



(19)
Bundesrepublik Deutschland
Deutsches Patent- und Markenamt

(10) **DE 699 38 579 T2** 2009.06.04

(12) **Übersetzung der europäischen Patentschrift**

(97) **EP 0 992 586 B1**

(21) Deutsches Aktenzeichen: **699 38 579.2**

(96) Europäisches Aktenzeichen: **99 119 184.2**

(96) Europäischer Anmeldetag: **07.10.1999**

(97) Erstveröffentlichung durch das EPA: **12.04.2000**

(97) Veröffentlichungstag
der Patenterteilung beim EPA: **23.04.2008**

(47) Veröffentlichungstag im Patentblatt: **04.06.2009**

(51) Int Cl.⁸: **C12N 15/12** (2006.01)

C07K 14/75 (2006.01)

C07K 14/78 (2006.01)

C07K 14/51 (2006.01)

C07K 14/47 (2006.01)

C07K 14/495 (2006.01)

C12P 21/02 (2006.01)

(30) Unionspriorität:

169768 09.10.1998 US

(84) Benannte Vertragsstaaten:

DE, ES, FR, GB, IT

(73) Patentinhaber:

United States Surgical Corp., Norwalk, Conn., US

(72) Erfinder:

Gruskin, Elliott A., Killingworth, CT 06419, US;

Buechter, Douglas D., Wallingford, CT 06492, US;

Zhang, Guanghui, Guilford, CT 06473, US;

Connolly, Kevin, Los Angeles, CA 90066, US

(74) Vertreter:

HOFFMANN & EITLE, 81925 München

(54) Bezeichnung: **Extrazelluläre Matrixproteine mit modifizierten Aminosäuren**

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

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

Beschreibung

Hintergrund

1. Technisches Gebiet

[0001] Technisch hergestellte Polypeptide und chimäre Polypeptide mit eingebauten Aminosäuren, die die Eigenschaften dieser Polypeptide verbessern oder in anderer Weise modifizieren.

2. Beschreibung im Zusammenhang stehender Techniken

[0002] Die Gentechnik ermöglicht die Herstellung von Polypeptiden, die von einem Organismus in einen anderen überführt werden. Wenn dies getan wird, wird ein Teil des Produktionsapparats, der dem ursprünglichen Wirt eigen ist, in einen Empfänger überführt. Häufig hat der ursprüngliche Wirt im Zusammenhang mit der Polypeptidherstellung einige einzigartige Prozessierungswege hervorgebracht, die in dem Empfänger nicht enthalten sind oder überführt wurden. Beispielsweise ist allseits bekannt, daß Säugerzellen eine komplexe Anordnung posttranslationaler Enzymsysteme enthalten, die Proteinprodukten dieser Systeme einzigartige Eigenschaften verleihen. Wenn ein Gen, das ein Protein codiert, normalerweise in Säugerzellen hergestellt wird, in eine bakterielle oder Hefezelle überführt wird, kann das Protein nicht einer derartigen posttranslationalen Modifizierung unterworfen werden, und das Protein kann nicht so funktionieren, wie ursprünglich beabsichtigt.

[0003] Normalerweise schließt der Prozeß der Polypeptid- oder Proteinsynthese in lebenden Zellen die Transkription von DNA in RNA und die Translation von RNA in Protein ein. Drei Formen von RNA sind an der Proteinsynthese beteiligt: Messenger-RNA (mRNA), welche die genetische Information zu den Ribosomen trägt, welche aus ribosomaler RNA (rRNA) bestehen, während die Transfer-RNA (tRNA) freie Aminosäuren in dem Pool einer Zelle verbindet. Aminosäuren/tRNA-Komplexe reihen sich neben den Codons der mRNA auf, wobei die tatsächliche Erkennung und die Bindung durch die tRNA vermittelt wird. Zellen können bis zu 20 Aminosäuren enthalten, die in Sequenzen mit verschiedenen Abfolgen in Proteinen kombiniert und eingebaut werden. Jede Aminosäure ist von den anderen 19 Aminosäuren verschieden und wird durch Enzyme, die als Aminoacyl-tRNA-Synthetasen bekannt sind, mit der tRNA beladen. Nach einer allgemeinen Regel sind Aminosäure/tRNA-Komplexe ziemlich spezifisch und normalerweise wird auf ein Molekül mit einer exakten stereochemischen Konfiguration nur durch eine besondere Aminoacyl-tRNA-Synthetase eingewirkt.

[0004] In vielen lebenden Zellen werden einige Aminosäuren von der umliegenden Umgebung aufgenommen, und einige werden innerhalb der Zelle aus Vorläufern synthetisiert, welche wiederum von außerhalb der Zelle assimiliert worden sind. Unter einigen Umständen ist eine Zelle auxotroph, d. h. sie benötigt eine spezifische Wachstumssubstanz neben dem Minimum, das für den normalen Metabolismus und Reproduktion benötigt wird, welche sie aus der umliegenden Umgebung gewinnen muß. Einige Auxotrophe hängen von der externen Umgebung an, einige Aminosäuren verabreicht zu bekommen. Dieses Merkmal ermöglicht ein Aminosäure-Analogon in Proteine, die von Auxotrophen hergestellt werden, eingebaut zu werden, indem von den relativ seltenen Ausnahmen der oben erwähnten Regel hinsichtlich der stereochemischen Spezifität von Aminoacyl-tRNA-Synthetasen profitiert wird. Beispielsweise ist Prolin eine derartige Ausnahme, d. h. die auf die Aminosäure einwirkenden Enzyme, die für die Synthese des Prolin-tRNA-Komplex verantwortlich sind, sind nicht so spezifisch wie andere. Infolgedessen sind einige Prolin-Analoga in bakterielle, pflanzliche und Tierzellsysteme eingebaut worden. Siehe Tan et al., Proline Analogues Inhibit Human Skin Fibroblast Growth and Collagen Production in Culture, Journal of Investigative Dermatology, 80: 261–267 (1983).

[0005] Ein Verfahren zum Einbau nicht-natürlicher Aminosäuren in Proteine wird beispielsweise in Noren et al., A General Method For Site-Specific Incorporation of Unnatural Amino Acids Into Proteins, Science, Bd. 244, S. 182–188 (1989) beschrieben, wobei chemisch acylierte Suppressor-tRNA verwendet wird, um eine Aminosäure infolge eines Stopp-Codons einzubauen, der ein Codon substituiert, der einen Rest von Interesse codiert. Siehe ebenfalls Dougherty et al., Synthesis of a Genetically Engineered Repetitive Polypeptide Containing Periodic Selenomethionin Residues, Macromolecules, Bd. 26, Nr. 7. S. 1779–1871 (1993), welche das Unterwerfen von Methionin-auxotrophen E. coli gegenüber einem Selenomethionin-haltigen Medium beschreibt, und auf der Grundlage experimenteller Daten beschreibt, daß Selenomethionin Methionin vollständig in allen durch die Zelle hergestellten Proteinen ersetzen kann.

[0006] cis-Hydroxy-L-prolin ist verwendet worden, um dessen Auswirkungen auf Kollagen durch Einbau in eukaryontische Zellen, wie beispielsweise kultivierte normale Hautfibroblasten (siehe Tan et al., oben) und in Sehnenzellen aus Hühnerembryonen (siehe z. B. Uitto et al., Prokollagen Polypeptides Containing cis-4-Hydro-

xy-L-prolin are Overglycosylated and Secreted as Nonhelical Pro- γ -Chains, *Archives of Biochemistry and Biophysics*, 182: 1: 214–221 (1978)) zu untersuchen. Jedoch haben Forscher herausgefunden, daß trans-4-Hydroxyprolin sich nicht mit der Prolin-spezifischen tRNA prokaryontischer *E. coli* verbindet. Siehe Papas et al., Analysis of the Amino Acid Binding to the Prolin Transfer Ribonucleic Acid Synthetase of *Escherichia coli*, *Journal of Biological Chemistry*, 245: 7: 1588–1595 (1970). Ein weiterer nicht erfolgreicher Versuch, trans-4-Hydroxyprolin in Prokaryonten einzubauen, wird von Deming et al. beschrieben. In *Vitro Incorporation of Proline Analogs into Artificial Proteins*, *Poly. Mater. Sci. Engin. Proceed.*, Bd. 71, S. 673–674 (1994). Deming et al. berichten über die Überwachung des Potentials zum Einbau bestimmter Prolin-Analoga, d. h. L-Azetidin-2-carbonsäure, L- γ -Thiaprolin, 3,4-Dihydroprolin und L-trans-4-Hydroxyprolin, in künstliche Proteine, die in *E. coli*-Zellen exprimiert werden. Nur L-Azetidin-2-carbonsäure, L- γ -Thioprolin und 3,4-Dehydroprolin werden als in Proteine in *E. coli* in vivo eingebaut, beschrieben.

[0007] Extrazelluläre Matrix-Proteine ("EMPs") werden in den Räumen um und nahe den Zellen multizellulärer Organismen gefunden und sind typischerweise fibröse Proteine zweier funktionaler Typen: hauptsächlich strukturell, z. B. Kollagen und Elastin und hauptsächlich adhäsiv, z. B. Fibronektin und Laminin. Kollagene sind eine Familie fibröser Proteine, die typischerweise von Bindegewebszellen sezerniert werden. 20 verschiedene Kollagen-Ketten sind identifiziert worden, die sich aus insgesamt ungefähr 10 verschiedenen Kollagen-Molekülen zusammenbauen. Eine allgemeine Diskussion von Kollagen wird von Alberts et al., *The Cell*, Garland Publishing, S. 802–823 (1989) bereitgestellt. Andere fibröse oder filamentöse Proteine schließen Typ I IF-Proteine, z. B. Keratine; Typ II-IF-Proteine, z. B. Vimentin, Desmin und gliales fibrilläres saures Protein; Typ III IF-Proteine, z. B. Neurofilament-Proteine; und Typ IV IF-Proteine, z. B. nukleare Laminine ein.

[0008] Typ I-Kollagen ist die häufigste Form der fibrillären, interstitiellen Kollagene und ist der Hauptbestandteil der extrazellulären Matrix. Kollagen-Monomere bestehen aus ungefähr 1000 Aminosäureresten in einer sich wiederholenden Anordnung aus Gly-X-Y-Triplets. Ungefähr 35% der X- und Y-Positionen werden durch Prolin und trans-4-Hydroxyprolin besetzt. Kollagen-Monomere verbinden sich in Dreifach-Helices, die aus einer $\alpha 2$ - und zwei $\alpha 1$ -Ketten bestehen. Die Dreifach-Helices verbinden sich in Fibrillen, die in festen Bündeln angeordnet sind. Bündel aus Kollagen-Fibrillen sind ferner so angeordnet, daß sie ein Gerüst für die extrazelluläre Matrix bilden.

