTTGAGA	1980
TAAACGAGGC	GCAGGCCGCC	CAAGTTCTAG	CGGTACAGCA	TGAATTGGGT	ATTGAGCACT	2040
CAAAACTTAA	TATTTGGGGC	GGGGCCATTG	CACTTGAGCA	CCCGCTTGCC	GCGACCGGAT	2100
TGCGTCTCTG	CATGACCCTC	GCTCACCAAT	TGCAAGCTAA	TAACTTTCGA	TATGGAATTG	2160

CCTCGGCATG	CATTGGTGGG	GGACAGGGGA	TGGCGGTTCT	TTTAGAGAAT	CCCCACTTCG	2220
GTTCGTCCCTC	TGCACGAAGT	TCGATGATTA	ACAGAGTTGA	CCACTATCCA	CTGAGCTAAC	2280
GGGCATCTCC	TTTGTTCCTT	TGAGGTGGCG	CACGAAGGAG	GGCTCGAAAA	TCTCTGCTAA	2340
AAACAAGAAG	AAGGAACAGG	GAACATGATT	AGTTTCGCTC	GTATGGCAGA	AAGTTTAGGA	2400
GTCCAGGCTA	AAC TTGCCCT	TGCCTTCGCA	CTCGTATTAT	GTGTCGGGCT	GATTGTTACC	2460
GGCACGGGTT	TCTACAGTGT	ACATACCTTG	TCAGGGTTGG	TGGGAATTC		2509

序列 18

GAATTCGCGG	GTCCGGCGAAA	GTTGATGCGC	TGTATCGTGG	TGAAGATCAA	TCCATGCTGC	60
GTGACGAGGC	CACACTGTGA	GTTGGTCAGG	GGGGGCTTAC	TCGGCGTTTT	CCGACACTGC	120
GTTGGTTGCG	GCAGTGCACA	CCCCCTGGAT	TGATTGCGGG	GGTGCCCTGT	CGCTGGTGTC	180
GCCTATCGAC	TTAGGGGTAA	AGGTCGCTCG	CGAAGTTCTG	ATGCGTGCGT	CGCTTGAACC	240
ACAAATGGTC	GATAGCGTAC	TCGCAGGCTC	TATGGCTCAA	GCAAGCTTTG	ATGCTTACCT	300
GCTCCCGCGG	CACATTGGCT	TGTACAGCGG	TGTTCCCAAG	TCGGTTCCGG	CCTTGGGGGT	360
GCAGCGCATT	TGCGGCACAG	GCTTCGAACT	GCTTCGGCAG	GCCGGCGAGC	AGATTTCCCA	420
AGGCGCTGAT	CACGTGCTGT	GTGTCGCGGC	AGAGTCCATG	TCGCGTAACC	CCATCGCGTC	480
GTATACACAC	CGGGGCGGGT	TCCGCCTCGG	TGCGCCCGTT	GAGTTCAAGG	ATTTTTTG TG	540
GGAGGCATTG	TTTGATCCTG	CTCCAGGACT	CGACATGATC	GCTACCGCAG	AAAACCTGGC	600
GCGCATTGAG	GGCGCAAGAG	GAGAAATGGA	TTGACCAAGA	GATCGTGGCT	GTTACGGATG	660
AACAGTTCGA	TTTAGAGGGC	TACAACAGTC	GAGCAATTGA	ACTGCCTCGG	AAGGCAAAAT	720
TGTTGATCGT	GACAGTCATC	CGCGGCCTAG	CAGTCTTTGA	AGCCCTTTCC	CGATTGAAGC	780
CTGTTTCAATC	TGGCGGGGTG	CAGACTGCGG	GCAACAGCTG	TGCCGTAAGT	GACGGCGCCG	840
CGGCGGCTTT	GGTGGCTCGA	GAGTCGTCTG	CGACACAGCC	GGTCTTGGCT	AGGATACTGG	900
CTACCTCCGT	AGTCGGGATC	GAGCCCGAGC	ATATGGGGCT	CGGCCCTGCG	CCCGCGATTG	960
GCCTGCTGCT	TGCGCGTAGT	GATCTTAGTT	TGAGGGATAT	CGACCTCTTT	GAGATAAAGC	1020
AGGCGCAGGC	CGCCCAAGTT	CTAGCGGTAC	AGCATGAATT	GGGTATTGAG	CAC'TCAAAAC	1080
TTAATATTTG	GGGCGGGGCC	ATTGCACTTG	GACACCCGCT	TGCCGCGACC	GGATTGCGTC	1140
TCTGCATGAC	CCTCGCTCAC	CAATTGCAAG	CTAATAACTT	TCGATATGGA	ATTGCCTCGG	1200
CATGCATTGG	TGGGGGACAG	GGGATGGCGG	TTCTTTTAGA	GAATCCCCAC	TTCGGTTCGT	1260
CCTCTGCACG	AAGTTCGATG	ATTAACAGAG	TTGACCACTA	TCCACTGAGC	TAACGGGCAT	1320
CTCCTTTGTT	GCTTTGAGGT	GGCGCACGAA	GGAGGGCTCG	AAAATCTCTG	CTAAAAACAA	1380
GAAGAAGGAA	CAGGGAACAT	GATTAGTTTC	GCTCGTATGG	CAGAAAGTTT	AGGAGTCCAG	1440
GCTAAACTTG	CCCTTGCTTT	CGCACTCGTA	TTATGTGTCG	GGCTGATTGT	TACCGGCACG	1500
GGTTTCTACA	GTGTACATAC	CTTGTCAGGG	TTGGTGGGAA	TTC		1543