[0009] In Säugerzellen hat die posttranslationale Modifizierung von Kollagen an den endgültigen chemischen und physikalischen Eigenschaften teil und schließt den proteolytischen Verdau von Pro-Regionen, die Hydroxylierung von Lysin und Prolin und die Glycosylierung hydroxylierten Lysins ein. Der proteolytische Verdau von Kollagen schließt das Abspalten der Pro-Region der N- und C-Termini ein. Es ist bekannt, daß die Hydroxylierung von Prolin für die mechanischen Eigenschaften von Kollagen wesentlich ist. Kollagen mit niedrigem Gehalt an 4-Hydroxyprolin hat schlechte mechanische Eigenschaften, wie durch die Folgen, die mit Skorbut verbunden sind, hervorgehoben wird. 4-Hydroxyprolin verleiht der Dreifachhelix durch eine Wasserstoffbindung und durch die Beschränkung der Rotation der C-N-Bindungen in dem Polypeptid-Rückgrat Stabilität. In Abwesenheit einer stabilen Struktur haben natürlich auftretende zelluläre Enzyme daran teil, das Kollagen-Polypeptid zu degradieren.

[0010] Die strukturellen Eigenschaften von Typ I-Kollagen zusammen mit dessen allgemein wahrgenommener Biokompatibilität machen es zu einem beliebten chirurgischen Implantatmaterial. Kollagen wird aus Rinderhaut oder -sehnen gereinigt und verwendet, um eine Vielzahl medizinischer Hilfsmittel zu modellieren, einschließlich Hämostaten, implantierbare Gele, Medikamentenverabreichungsvehikel und Knochenersatz. Wenn es jedoch in Menschen implantiert wird, kann Rinderkollagen akute und verzögerte Immunreaktionen verursachen.

[0011] Infolgedessen haben Forscher versucht, menschliches rekombinantes Kollagen mit allen seinen strukturellen Eigenschaften mit Hilfe der Gentechnik in kommerziellen Mengen herzustellen. Unglücklicherweise war die Produktion von Kollagen durch kommerzielle Massenproduzenten von Proteinen, wie beispielsweise *E. coli*, nicht erfolgreich. Ein Hauptproblem liegt darin, daß die umfangreiche posttranslationale Modifizierung von Kollagen durch Enzyme erfolgt, die in *E. coli* nicht vorhanden sind. Da *E. coli*-Zellen es nicht schaffen, eine Prolinhydroxylierung nicht-hydroxylierten Kollagens bereitzustellen, verhindert die Herstellung strukturell einwandfreien Kollagen in kommerziellen Mengen.

[0012] Ein weiteres Problem bei dem Versuch *E. coli* zu verwenden, um menschliches Kollagen herzustellen liegt darin, daß *E. coli* bestimmte Codons bei der Herstellung von Polypeptiden bevorzugen. Obwohl der genetische Code sowohl in prokaryontischen und eukaryontischen Organismen identisch ist, variiert der jeweilige Codon (von den mehreren, die für die meisten Aminosäuren möglich sind), welcher am häufigsten verwendet

wird zwischen Prokaryonten und Eukaryonten beachtlich. Siehe K.-N. Wada, Y. Wada, F. Ishibashi, T. Gojobori und T. Ikemura. *Nucleic Acids Res.* 20. Supplement: 2111–2118, 1992. Die effiziente Expression heterologer (z. B. Säuger-)Gene in Prokaryonten, wie beispielsweise *E. coli*, kann durch das Vorhandensein von Codons in dem Gen, die von *E. coli* selten verwendet werden, negativ beeinflusst werden, und die Expressionsniveaus des heterologen Proteins steigen häufig an, wenn seltene Codons durch häufigere Codons ausgetauscht werden. Siehe z. B. D. P. Williams, D. Regier, D. Akiyoshi, F. Genbauffe und J. R. Murphy. *Nucleic Acids Res.* 16: 10453–10467, 1988 und J.-O. Höög, H. v. Bahr-Lindström, H. Jörnvall und A. Holmgren. *Gene* 43: 13–21, 1986. Man vermutet, daß dieses Phänomen mindestens teilweise mit der Beobachtung zusammenhängt, daß eine geringere Häufigkeit des Auftretens eines besonderen Codons mit einer geringen zellulären Konzentration der Transfer-RNA für diesen Codon zusammenhängt. Siehe T. J. Ikemura, *Mol. Biol.* 158: 573–597, 1982 und T. J. Ikemura, *Mol. Biol.* 146: 1–21, 1–81. So kann die zelluläre tRNA-Konzentration die Translationsrate des Codons limitieren, und dadurch die Gesamttranslationsrate des Proteins vollständiger Länge beeinflussen. Siehe T. J. Ikemura, *Biol.* 146: 1–21, 1981, F. Bonekamp und F. K. Jensen. *Nucleic Acids Res.* 16: 3013–3024, 1988; R. Misra und P. Reeves, *Eur. J. Biochem.* 152: 151–155, 1985; und L. E. Post, G. D. Strycharz, M. Nomura, H. Lewis und P. P. Lewis, *Proc. Natl. Acad. Sci. USA* 76: 1697–1701, 1979. Als Stütze dieser Hypothese dient die Beobachtung, daß Gene für häufige *E. coli*-Proteine allgemein eine Neigung zu gewöhnlich verwendeten Codons zeigen, welche häufige tRNAs darstellen. Siehe T. J. Ikemura, *Mol. Biol.* 146: 1–21, 1981; F. Bonekamp und F. K. Jensen, *Nucleic Acids Res.* 16: 3013–3024, 1988; R. Misra und P. Reeves, *Eur. J. Biochem.* 152: 151–155, 1985; und L. E. Post, G. D. Strycharz, M. Nomura, H. Lewis und P. P. Lewis, *Proc. Natl. Acad. Sci. USA* 76: 1697–1701, 1979. Zusätzlich zur Codonhäufigkeit kann auch das Codonumfeld (d. h. die umliegenden Nukleotide) die Expression beeinflussen.

[0013] Obwohl es scheinen sollte, daß die Substituierung bevorzugter Codons für seltene Codons die Expression heterologer Proteine in Wirtsorganismen den Erwartungen nach steigern sollte, ist dies nicht der Fall. Tatsächlich "ist es nicht möglich gewesen, allgemeine und unzweideutige Regeln zu formulieren, um vorherzusagen, ob der Gehalt an selten verwendeten Codons in einem spezifischen Gen gegenteilig die Effizienz dessen Expression in *E. coli* beeinflussen kann". Siehe Seite 524 von S. C. Makrides (1996), *Strategies for Achieving High-Level Expression of Genes in Escherichia coli*. *Microbiological Reviews* 60, 512–538. So wurden beispielsweise in einem Fall zahlreiche Genfusionen zwischen dem Hefe- α -Faktor und Somatomedin C gemacht, die nur in der codierenden Sequenz verschieden waren. In diesen Experimenten wurden keine Korrelation zwischen der Neigung der Codons und den Expressionsniveaus in *E. coli* gefunden. J. F. Ernst und E. Kawashima (1988), *J. Biotechnology*, 7, 1–10. In einem anderen Fall wurde gezeigt, daß trotz der größeren Häufigkeit optimaler Codons in einem synthetischen β -Globin-Gen, verglichen mit der nativen Sequenz, kein Unterschied in der Protein-Expression bei diesen zwei Konstrukten gefunden wurde, wenn sie hinter dem T7-Promotor plaziert wurden. Hernan et al. (1992), *Biochemistry*, 31, 8619–8628. Andererseits gibt es viele Beispiele für Proteine mit relativ hohem Prozentsatz seltener Codons, die gut in *E. coli* exprimiert werden. Eine Tabelle, die einige dieser Beispiele auflistet und eine allgemeine Diskussion kann in A. J. Makoff et al. (1989), *Nucleic Acids Research*, 17, 10191–10202 gefunden werden. In einem Fall hat sogar die Einschleusung nicht-optimaler, seltener Arginin-Codons am 3'-Ende eines Gens die Ausbeute des exprimierten Proteins erhöht. Y. G. Gursky und R. Sh. Beabealashvili, *Gene* 148, 15–21.

[0014] Wenn man es nicht schafft posttranslationale Modifizierungen, wie eine Hydroxylierung von Prolin beizubringen, kann das Vorhandensein von seltenen Codons für *E. coli* im menschlichen Kollagen an den Schwierigkeiten teilhaben, die bei der Expression menschlicher Kollagen-Gene in *E. coli* vorgefunden werden.

Zusammenfassung

[0015] Ein Verfahren zum Einbau eines Aminosäure-Analogons in ein Polypeptid, das von einer Zelle hergestellt wird, wird bereitgestellt, welches das Bereitstellen einer Zelle einschließt, die aus der Gruppe der prokaryontischen Zellen und eukaryontischen Zellen besteht, wobei hypertones Wachstumsmedium verwendet wird, das mindestens ein Aminosäure-Analogon enthält, das aus der Gruppe ausgewählt ist, die aus trans-4-Hydroxyprolin, 3-Hydroxyprolin und Kombinationen daraus besteht, und das Inkontaktbringen der Zelle mit dem Wachstumsmedium, wobei das mindestens eine Aminosäure-Analogon in der Zelle assimiliert wird und in mindestens ein Polypeptid eingebaut wird.

[0016] Die internationale PTC-Anmeldung WO 97/38710 beschäftigt sich mit der Synthese menschlichen Prokollagens und der Kollagene in rekombinanten DNA-Systemen. Die europäische Patentanmeldung EP 0 704 532 A2 beschäftigt sich mit rekombinanten chimären Proteinen und Verfahren hierfür. M. Haardt et al., *Mol. Gen. Genet.* (1995) 246: 783–786 beschäftigt sich mit osmoprotektivem Prolinbetain als Hauptsubstrat für das Bindungsprotein-abhängige Transportsystem ProU von *Escherichia coli* K-12.

[0017] Ebenfalls bereitgestellt wird ein Verfahren der Substituierung eines Aminosäure-Analogons einer Aminosäure in einem Polypeptid, das in einer Zelle hergestellt wird, die aus der Gruppe ausgewählt wird, die aus einer prokaryontischen Zelle und eukaryontischen Zelle besteht, welches das Bereitstellen einer Zelle einschließt, die aus der Gruppe ausgewählt ist, die aus einer prokaryontischen Zelle und einer eukaryontischen Zelle besteht, das Bereitstellen eines hypertonen Wachstumsmediums, das mindestens ein Aminosäure-Analogon enthält, das aus der Gruppe ausgewählt ist, die aus trans-4-Hydroxyprolin, 3-Hydroxyprolin und Kombinationen daraus besteht, und das Inkontaktbringen der Zelle mit dem Wachstumsmedium, wobei das mindestens eine Aminosäure-Analogon in die Zelle assimiliert wird und als Substituent für mindestens eine natürlich auftretende Aminosäure in mindestens einem Polypeptid eingebaut wird.

[0018] Ein Verfahren zum Kontrollieren der Menge des in ein Polypeptid eingebauten Aminosäuresequenzanalogons wird ebenfalls bereitgestellt, welches das Bereitstellen mindestens einer ersten Zelle einschließt, die aus der Gruppe ausgewählt ist, die aus einer prokaryontischen Zelle und einer eukaryontischen Zelle besteht, das Bereitstellen eines ersten hypertonen Wachstumsmediums, das eine erste vorherbestimmte Menge mindestens eines Aminosäure-Analogons enthält, das aus der Gruppe ausgewählt ist, die aus trans-4-Hydroxyprolin, 3-Hydroxyprolin und Kombinationen daraus besteht, und das Kontaktieren der ersten Zelle mit dem ersten Wachstumsmedium, wobei eine erste Menge eines Aminosäure-Analogons in die erste Zelle assimiliert wird und in mindestens ein Polypeptid eingebaut wird. Mindestens eine zweite Zelle, die aus der Gruppe ausgewählt ist, die aus einer prokaryontischen Zelle und einer eukaryontischen Zelle besteht, wird ebenfalls zusammen mit einem zweiten hypertonen Wachstumsmedium bereitgestellt, welches mindestens eine zweite vorherbestimmte Menge eines Aminosäure-Analogons enthält, das aus der Gruppe ausgewählt ist, die aus trans-4-Hydroxyprolin, 3-Hydroxyprolin und Kombinationen daraus besteht, und die mindestens zweite Zelle wird mit dem zweiten Wachstumsmedium in Kontakt gebracht, wobei eine zweite Menge eines Aminosäure-Analogons durch die zweite Zelle assimiliert wird und in mindestens ein Polypeptid eingebaut wird.

[0019] Ebenfalls bereitgestellt wird ein Verfahren zum Erhöhen der Stabilität eines rekombinanten Polypeptids, das von einer Zelle hergestellt wird, welches das Bereitstellen einer Zelle einschließt, die aus der Gruppe ausgewählt ist, die aus einer prokaryontischen und einer eukaryontischen Zelle besteht, und das Bereitstellen eines hypertonen Wachstumsmediums, das ein Aminosäure-Analogon enthält, welches aus der Gruppe ausgewählt ist, die aus trans-4-Hydroxyprolin, 3-Hydroxyprolin und Kombinationen daraus besteht und das Inkontaktbringen der Zelle mit dem Wachstumsmedium, wobei das Aminosäure-Analogon durch die Zelle assimiliert wird und in ein rekombinantes Polypeptid eingebaut wird, wodurch das Polypeptid stabilisiert wird.

[0020] Ein Verfahren zur Erhöhung der Aufnahme eines Aminosäure-Analogons durch eine Zelle und die Verursachung der Bildung eines Aminosäuresequenzanalogons/tRNA-Komplexes wird ebenfalls bereitgestellt, welches das Bereitstellen einer Zelle einschließt, die aus der Gruppe ausgewählt ist, die aus einer prokaryontischen und einer eukaryontischen Zelle besteht, das Bereitstellen eines hypertonen Wachstumsmediums, das ein Aminosäure-Analogon enthält, das aus der Gruppe ausgewählt ist, die aus trans-4-Hydroxyprolin, 3-Hydroxyprolin und Kombinationen daraus besteht, und das Kontaktieren der Zelle mit dem hypertonen Wachstumsmedium, wobei das Aminosäure-Analogon durch die Zelle assimiliert wird und in einen Aminosäure-Analogon/tRNA-Komplex eingebaut wird. In irgendeiner der oben erwähnten Verfahren kann ein hypertones Wachstumsmedium gegebenenfalls eingebaut werden, um die Aufnahme eines Aminosäure-Analogons in eine Zelle zu erhöhen.

[0021] Eine Zusammensetzung wird beschrieben, die eine Zelle einschließt, die aus der Gruppe ausgewählt ist, die aus einer prokaryontischen Zelle und einer eukaryontischen Zelle besteht, und ein hypertones Medium einschließlich eines Aminosäure-Analogons, das aus der Gruppe ausgewählt wird, die aus trans-4-Hydroxyprolin, 3-Hydroxyprolin und Kombinationen daraus besteht.

[0022] Ebenfalls bereitgestellt wird ein Verfahren zur Herstellung eines extrazellulären Matrix-Proteins (EMP) oder eines Fragments davon, das in der Lage ist, ein Selbstaggregat in einer Zelle bereitzustellen, welche gewöhnlicherweise Prolin nicht hydroxyliert, welches das Bereitstellen einer Nukleinsäuresequenz einschließt, die das EMP oder ein Fragment davon codiert, welches hinsichtlich der Expression in der Zelle durch Substituierung von durch die Zelle bevorzugte Codons für natürlicherweise vorkommende Codons, die von der Zelle nicht bevorzugt werden, optimiert worden ist, das Einbauen der Nukleinsäuresequenz in der Zelle, das Bereitstellen eines hypertonen Wachstumsmediums, das mindestens eine Aminosäure enthält, die aus der Gruppe ausgewählt ist, die aus trans-4-Hydroxyprolin und 3-Hydroxyprolin besteht, und das Kontaktieren der Zelle mit dem Wachstumsmedium, wobei die mindestens eine Aminosäure durch die Zelle assimiliert wird und in das EMP oder Fragment davon eingebaut wird.

[0023] Eine Nukleinsäure, die ein chimäres Protein codiert, wird beschrieben, welches eine Domäne einschließt, die aus einem physiologisch aktiven Peptid stammt, und eine Domäne eines extrazellulären Matrixproteins (EMP), welches in der Lage ist, selbst zu aggregieren. Die Nukleinsäure kann in einen Klonierungsvektor eingebaut werden, welcher dann in die Zelle eingebaut werden kann.

[0024] Ebenfalls beschrieben ist ein chimäres Protein, das eine Domäne eines physiologisch aktiven Peptids und eine Domäne eines extrazellulären Matrixproteins (EMP) einschließt, welches in der Lage ist, ein Selbstaggregat bereitzustellen.

[0025] Ebenfalls beschrieben ist menschliches Kollagen, das von einer prokaryontischen Zelle hergestellt wird, wobei das menschliche Kollagen in der Lage ist, ein Selbstaggregat bereitzustellen.

[0026] Ebenfalls beschrieben wird eine Nukleinsäure, welche ein menschliches extrazelluläres Matrixprotein (EMP) codiert, wobei die Codonverwendung in der Nukleinsäuresequenz die bevorzugte Codonverwendung in einer prokaryontischen Zelle widerspiegelt.

Kurze Beschreibung der Zeichnungen

[0027] [Fig. 1](#) ist eine Plasmidkarte, die pMAL-c2 darstellt.

[0028] [Fig. 2](#) ist eine graphische Darstellung der Konzentration des intrazellulären Hydroxyprolins, auf Grundlage der Konzentration von trans-4-Hydroxyprolin in einer Wachstumskultur über die Zeit.

[0029] [Fig. 2A](#) ist eine graphische Darstellung der Konzentration des intrazellulären Hydroxyprolins als Funktion der Natriumchlorid-Konzentration.

[0030] [Fig. 3A](#) und [Fig. 3B](#) spiegeln die DNA-Sequenz wider, welche menschliches Typ 1 (α_1)-Kollagen codiert (SEQ ID NO: 1).

[0031] [Fig. 4](#) ist eine Plasmidkarte, die pHuCol darstellt.

[0032] [Fig. 5](#) stellt eine DNA-Sequenz dar, welche ein Fragment des menschlichen Typ 1 (α_1)-Kollagens codiert (SEQ ID NO: 2).

[0033] [Fig. 6](#) ist eine Plasmidkarte, die pHuCol-FI darstellt.

[0034] [Fig. 7](#) gibt eine DNA-Sequenz wider, welche ein Kollagen-artiges Peptid codiert, wobei die Region die für das Kollagen-Gen-artige Peptid codiert unterstrichen ist (SEQ ID NO: 3).

[0035] [Fig. 8](#) gibt eine Aminosäuresequenz eines Kollagen-artigen Peptids (SEQ ID NO: 4) wider.

[0036] [Fig. 9](#) ist eine Plasmidkarte, die pCLP darstellt.

[0037] [Fig. 10](#) gibt die DNA-Sequenz wider, die reifes Knochenmorphogenese-Protein codiert (SEQ ID NO: 5).

[0038] [Fig. 11](#) ist eine Plasmidkarte, die pCBC darstellt.

[0039] [Fig. 12](#) ist eine graphische Darstellung des prozentualen Einbaus von Prolin und trans-4-Hydroxyprolin in das Maltose-bindende Protein unter verschiedenen Bedingungen.

[0040] [Fig. 13](#) gibt die Kollagen I (α_1)/BMP-2B chimäre Aminosäuresequenz (SEQ ID NO: 6) wider.

[0041] [Fig. 14A](#) bis [Fig. 14C](#) geben eine chimäre Kollagen I (α_1)/BMP-2B-Nukleotidsequenz wider (SEQ ID NO: 7).

[0042] [Fig. 15](#) gibt die Kollagen I (α_1)/TGF- β_1 -Aminosäuresequenz (SEQ ID NO: 8) wider.

[0043] [Fig. 16A](#) bis [Fig. 16C](#) geben die Kollagen I ($\alpha\alpha$)-TGF- β_1 -Nukleotidsequenz wider. Die kleinen Buchstaben zeigen die nicht-codierende Sequenz an.