说明书附图



图 1a



图 1b

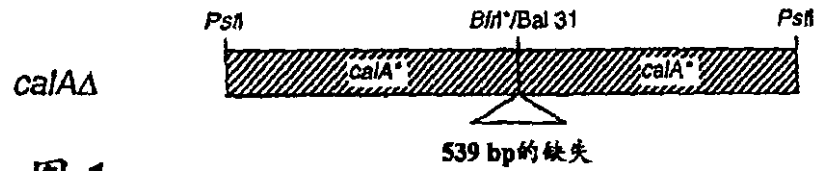


图 1c



图 1d



图 1e

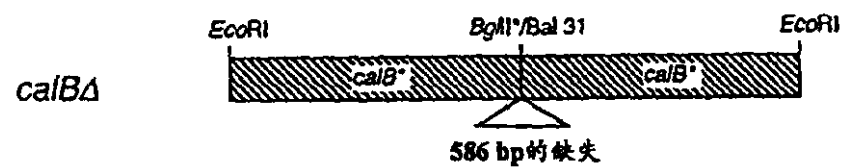


图 1f

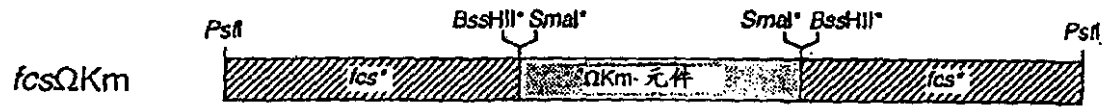


图 1g



图 1h

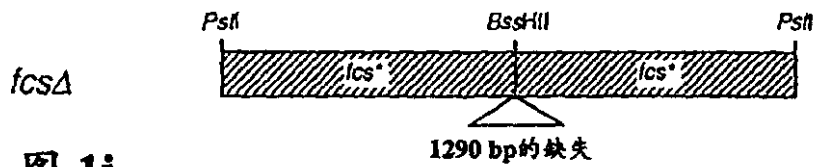


图 1i



图 1j

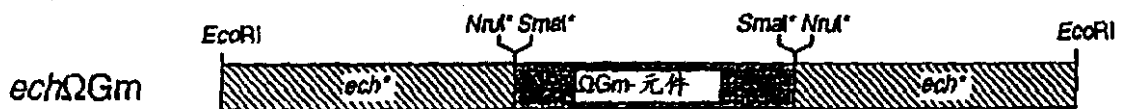


图 1k

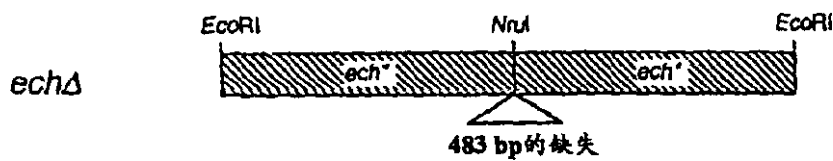


图 1l



图 1m



图 1n

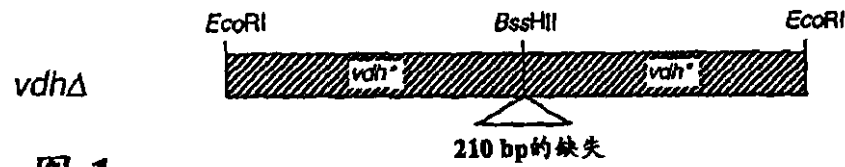


图 1o



图 1p



图 1q

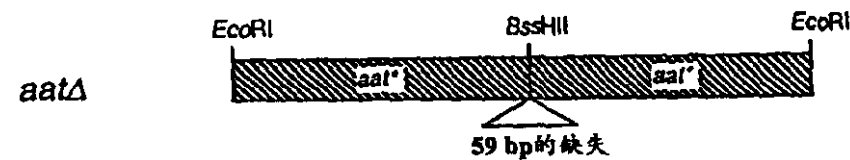


图 1r