- [0044] [Fig. 17A](#) bis [Fig. 17B](#) geben eine Kollagen I ($\alpha\alpha$)/Decorin-Aminosäure-Sequenz (SEQ ID NO: 10) wider.
- [0045] [Fig. 18](#) gibt eine Kollagen I (α_1)/Decorin-Peptidaminosäure-Sequenz (SEQ ID NO: 11) wider.
- [0046] [Fig. 19A](#) bis [Fig. 19D](#) geben eine Kollagen I (α_1)/Decorin-Nukleotid-Sequenz wider (SEQ ID NO: 12).
- [0047] [Fig. 20A](#) bis [Fig. 20C](#) geben eine Kollagen/Decorin-Peptidnukleotid-Sequenz (SEQ ID NO: 13) wider. Kleine Buchstaben geben die nicht-codierende Sequenz an.
- [0048] [Fig. 21](#) gibt den pMal-Klonierungsvektor und die Polylinker-Klonierungsstelle wider.
- [0049] [Fig. 22](#) gibt eine Polylinker-Klonierungsstelle wider, die im pMal-Klonierungsvektor der [Fig. 21](#) enthalten ist (SEQ ID NO: 14).
- [0050] [Fig. 23](#) gibt den pMal-Klonierungsvektor, der ein BMP/Kollagennukleotid-chimären Konstrukt enthält, wider.
- [0051] [Fig. 24](#) gibt den pMal-Klonierungsvektor wider, der ein chimäres Konstrukt aus TGF- β_1 /Kollagen-Nukleotid enthält.
- [0052] [Fig. 25](#) gibt den pMal-Klonierungsvektor wider, der ein chimäres Konstrukt aus Decorin/Kollagen-Nukleotiden enthält.
- [0053] [Fig. 26](#) gibt einen pMal-Klonierungsvektor wider, der ein chimäres Konstrukt aus Decorin-Peptid/Kollagen-Nukleotiden enthält.
- [0054] [Fig. 27A](#) bis [Fig. 27E](#) geben die menschliche Kollagen-Typ I (α_1)-Nukleotidsequenz (SEQ ID NO: 15) und die entsprechende Aminosäuresequenz (SEQ ID NO: 16) wider.
- [0055] [Fig. 28](#) ist eine schematische Darstellung der Herstellung des menschlichen Kollagen-Gens aus synthetischen Oligonukleotiden.
- [0056] [Fig. 29](#) ist eine schematische Darstellung der Aminosäuresequenz der chimären Proteine GST-ColE-Col (SEQ ID NO: 17) und GST-D4 (SEQ ID NO: 18).
- [0057] [Fig. 30](#) ist eine Tabelle, die das Auftreten von vier Prolin- und vier Glycin-Codons in menschlichem Kollagen-Typ I (α_1)-Gen mit der optimierten Codonverwendung (ColECol) wiedergibt.
- [0058] [Fig. 31](#) gibt ein Gel wider, das die Expression und Abhängigkeit der Expression von GST-D4 von Hydroxyprolin widerspiegelt.
- [0059] [Fig. 32](#) stellt ein Gel dar, welches die Expression von GST-D4 in hypertonem Medium zeigt.
- [0060] [Fig. 33](#) ist ein Graph, welcher das zirkuläre Dichroismus-Spektrum von nativem und denaturiertem D4 in neutralem Phosphatpuffer zeigt.
- [0061] [Fig. 34](#) stellt ein Gel dar, welches den Verdau von D4 mit Rinderpepsin wiedergibt.
- [0062] [Fig. 35](#) stellt ein Gel dar, welches die Expression von GS-H Col und GST-ColEcol unter angegebenen Bedingungen darstellt.
- [0063] [Fig. 36](#) stellt ein Gel dar, das die Expression von GST-CM4 im Medium mit oder ohne NaCl und entweder Prolin oder Hydroxyprolin darstellt.
- [0064] [Fig. 37](#) stellt ein Gel mit Proben 6 Stunden nach der Induktion von GST-CM4 dar, welches in E. coli mit variierenden Konzentrationen von NaCl exprimiert wird.
- [0065] [Fig. 38](#) stellt ein Gel 4 Stunden nach der Induktion der Proben von GST-CM4 dar, welche in E. coli mit konstanten Mengen an Hydroxyprolin und variierenden Mengen an Prolin exprimiert werden.

[0066] Die [Fig. 39A–Fig. 39E](#) geben die Nukleotidsequenz (SEQ ID NO: 19) und Aminosäuresequenz (SEQ ID NO: 20) von HuCol^{Ec}, den helikalen Bereich von menschlichen Typ I (α_1)-Kollagen plus der 17 Amino-terminalen extrahelikalen Aminosäuren und 26 Carboxy-terminalen extrahelikalen Aminosäuren wieder, wobei die Codonverwendung für E. coli optimiert wurde.

[0067] [Fig. 40](#) stellt die Sequenz und die Restriktionskarten synthetischer Oligos wieder, die für die Rekonstruktion der ersten 243 Basenpaare des menschlichen Typ I (α_1)-Kollagen-Gens mit der optimierten E. coli-Codon-Verwendung verwendet wurden. Die synthetischen Oligos werden als N1-1 (SEQ ID NO: 21), N1-2 (SEQ ID NO: 22), N1-3 (SEQ ID NO: 23) und N1-4 (SEQ ID NO: 24) bezeichnet.

[0068] [Fig. 41](#) stellt eine Plasmidkarte von pBSN1-1 dar, welches ein 114 Basenpaarfragment des menschlichen Kollagens vom Typ I (α_1) mit optimierter E. coli-Codonverwendung enthält.

[0069] [Fig. 42](#) stellt die Nukleotid (SEQ ID NO: 25)- und Aminosäure (SEQ ID NO: 26)-SEQ ID NO: eines Fragments des menschlichen Kollagens Typ I (α_1)-Gens mit optimierter E. coli-Codonverwendung dar, welches vom Plasmid pBSN1-1 codiert wird.

[0070] [Fig. 43](#) stellt eine Plasmidkarte von pBSN1-2 dar, welches ein 243 Basenpaare großes Fragment des menschlichen Kollagens Typ I (α_1) mit optimierter E. coli-Codonverwendung enthält.

[0071] [Fig. 44](#) stellt die Nukleotid (SEQ ID NO: 27)- und Aminosäure (SEQ ID NO: 28)-Sequenz eines Fragments des menschlichen Kollagens vom Typ I (α_1)-Gens mit optimierter E. coli-Codonverwendung dar, welches vom Plasmid pBSN1-2 codiert wird.

[0072] [Fig. 45](#) stellt eine Plasmidkarte von pHuCol^{Ec} dar, welches menschliches Kollagen Typ I (α_1) mit optimierter E. coli-Codonverwendung enthält.

[0073] [Fig. 46](#) stellt eine Plasmidkarte von pTrcN1-2 dar, welches ein 234 Nukleotide großes Fragment des menschlichen Kollagens vom Typ I (α_1) mit optimierter E. coli-Codonverwendung enthält.

[0074] [Fig. 47](#) stellt eine Plasmidkarte von pN1-3 dar, welches ein Fraktion aus 360 Nukleotiden des menschlichen Kollagens vom Typ I (α_1) mit optimierter E. coli-Codonverwendung enthält.

[0075] [Fig. 48](#) stellt eine Plasmidkarte von pD4 dar, welches ein 3'-Fragment aus 657 Nukleotiden des menschlichen Kollagens vom Typ I (α_1) mit optimierter E. coli-Codonverwendung enthält.

[0076] [Fig. 49A–Fig. 49E](#) stellen die Nukleotid (SEQ ID NO: 29)- und Aminosäure (SEQ ID NO: 30)-Sequenz eines helikalen Bereichs des menschlichen Typ I (α_2)-Kollagens plus 11 Amino-terminaler extrahelikaler Aminosäuren und 12 Carboxy-terminaler extrahelikaler Aminosäuren enthält.

[0077] [Fig. 50A](#) bis [Fig. 50E](#) stellen den Nukleotid (SEQ ID NO: 31) und Aminosäure (SEQ ID NO: 32) von HuCol(α_2)^{Ec} dar, der helikalen Region des menschlichen Typ I (α_2)-Kollagens plus 11 Amino-terminaler extrahelikaler Aminosäuren und 12 Carboxy-terminaler extrahelikaler Aminosäuren mit der Codonverwendung, die für E. coli optimiert wurde.

[0078] [Fig. 51](#) stellt die Sequenz und Restriktionskarten synthetischer Oligos, die für die Rekonstruktion der ersten 240 Basenpaare des menschlichen Typ I (α_2)-Kollagen-Gens verwendet wurden, unter optimierter E. coli-Codonverwendung dar. Die synthetischen Oligos werden N1-1 (α_2) (SEQ ID NO: 33), N1-2 (α_2) (SEQ ID NO: 34), N1-3 (α_2) (SEQ ID NO: 35) und N1-4 (α_2) (SEQ ID NO: 36) bezeichnet.

[0079] [Fig. 52](#) stellt eine Plasmidkarte von pBSN1-1(α_2) dar, welches ein 117 Basenpaare-Fragment des menschlichen Kollagen-Typ I (α_2) mit optimierter E. coli-Codonverwendung enthält.

[0080] [Fig. 53](#) stellt eine Plasmidkarte von pBSN1-2(α_2) dar, welches ein 240 Basenpaare langes Fragment des menschlichen Kollagens vom Typ I (α_2) mit optimierter E. coli-Codonverwendung enthält.

[0081] [Fig. 54](#) stellt die Nukleotid (SEQ ID NO: 37)- und Aminosäure (SEQ ID NO: 38)-Sequenz eines Fragments des menschlichen Kollagens vom Typ I (α_2)-Gens mit optimierter E. coli-Verwendung dar, welches vom Plasmid pBSN1-2(α_2) codiert wird.

- [0082] [Fig. 55](#) stellt eine Plasmidkarte vom pHuCol(α_2)^{Ec} dar, welches das gesamte Gen des menschlichen Kollagens vom Typ I (α_2) mit optimierter E. coli-Codonverwendung enthält.
- [0083] [Fig. 56](#) stellt eine Plasmidkarte von pN1-2(α_2) dar, welches ein 240 Basenpaare langes Fragment des menschlichen Kollagens vom Typ I (α_2) mit optimierter E. coli-Codonverwendung enthält.
- [0084] [Fig. 57](#) stellt ein Gel dar, welches die Expression von GST und TGF- β_1 unter angegebenen Bedingungen widerspiegelt.
- [0085] [Fig. 58](#) stellt ein Gel dar, welches die Expression von MBP, FN-BMP-2A, FN-TGF- β_1 und FN unter angegebenen Bedingungen widerspiegelt.
- [0086] [Fig. 59](#) stellt ein Gel dar, das die Expression von GST-Coll unter angegebenen Bedingungen widerspiegelt.
- [0087] [Fig. 60](#) stellt eine Plasmidkarte von pGST-CM4 dar, welches das Gen für Glutathion-S-Transferase enthält, das mit dem Gen für das Kollagen-Mimetikum 4 fusioniert ist.
- [0088] [Fig. 61](#) stellt die Nukleotid (SEQ ID NO: 39)- und Aminosäure (SEQ ID NO: 40)-SEQ ID NO: des Kollagen-Mimetikums 4 dar.
- [0089] [Fig. 62A](#) stellt ein Chromatogramm der Eluierung von Hydroxyprolin-haltigem Kollagen-Mimetikum 4 von einer Poros RP2-Säule dar. Der Pfeil gibt den Peak an, der das Hydroxyprolin-haltige Kollagen-Mimetikum 4 enthält.
- [0090] [Fig. 62B](#) stellt ein Chromatogramm der Eluierung von Prolin-haltigem Kollagen-Mimetikum 4 von einer Poros RP2-Säule dar. Der Pfeil gibt den Peak an, der das Prolin-haltige Mimetikum 4 enthält.
- [0091] [Fig. 63A](#) gibt ein Chromatogramm einer Prolinaminosäure-Standards (250 pmol) wider.
- [0092] [Fig. 63B](#) gibt ein Chromatogramm eines Hydroxyprolinaminosäure-Standards (250 pmol) wider.
- [0093] [Fig. 63C](#) gibt ein Aminosäure-Analysechromatogramm des Prolin-haltigen Kollagen-Mimetikums 4 wider.
- [0094] [Fig. 63D](#) gibt das Aminosäure-Analysechromatogramm der Hydrolyse des Hydroxyprolin-haltigen Kollagen-Mimetikums 4 wider.
- [0095] [Fig. 64](#) ist ein Graph bei OD600 gegenüber der Zeit für Kulturen von E. coli JM109 (F⁻), die bis zum Plateau gezüchtet wurden und dann mit zahlreichen Aminosäuren ergänzt wurden.
- [0096] [Fig. 65](#) stellt eine Plasmidkarte von pcEc- α_1 dar, welche das Gen für HuCol(α_1)^{Ec} enthält.
- [0097] [Fig. 66](#) stellt eine Plasmidkarte von pcEc- α_2 dar, welche das Gen für HuCol(α_2)^{Ec} enthält.
- [0098] [Fig. 67](#) gibt eine Plasmidkarte von pD4- α_1 wider, welche das Gen für ein C-terminales Fragment aus 219 Aminosäuren des menschlichen Kollagens vom Typ I (α_1) mit optimierter E. coli-Codonverwendung wider, das mit dem Gen der Glutathion-S-Transferase fusioniert ist.
- [0099] [Fig. 68](#) gibt eine Plasmid-Karte von pD4- α_2 wider, welche das Gen für ein C-terminales Fragment aus 207 Aminosäuren des menschlichen Kollagens Typ I (α_2) mit optimierter E. coli-Codonverwendung wider, welche mit dem Gen von Glutathion-S-Transfersase fusioniert ist.
- [0100] [Fig. 69](#) gibt die vorhergesagte Aminosäuresequenz der DNA-Sequenz der ersten 13 Aminosäuren des Proteins D4- α_1 (SEQ ID NO: 41) und der Aminosäuresequenz wider, die experimentell bestimmt wurde (SEQ ID NO: 42).
- [0101] [Fig. 70](#) gibt das Massenspektrum von Hydroxyprolin-haltigem D4- α_1 wider.
- [0102] [Fig. 71](#) gibt die Nukleotidsequenz eines 657 Nukleotide langen Fragments vom menschlichen Kolla-

gen von Typ I ($\alpha 1$)3' mit optimierter E. coli-Codonverwendung wider, welche als D4 bezeichnet wird (SEQ ID NO: 43).

[0103] [Fig. 72](#) gibt die Aminosäuresequenz eines C-terminalen 219 Aminosäuren langen Fragments des menschlichen Kollagens vom Typ I ($\alpha 1$) wider, welches als D4 bezeichnet wird (SEQ ID NO: 44).

[0104] [Fig. 73](#) ist eine Plasmidkarte, die pGEX-4T.1 darstellt, welches das Gen für Glutathion-S-Transferase enthält.

[0105] [Fig. 74](#) ist eine Plasmidkarte, die pTrc-TGF darstellt, welches das Gen für das reife menschliche TGF- $\beta 1$ -Gen Polypeptid enthält.

[0106] [Fig. 75](#) ist eine Plasmidkarte, die pTrc-Fn darstellt, welches das Gen für ein 70 kDa-Fragment vom menschlichen Fibronectin enthält.

[0107] [Fig. 76](#) ist eine Plasmidkarte, die pTrc-Fn-TGF darstellt, welches das Gen für ein Fusionsprotein aus einem 70 kDa-Fragment des menschlichen Fibronectins und des reifen menschlichen TGF- $\beta 1$ Polypeptid enthält.

[0108] [Fig. 77](#) ist eine Plasmidkarte, die pTrc-Fn-BMP darstellt, welches das Gen eines Fusionsproteins eines 70 kDa-Fragments des menschlichen Fibronectins und des menschlichen Knochenmorphogenese-Proteins 2A enthält.

[0109] [Fig. 78](#) ist eine Plasmidkarte, die pGEX-HuCol1^{Ec} darstellt, welches das Gen für ein Fusionsprotein zwischen Glutathion-S-Transferase und Typ I ($\alpha 1$) menschlichen Kollagen mit optimierter E. coli-Codonverwendung enthält.

[0110] [Fig. 79](#) gibt die Nukleotidsequenz eines 3'-Fragments aus 627 Nukleotiden des menschlichen Kollagens vom Typ I ($\alpha 2$)3'-Fragments mit optimierter E. coli-Codonverwendung (SEQ ID NO: 45) wider.

[0111] [Fig. 80](#) gibt die Aminosäuresequenz eines C-terminalen 209 Aminosäure langen Fragments des menschlichen Kollagens vom Typ I ($\alpha 2$) (SEQ ID NO: 46) wider.

[0112] [Fig. 81](#) gibt die Sequenz synthetischer Oligos wider, die verwendet wurden, um die ersten 282 Basenpaare des Gens für die Carboxy-terminalen 219 Aminosäuren vom menschlichen Typ I ($\alpha 1$)-Kollagen mit optimierter E. coli-Codonverwendung zu rekonstruieren, welche als N4-1 (SEQ ID NO: 47), N4-2 (SEQ ID NO: 48), N4-3 (SEQ ID NO: 49) und N4-4 (SEQ ID NO: 50) bezeichnet werden.

Detaillierte Beschreibung der bevorzugten Ausführungsformen

[0113] Prokaryontische Zellen und eukaryontische Zellen können unerwarteterweise dazu gebracht werden trans-4-Hydroxyprolin in Proteine zu assimilieren und einzubauen, entgegen von sowohl Papas et al. als auch Deming et al., oben. Eine derartige Assimilierung und der Einbau ist besonders nützlich, wenn die Struktur und Funktion eines Polypeptids von einer posttranslationalen Hydroxylierung von Prolin abhängt, welche durch das native Protein-Produktionssystem eines rekombinanten Wirts nicht bereitgestellt wird. So können prokaryontische Bakterien, wie beispielsweise E. coli und eukaryontische Zellen wie beispielsweise Saccharomyces cerevisiae, Saccharomyces carlsbergensis und Schizosaccharomyces pombe, die für gewöhnlich nicht Prolin hydroxylieren und zusätzliche Eukaryonten wie beispielsweise Insektenzellen, einschließlich Lepidoptera-Zelllinien einschließlich Spodoptera frugiperda, Trichoplasia ni, Heliothis virescens, Bombyx mori, die mit einem Baculovirus infiziert sind; CHO-Zellen, COS-Zellen und NIH 3T3-Zellen, die es nicht schaffen, einige Polypeptide adäquat herzustellen, deren Strukturen und Funktion von einer derartigen Hydroxylierung abhängt, dazu gebracht werden, Polypeptide mit hydroxylierten Prolinen herzustellen. Der Einbau schließt die Addition von trans-4-Hydroxyprolin in einem Polypeptid ein, beispielsweise zunächst durch das Ändern einer Aminosäure in Prolin, wodurch eine neue Prolin-Position erzeugt wird, die wiederum mit trans-4-Hydroxyprolin substituiert werden kann, oder das Substituieren eines natürlich auftretenden Prolins in einem Polypeptid durch trans-4-Hydroxyprolin.

[0114] Das Verfahren zur Herstellung rekombinanter Polypeptide in Massen produzierenden Organismen ist gut bekannt. Replizierbare Expressionsvektoren, wie beispielsweise Plasmide, Viren, Cosmide und künstliche Chromosomen werden für gewöhnlich verwendet, um Gene, die gewünschte Proteine codieren, von einem

Wirt in einen anderen zu überführen. Es wird in Erwägung gezogen, daß jedes bekannte Verfahren zur Klonierung eines Gens, das Ligieren des Gens in einen Expressionsvektor und das Transformieren einer Wirtszelle mit einem derartigen Expressionsvektor nachfolgend in der vorliegenden Offenbarung verwendet werden kann.

[0115] Der Einbau von trans-4-Hydroxyprolin in Polypeptide, die von trans-4-Hydroxyprolin bezüglich ihrer chemischen und physikalischen Eigenschaften abhängen, ist nicht nur nützlich bei Produktionssystemen, die nicht die geeigneten Systeme zur Umwandlung von Prolin in trans-4-Hydroxyprolin haben, sondern auch nützlich zum Untersuchen der Struktur und Funktion von Polypeptiden, die normalerweise kein trans-4-Hydroxyprolin enthalten. Es wird in Erwägung gezogen, daß die folgenden Aminosäure-Analoga ebenfalls in Übereinstimmung der vorliegenden Offenbarung eingebaut werden können: trans-4-Hydroxyprolin, 3-Hydroxyprolin und Kombinationen daraus (nachfolgend bezeichnet als "Aminosäure-Analoga"). Die Verwendung von Prokaryonten und Eukaryonten ist wünschenswert, da sie eine relativ günstige Massenproduktion ermöglichen. Es wird in Erwägung gezogen, daß die Aminosäure-Analoga in jedes gewünschte Polypeptid eingebaut werden können. In einer bevorzugten Ausführungsform werden die prokaryontischen Zellen und eukaryontischen Zellen hinsichtlich Prolin hungern gelassen, indem die Menge an Prolin im Wachstumsmedium vor der Zugabe eines Aminosäure-Analogons dazu abgesenkt oder es entfernt wird.

[0116] Expressionsvektoren, die das Gen für das Maltose-bindende Protein (MBP) enthalten, z. B. siehe [Fig. 1](#), welche das Plasmid pMAL-c2 darstellt, welches kommerziell von New England Bio-Labs erhältlich ist, werden in Prokaryonten wie beispielsweise E. coli Prolin-Auxotrophe oder Eukaryonten wie beispielsweise S. cerevisiae-Auxotrophe transformiert, die vom extern zugeführten Prolin für die Proteinsynthese und den Anabolismus abhängen. Andere bevorzugte Expressionsvektoren zur Verwendung in Prokaryonten sind die kommerziell erhältlichen Plasmide, wie pKK-223 (Pharmacia), pTRC (Invitrogen), pGEX (Pharmacia), pET (Novagen) und pQE (Quiagen). Es ist klar, daß jeder geeignete Expressionsvektor von Fachleuten auf dem Gebiet verwendet werden kann.

[0117] Die Substituierung der Aminosäure-Analoga für Prolin in der Proteinsynthese tritt auf, da Prolyl-tRNA-Synthetase hinreichend promiskuitiv ist, und eine Mißacylierung von Prolin-tRNA mit irgendeinem der Aminosäure-Analoga ermöglicht. Eine ausreichende Menge, d. h. typischerweise von 0,001 M bis ungefähr 1,0 M reichen, aber mehr bevorzugt von ungefähr 0,005 M bis ungefähr 0,5 M der Aminosäure-Analoga wird zu dem Wachstumsmedium der transformierten Zelle hinzugegeben, um mit dem Prolin bei der zellulären Aufnahme zu wetteifern. Nach ausreichender Dauer, im allgemeinen von ungefähr 30 Minuten bis ungefähr 24 Stunden oder mehr, sind die Aminosäure-Analoga durch die Zelle assimiliert und in die Proteinsynthesewege eingebaut worden. Wie aus [Fig. 2](#) und [Fig. 2A](#) erkannt werden kann, steigt die intrazelluläre Konzentration von trans-4-Hydroxyprolin durch den Anstieg der Konzentration von Natriumchlorid im Wachstumsmedium an. In einer bevorzugten Ausführungsform werden die prokaryontischen Zellen und/oder eukaryontischen Zellen hinsichtlich Prolin hungern lassen, indem die Menge an Prolin im Wachstumsmedium vor der Zugabe eines Aminosäure-Analogons dazu gesenkt oder es eliminiert wird.

[0118] Expressionsvektoren, die das Gen für menschliches Kollagen vom Typ I ($\alpha 1$) enthalten (DNA-Sequenz, die in den [Fig. 3](#) und [Fig. 3A](#) dargestellt ist; Plasmidkarte, die in [Fig. 4](#) dargestellt ist), werden in prokaryontische oder eukaryontische Prolin-Auxotrophe transformiert, die vom extern zugeführten Prolin für die Proteinsynthese und den Anabolismus abhängen. Wie oben erwähnt, tritt die Substituierung der Aminosäure-Analoga auf, da Prolyl-tRNA-Synthetase hinreichend promiskuitiv ist, um eine Mißacylierung der Prolin-tRNA mit den Aminosäure-Analoga zu ermöglichen. Die Menge der Aminosäure-Analoga im Medium, die oben angegeben wurde, ist ebenfalls anwendbar.

[0119] Expressionsvektoren, die DNA enthalten, welche Fragmente des menschlichen Kollagens vom Typ I ($\alpha 1$) codieren (z. B. die DNA-Sequenz, die in [Fig. 5](#) dargestellt ist, und die Plasmidkarte, die in [Fig. 6](#) dargestellt ist), werden in prokaryontische oder eukaryontische Auxotrophe, wie oben erwähnt, transformiert. Gleichermaßen können Expressionsvektoren, die DNA enthalten, welche das Kollagen-artige Polypeptid codieren (z. B. die DNA-Sequenz, die in [Fig. 7](#) dargestellt ist, die Aminosäuresequenz, die in [Fig. 8](#) dargestellt ist sowie die Plasmidkarte, die in [Fig. 9](#) dargestellt ist) verwendet werden, um prokaryontische oder eukaryontische Auxotrophe, wie oben angegeben, zu transformieren. Kollagen-artige Peptide sind solche, die mindestens eine Teilhomologie mit Kollagen aufweisen und ähnliche chemische und physikalische Eigenschaften wie Kollagen aufweisen. So bestehen Kollagen-artige Peptide beispielsweise aus sich wiederholenden Anordnungen von Gly-X-Y-Triplets, in denen ungefähr 35% der X- und Y-Positionen durch Prolin und 4-Hydroxyprolin besetzt sind. Kollagen-artige Peptide werden hier austauschbar als Kollagen-artige Proteine, Kollagen-artige Polypeptide, Kollagen-mimetische Polypeptid und Kollagen-Mimetika bezeichnet. Bevorzugte Kollagen-Fragmente

und Kollagen-artige Peptide in Übereinstimmung hiermit, sind in der Lage, sich in einer extrazellulären Matrix zusammenzufinden. Sowohl in Kollagen-freien Fragmenten und Kollagen-artigen Peptiden, die oben beschrieben wurden, tritt eine Substituierung durch Aminosäure-Analoga auf, da Prolyl-tRNA-Synthetase hinreichend promiskuitiv ist, um eine Mißacylierung der Prolin-tRNA durch ein oder mehrere der Aminosäure-Analoga zu ermöglichen. Die Menge der Aminosäure-Analoga, die oben angegeben wurde, ist ebenfalls anwendbar.

[0120] Es wird in Erwägung gezogen, daß jedes Polypeptid, das eine extrazelluläre Matrix-Proteindomäne hat, wie beispielsweise Kollagen, Kollagen-Fragmente oder eine Kollagen-artige Peptiddomäne dazu gebracht werden kann, Aminosäure-Analoga gemäß der Offenbarung hierin einzubauen. Diese Polypeptide schließen Kollagen, ein Kollagen-Fragment oder eine Kollagen-artige Peptiddomäne ein sowie eine Domäne mit einem Bereich, der ein oder mehrere physiologisch wirksame Mittel einbaut, wie beispielsweise Glycoproteine, Proteine, Peptide und Proteoglycane. So wie es hier verwendet wird, üben physiologisch aktive Mittel Kontrolle aus oder modifizieren vorhandene physiologische Eigenschaften lebender Dinge. Physiologisch aktive Mittel schließen Hormone, Wachstumsfaktoren, Enzyme, Liganden und Rezeptoren ein. Viele aktive Domänen physiologisch aktiver Mittel sind definiert und isoliert worden. Es wird beschrieben, daß Polypeptide mit einem Kollagen, einem Kollagen-Fragment oder einer Kollagen-artigen Peptiddomäne ebenfalls eine Domäne haben können, die ein oder mehrere physiologisch aktive Domänen einschließt, welche aktive Fragmente dieser physiologisch wirksamen Mittel sind. So wie es hier verwendet wird, bedeutet physiologisch wirksames Mittel vollständige Peptide, Polypeptide, Proteine, Glycoproteine, Proteoglycane und aktive Fragmente irgendeines davon. So können chimäre Proteine dazu gebracht werden, durch Transformierung eines prokaryontischen Prolin-Auxotrophen oder eines eukaryontischen Prolin-Auxotrophen mit einem geeigneten Expressionsvektor, und das Kontaktieren des transformierten Auxotrophen mit einem Wachstumsmedium, das mindestens eins der Aminosäure-Analoga enthält Aminosäure-Analoga einzubauen. Beispielsweise kann ein chimäres Kollagen/Knochenmorphogenese-Protein(BMP)-Konstrukt oder verschiedene chimäre Kollagen/Wachstumsfaktor-Konstrukte erfindungsgemäß nützlich sein. Derartige Wachstumsfaktoren sind gut bekannt und schließen Insulin-artigen Wachstumsfaktor, transformierenden Wachstumsfaktor, aus Plättchen-stammender Wachstumsfaktor und ähnliche ein. [Fig. 10](#) stellt die DNA von BMP dar, welche mit dem 3'-Terminus der DNA, welche Kollagen codiert, DNA, die ein Kollagen-Fragment codiert, oder DNA, die ein Kollagen-artiges Peptid codiert fusioniert werden kann. [Fig. 11](#) stellt eine Karte des Plasmids pCBC dar, welche ein Kollagen/BMP-Konstrukt enthält. In einer bevorzugten Ausführungsform fügen sich Proteine mit einem Kollagen, Kollagen-Fragment oder einer Kollagen-artigen Peptiddomäne zusammen oder aggregieren, um eine extrazelluläre Matrix zu bilden, die als chirurgisches Implantat verwendet werden kann. Die Eigenschaft der Selbstaggregation, die hier verwendet wird, schließt die Fähigkeit ein, mit den gleichen oder ähnlichen Molekülen ein Aggregat zu bilden, oder ein Aggregat mit anderen Molekülen zu bilden, die die Eigenschaft der Aggregation teilen, um beispielsweise eine Doppel- oder Tripelhelix zu bilden. Ein Beispiel einer derartigen Aggregation ist die Struktur aneinandergelagerter Kollagen-Matrizes.

[0121] So können chimäre Polypeptide, die hier ebenfalls als chimäre Proteine bezeichnet werden können, eine integrierte Zusammensetzung aus einer therapeutisch wirksamen Domäne eines physiologisch wirksamen Mittels und einem oder mehreren EMP-Resten bereitstellen. Die EMP-Domäne stellt ein integrales Vehikel zur Verabreichung des therapeutisch wirksamen Restes an einen Zielort bereit. Die zwei Domänen sind durch eine oder mehrere Peptidbindungen, die in einer Linkerregion enthalten sind, kovalent verbunden. So wie es hier verwendet wird, bedeutet integriert oder integral Eigenschaften, die aus der kovalenten Verbindung einer oder mehrerer Domänen des chimären Proteins hervorgehen. Die therapeutisch wirksamen Reste, die hier offenbart sind, bestehen typischerweise aus Aminosäuren, die verbunden sind, um Peptide, Polypeptide, Proteine, Glycoproteine oder Proteoglycane zu bilden. So wie es hier verwendet wird, umfaßt "Peptid" Polypeptide und Proteine.

[0122] Die inhärenten Eigenschaften von EMPs sind ideal zur Verwendung als Vehikel für den therapeutischen Rest. Eine derartige Eigenschaft ist die Fähigkeit der EMPs, das Selbstaggregat zu bilden. Beispiele für geeignete EMPs sind Kollagen, Elastin, Fibronectin, Fibrinogen und Fibrin. Fibrilläre Kollagene (Typ I, II und III) lagern sich in geordneten Polymeren zusammen und aggregieren häufiger zu größeren Bündeln. Typ IV-Kollagen setzt sich in schichtartigen Netzwerken zusammen. Elastin-Moleküle bilden Filamente und Schichten, in denen die Elastin-Moleküle miteinander stark vernetzt sind, um eine gute Elastizität und eine hohe Reißstärke bereitzustellen. Die vernetzte, zufallsmäßig verknäuelte Struktur des Fasernetzwerks ermöglicht es, wie ein Gummiband gedehnt zu werden und sich wieder zusammenzuziehen. Fibronectin ist ein großes Fibrillen-bildendes Glycoprotein, welches in einer seiner Formen aus stark unlöslichen Fibrillen besteht, die miteinander durch Disulfid-Bindungen verbunden sind. Fibrin ist ein unlösliches Protein, das aus Fibrinogen durch proteolytische Einwirkung von Thrombin während der normalen Verklumpung von Blut gebildet wird.

[0123] Die molekulare und makromolekulare Morphologie der oben erwähnten EMPs definiert Netzwerke oder Matrices, um ein Substrat oder ein Gerüst in integraler kovalenter Verbindung mit dem therapeutisch wirksamen Rest bereitzustellen. Die Netzwerke oder Matrices, die durch die EMP-Domäne gebildet werden, stellen eine Umgebung bereit, die besonders gut für das Einwachsen autologer Zellen geeignet ist, die am Wachstum, der Reparatur und dem Austausch von vorhandenem Gewebe beteiligt sind. Die integralen therapeutisch wirksamen Reste, die kovalent in die Netzwerke oder Matrices gebunden sind, stellen eine maximale Exposition der wirksamen Mittel an ihrem Ziel bereit, um eine gewünschte Reaktion auszulösen.

[0124] Implantate, die aus oder von den vorliegenden chimären Proteinen gebildet werden, stellen die Wirkung der verzögerten Freisetzung in oder am gewünschten Ort oder einer Zielstelle bereit. Da sie an eine EMP-Domäne gebunden sind, ist die therapeutisch wirksame Domäne des vorliegenden chimären Proteins nicht frei um getrennt zu diffundieren oder in sonstiger Weise von dem Träger, die es trägt, fort transportiert zu werden, ausgenommen von der Abspaltung der Peptidbindungen. Infolgedessen stellen die hierin bereitgestellten chimären Proteine eine wirksame Verankerung für eine therapeutische Wirkung bereit, welche ermöglicht, daß die Aktivität an einem Zielort für einen verlängerten Zeitraum beschränkt ist. Da die Zufuhr des therapeutisch wirksamen Mittels nicht so häufig wieder aufgefüllt werden muß, wenn es mit nicht-verzögerten Freisetzungsdosierungsformen verglichen wird, können geringere Mengen eines therapeutisch wirksamen Mittels im Verlaufe der Therapie verwendet werden. Infolgedessen sind einige Vorteile, die durch die vorliegenden chimären Proteine bereitgestellt werden, die Absenkung oder die Eliminierung lokaler und systemischer Nebenwirkungen, die geringere Potenzierung oder Reduzierung der therapeutischen Wirkung bei chronischer Verwendung, und die Minimierung der Medikamentenanhäufung in einem Körpergewebe bei chronischer Dosierung.

[0125] Die Verwendung der rekombinanten Technologie ermöglicht die Herstellung nicht-immuogener chimärer Proteine. Die DNA, die sowohl den therapeutisch wirksamen Rest und den EMP-Rest codiert, sollte vorzugsweise von der gleichen Art stammen, wie der Patient, der zu behandeln ist, um eine immunogene Reaktion zu vermeiden. Beispielsweise wird, wenn der Patient ein Mensch ist, der therapeutische wirksame Rest wie auch der EMP-Rest vorzugsweise aus menschlicher DNA stammen.

[0126] Chimäre osteogene/EMP Proteine stellen biodegradierbare und biokompatible Mittel zur Induzierung der Knochenbildung an einer gewünschten Stelle bereit. Wie oben erwähnt, ist in einer Ausführungsform ein BMP-Rest kovalent mit einem EMP verbunden, um ein chimäres Protein zu bilden. Der BMP-Rest induziert die Osteogenese, und der Proteinrest der extrazellulären Matrix stellt ein integrales Substrat oder Gerüst für den BMP-Rest und Zellen, die an der Rekonstruktion und dem Wachstum beteiligt sind, bereit. Zusammensetzungen, die das chimäre BMP/EMP-Protein enthalten, stellen die wirksame verzögerte Freisetzungsverabreichung des BMP-Restes an gewünschte Zielorte bereit. Das Verfahren zur Herstellung eines solchen osteogenen Mittels ist effizient, da die Notwendigkeit für zusätzliche zeitaufwendige Schritte wie die Reinigung von EMP und dann die Vermischung desselben mit dem gereinigten BMP ausgeschaltet ist. Ein zusätzlicher Vorteil des chimären BMP/EMP-Proteins stammt aus der Stabilität, die durch die kovalente Bindung zwischen BMP und dem EMP gebildet wird, d. h. der BMP-Teil ist nicht frei in der Lage separat vom EMP weg zu diffundieren, wodurch ein stabileres therapeutisches Mittel bereitgestellt wird.

[0127] Knochenmorphogenese-Proteine sind eine Klasse, die als BMP-1 bis BMP-9 identifiziert wurden. Ein bevorzugtes osteogenes Protein zur Verwendung in menschlichen Patienten ist menschliches BMP-2B. Ein chimäres BMP-2B/Kollagen IA-Protein ist in [Fig. 13](#) (SEQ ID NO: 6) dargestellt. Die Proteinsequenz, die in [Fig. 15](#) dargestellt ist (SEQ ID NO: 8) schließt eine helikale Kollagen-Domäne ein, die in den Aminosäuren 1 bis 1057 wiedergegeben ist und eine reife Form von BMP-2B in den Aminosäuren 1060 bis 1169. Die physikalischen Eigenschaften des chimären Proteins werden teilweise durch den EMP-Anteil bestimmt. Im Fall eines Kollagen-Restes, hat eine konzentrierte Lösung des chimären Proteins eine gelatineartige Konsistenz, die die leichte Handhabung durch den Mediziner ermöglicht. Der EMP-Rest wirkt als Sequestrierungsmittel, um die schnelle Desorption des BMP-Restes von der gewünschten Stelle zu verhindern, und um eine verzögerte Freisetzung der BMP-Aktivität bereitzustellen. Infolge dessen bleibt der BMP-Rest an der gewünschten Stelle und stellt einer verlängerten Freisetzung der BMP-Aktivität an der gewünschten Stelle für einen notwendigen Zeitraum bereit, um die Knochenbildung wirksam zu induzieren. Der EMP-Rest stellt ebenfalls eine Matrix bereit, welche ermöglicht, daß die autologen Zellen des Patienten, z. B. Chondrozyten und ähnliche, die normalerweise an der Osteogenese beteiligt sind, sich hierin zu sammeln, um ein autologes Netzwerk für neues Gewebewachstum zu bilden. Die gelatineartige Konsistenz des chimären Proteins stellt auch eine günstige und nützliche therapeutische Art der Immobilisierung von aktivem BMP auf einem geeigneten Vehikel oder Implantat zur Verabreichung des BMP-Restes an einer Stelle bereit, an der das Knochenwachstum gewünscht ist.

[0128] Der BMP-Rest und der EMP-Rest werden gegebenenfalls zusammen durch Linkersequenzen aus Aminosäuren verbunden. Beispiele für Linkersequenzen, die verwendet werden, liegen innerhalb der Sequenz, die in [Fig. 14A](#) bis [Fig. 14C](#) (SEQ ID NO: 7), [Fig. 14A](#) bis [Fig. 16C](#) (SEQ ID NO: 9), [Fig. 19A](#) bis [Fig. 19C](#) (SEQ ID NO: 12) und [Fig. 20A](#) bis [Fig. 20C](#) (SEQ ID NO: 13) wiedergegeben sind und werden im genaueren Detail unten beschrieben. Linkersequenzen können auf Grundlage bestimmter Eigenschaften ausgewählt werden, die sie dem chimären Protein verleihen. Beispielsweise werden Aminosäuresequenzen wie beispielsweise Ile-Glu-Gly-Arg und Leu-Val-Pro-Arg durch Faktor XA und Thrombin-Enzyme gespalten. Der Einbau von Sequenzen, die von proteolytischen Enzymen gespalten werden, in chimäre Proteine stellt hier die Spaltung an der Linkerstelle nach Exposition gegenüber einem geeigneten Enzym, und die Trennung der zwei Domänen in getrennte Einheiten bereit. Es wird in Erwägung gezogen, daß verschiedene Linkersequenzen in irgendeines der chimären Proteine eingebaut werden können.

[0129] Ein chimäres DNA-Konstrukt schließt ein Gen ein, das ein Osteogenese-Protein oder ein Fragment davon einschließt, das mit einem Gen verbunden ist, das ein EMP oder ein Fragment davon codiert. Die Gensequenz verschiedener BMPs sind bekannt, siehe z. B. US-Patent Nrn. 4,294,753, 4,761,471, 5,106,748, 5,187,076, 5,141,905, 5,108,922, 5,116,738 und 5,168,050. Das BMP-2B-Gen zur Verwendung hierin wird durch Ligieren von Oligonukleotiden, die ein BMP-Protein codieren, synthetisiert. Die Oligonukleotide, die BMP-2B codieren, werden unter Verwendung eines automatisierten DNA-Synthesegerätes (Beckman Oligo-1000) synthetisiert. Die Nukleotidsequenz, die das BMP codiert, wird zur Expression in *E. coli* maximiert. Dies wird durch die Verwendung der *E. coli*-Verwendungstabellen erreicht, um die Sequenz der Aminosäuren von BMP in Codons zu transportieren, die am häufigsten durch *E. coli* verwendet werden. Alternativ dazu kann native DNA, die BMP codiert, welche aus Säugern, einschließlich Menschen, isoliert wurde, gereinigt und verwendet werden.

[0130] Das BMP-Gen und die DNA-Sequenz, die ein extrazelluläres Matrixprotein codiert, wird durch Standard-Gentechnikverfahren kloniert, die in Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor 1989 beschrieben sind.

[0131] Die DNA-Sequenz, die der helikalen und Telepeptid-Region von Kollagen I ($\alpha 1$) entspricht wird aus einer menschlichen Fibroblasten-Zelllinie kloniert. Zwei Ansätze der Polymerasekettenreaktion werden unter Verwendung von cDNA durchgeführt, die durch Standardverfahren aus AG02261A-Zellen präpariert wurde. Das erste PCR-Primerpaar schließt einen 5'-Primer ein, der eine XmnI-Linkersequenz trägt und einen 3'-Primer, der die BsmI-Stelle an der Nukleotidposition 1722 trägt. Das daraus hervorgehende PCR-Produkt besteht aus einer Sequenz von der Position 1 bis 1722. Das zweite Primerpaar schließt die BsmI-Schnittstelle bei 1722 ein, und eine Linkersequenz am 3'-Ende welche eine BglII-Schnittstelle trägt. Das erhaltene PCR-Produkt besteht aus einer Sequenz von Position 1722 bis 3196. Die vollständige Sequenz wird durch Standard-Klonierungstechniken zusammengebaut. Die zwei PCR-Produkte werden an der BsmI-Schnittstelle ligiert, und der kombinierte Klon wird in irgendeinen Vektor mit einer mit XmnI-BglII-Schnittstellen inseriert, wie beispielsweise den pMal-c2-Vektor.

[0132] Um das BMP-2B-Gen zu klonieren, wird die gesamte zelluläre RNA aus menschlichen Osteosarkom-Zellen (U-2OS) durch das von Robert E. Farrel Jr. (*Academic Press*, CA, 1993, S. 68–69) beschriebene Verfahren isoliert. Die Integrität der RNA wird mittels spektrophotometrischer Analyse und Elektrophorese mit Agarosegelen überprüft. Typische Ausbeuten an Gesamt-RNA sind 50 µg aus 100 mm konfluenten Gewebekulturschälchen. Die RNA wird verwendet, um cDNA durch reverse Transkription unter Verwendung des Superscript-Präamplifizierungssystem von Gibco BRL zu erzeugen. Die cDNA wird als Matrize für die PCR-Amplifizierung unter Verwendung von stromaufwärts und stromabwärts liegenden Primern verwendet, die spezifisch für BMP-2B (GenBank HUMBMP2B, Zugangs-Nr. #M22490) sind. Das daraus hervorgehende PCR-Produkt besteht aus der BMP-2B-Sequenz von Position 1289 bis 1619. Das PCR-Produkt wird mittels Elektrophorese durch Agarosegele gelöst, mit gene clean (BIO 101) gereinigt und in den pMal-c2-Vektor (New England Biolabs) ligiert. Die Domäne der menschlichen Kollagen I (21)-Kette wird in ähnlicher Weise kloniert. Jedoch wird die gesamte zelluläre RNA aus einer menschlichen Fibroblasten-Zelllinie isoliert (AG02261A menschliche Hautfibroblasten).

[0133] Ein chimäres BMP/EMP-DNA-Konstrukt wird durch Ligieren des synthetischen BMP-Gens an eine DNA-Sequenz, die ein EMP wie beispielsweise Kollagen, Fibrinogen, Fibrin, Fibronectin, Elastin oder Laminin codiert, gewonnen. Jedoch sind die chimären Polypeptide hier nicht auf diese bestimmten Protein limitiert. Die [Fig. 14A](#) bis [Fig. 14C](#) (se17) stellen ein DNA-Konstrukt dar, welches ein BMP-2B/Kollagen I ($\alpha 1$)-chimäres Protein darstellt. Die codierende Sequenz eines EMP kann stromaufwärts und/oder stromabwärts und im Leseraster mit einer codierenden Sequenz für das BMP ligiert werden. Die DNA, welche ein EMP codiert, kann ein

Teil des Gens des gesamten EMP-Gens sein. Außerdem können zwei verschiedene EMPs stromaufwärts und stromabwärts von BMP ligiert werden.

[0134] Das chimäre BMP-2/Kollagen I ($\alpha 1$)-Protein, das in den [Fig. 14A](#) bis [Fig. 14C](#) dargestellt ist, schließt eine XmnI-Linkersequenz an den Basenpaaren (bp) 1–19, eine Kollagen-Domäne (bp 20–3190), eine BglII/BamHI-Linkersequenz (bp 3191–3196), die reife Form von BMP2b (bp 3197–3529) und eine HindIII-Linkersequenz (bp 3530–3535) ein.

[0135] Jede Kombination aus Wachstumsfaktor-Matrixproteinsequenzen wird beschrieben, einschließlich sich wiederholender Einheiten, oder multipler Anordnungen jedes Segments in irgendeiner Reihenfolge.

[0136] Der Einbau von Fragmenten von sowohl Matrix- und Wachstumsfaktor-Proteinen wird ebenfalls beschrieben. Beispielsweise kann im Fall von Kollagen nur die helikale Domäne eingeschlossen sein. Andere Matrix-Proteine haben definierte Domänen, wie beispielsweise Laminin, welches EGF-artige Domänen hat. In diesen Fällen können spezifische Funktionalitäten ausgewählt werden, um gewünschte Wirkungen zu erzielen. Außerdem kann es nützlich sein, Domänen verschiedener Matrixproteine zu kombinieren, beispielsweise die helikale Region von Kollagen und die Zellanhaftungsregionen von Fibronectin. Im Falle von Wachstumsfaktoren ist bei spezifischen Segmenten gezeigt worden, daß sie durch posttranslationale Prozessierung vom reifen Protein entfernt werden. Chimäre Proteine können entworfen werden, um nur den reifen biologisch wirksamen Bereich einzuschließen. Beispielsweise werden im Fall von BMP-2B nur die letzten 110 Aminosäuren in dem aktiven Protein gefunden.

[0137] In einer anderen Ausführungsform wird ein Rest des transformierenden Wachstumsfaktors (TGF) kovalent mit einem EMP verbunden, um ein chimäres Protein zu bilden. Der TGF-Rest erhöht die Wirksamkeit der normalen Reparaturantwort von Weichteilgeweben im Körper, und induziert ebenfalls die Osteogenese. Infolgedessen können chimäre TGF/EMP-Proteine für irgendeine oder beide Funktionen verwendet werden. Eine der fundamentalen Eigenschaften der TGF- β s ist deren Fähigkeit, zahlreiche Aktivitäten anzuschalten, die zur Synthese neuen Bindegewebes führen. Siehe Piez und Sporn Hrsg., Transforming Growth Factor- β s Chemistry, Biology and Therapeutics, Annals of the New York Academy of Sciences, Bd. 593 (1990). TGF- β ist dafür bekannt, daß er in mindestens fünf verschiedenen Isoformen vorliegt. Die DNA-Sequenz für menschliches TGF- β_1 ist bekannt und ist kloniert worden. Siehe Derynck et al., Human Transforming Growth Factor-Beta cDNA Sequence and Expression in Tumour Cell Lines, Nature, Bd. 316, S. 701–705 (1985). TGF- β_2 ist aus Rinderknochen, menschlichen Glioblastom-Zellen und Plättchen des Schweins isoliert worden. TGF- β_3 ist ebenfalls kloniert worden. Siehe ten Dijke et al., Identification of a New Member of the Transforming Growth Factor- β Gene Family, Proc. Natl. Acad. Sci. (USA), Bd. 85, S. 4715–4719.

[0138] Ein chimäres TGF- β /EMP-Protein hat die bekannten Aktivitäten von TGF- β s und stellt die integrale Gerüstfunktion als Substrat des EMPs dar, das oben beschrieben wurde, um eine Zusammensetzung bereitzustellen, die ferner die verzögerte Freisetzung in einer fokalen Verabreichung an Zielorten ermöglicht.

[0139] Das TGF- β -Rest und der EMP-Rest sind gegebenenfalls durch Linkersequenz-Aminosäuren verbunden. Linkersequenzen können auf Grundlage ihrer besonderen Eigenschaften, die sie dem chimären Protein verleihen, ausgewählt werden. Beispielsweise werden Aminosäuresequenzen wie Ile-Glu-Gly-Arg und Leu-Val-Pro-Arg durch Faktor XA und Thrombin-Enzyme gespalten. Der Einbau von Sequenzen, die durch proteolytische Enzyme gespalten werden in das chimäre Protein stellt eine Spaltung an der Bindungsstelle nach Exposition gegenüber dem geeigneten Enzym und die Trennung der Domänen in getrennte Einheiten bereit. [Fig. 15](#) stellt eine Aminosäuresequenz für ein TGF- β_1 /Kollagen IA-chimäres Protein (SEQ ID NO: 8) dar. Die dargestellte Aminosäuresequenz schließt die Kollagen-Domäne (1–1057) und eine reife Form von TGF- β_1 (1060–1171) ein.

[0140] Ein chimäres DNA-Konstrukt schließt ein Gen ein, das TGF- β_1 codiert oder ein Fragment davon, oder ein Gen, das TGF- β_2 oder ein Fragment davon codiert, oder ein Gen, das TGF- β_3 oder ein Fragment davon codiert, welches mit einer DNA-Sequenz, welche ein EMP-Protein codiert, ligiert ist, wie beispielsweise Kollagen (I–IV), Fibrin, Fibrinogen, Fibronectin, Elasin oder Laminin. Ein bevorzugtes chimäres DNA-Konstrukt kombiniert DNA, die TGF- β_1 codiert, eine DNA-Linkersequenz, und eine DNA, die Kollagen IA codiert. Ein chimäres DNA-Konstrukt, das das TGF- β_1 -Gen und ein Kollagen I (α)-Gen codiert, ist in den [Fig. 16A](#) bis [Fig. 16C](#) (SEQ ID NO: 9) gezeigt. Das dargestellte Konstrukt schließt eine XmnI-Linkersequenz (bp 1–19) ein, DNA, die eine Kollagen-Domäne codiert (bp 20–3190), eine BglII-Linkersequenz (bp 3191–3196), eine DNA, die die reife Form von TGF- β_1 codiert (3197–3535) und eine XbaI-Linkersequenz (bp 3536–3541).

[0141] Die für EMP codierende Sequenz kann stromaufwärts und/oder stromabwärts und im Leseraster mit einer codierenden Sequenz für den TGF- β ligiert sein. Die DNA, die das extrazelluläre Matrixprotein codiert, kann einen Teil eines Fragments des EMP codieren oder das gesamte EMP codieren. Gleichmaßen kann die DNA, die das TGF- β codiert, eines oder mehr Fragmente davon oder das gesamte Gen sein. Außerdem können zwei oder mehr verschiedene TGF- β s oder zwei oder mehr verschiedene EMPs stromaufwärts oder stromabwärts abwechselnder Reste ligiert sein.

[0142] In einer weiteren Ausführungsform ist ein Dermatansulfaproteoglycan-Rest, ebenfalls als Decorin oder Proteoglycan II bekannt, kovalent mit einem EMP verbunden, um ein chimäres Protein zu bilden. Decorin ist dafür bekannt, daß es an Typ I-Kollagen bindet, und daher die Fibrillenbildung beeinflusst, und die Zellanhaftungs-fördernde Aktivität von Kollagen und Fibrinogen durch Bindung an derartige Moleküle in der Nähe der Zellbindungsstellen inhibiert. Chimäre Proteine, die einen Decorin-Rest enthalten, wirken so, daß sie die Vernarbung heilenden Gewebes induzieren. Die Primärstruktur des Kernproteins von Decorin ist von der klonierten cDNA abgeleitet worden. Siehe Krusius et al., Primary Structure of an Extracellular Matrix Proteoglycan Core Protein-Deduced from Clones cDNA, Natl. Acad. Scil (USA), Bd. 83, S. 7683–7687 (1986).

[0143] Ein chimäres Decorin/(EMP-Protein schließt die bekannten Aktivitäten von Decorin ein, und stellt ein integrales Gerüst oder ein Substrat des EMPs, das oben beschrieben wurde, bereit, um eine Zusammensetzung hervorzubringen, die eine verzögerte Freisetzung durch fokale Verabreichung an Zielorte ermöglicht. [Fig. 17A](#) bis [Fig. 17B](#) stellen ein chimäres Decorin/Kollagen 1A-Protein (SEQ ID NO: 10) dar, bei dem die Kollagen-Domäne die Aminosäuren 1–1057 einschließt und das reife Decorin-Protein die Aminosäuren 1060–1388 einschließt. [Fig. 18](#) stellt ein chimäres Decorin-Peptid/Kollagen 1A-Protein (SEQ ID NO: 11) dar, in dem die helikale Domäne die Aminosäuren 1–1057 einschließt, und das Decorin-Peptidfragment die Aminosäuren 1060–1107 einschließt. Das Decorin-Peptidfragment setzt sich aus P46 bis G93 der reifen Form des Decorins zusammen.

[0144] Außerdem wird ein chimäres DNA-Konstrukt bereitgestellt, das ein Gen einschließt, das Decorin oder ein oder mehrere Fragmente davon codiert, welches gegebenenfalls über einen DNA-Linker mit einer DNA-Sequenz verbunden ist, die ein EMP codiert, wie beispielsweise Kollagen (I–IV), Fibrin, Fibrinogen, Fibronectin, Elastin oder Laminin. Ein bevorzugtes chimäres DNA-Konstrukt, kombiniert DNA, die Decorin codiert, eine DNA-Linkersequenz und DNA, die Kollagen I ($\alpha 1$) codiert. Ein chimäres DNA-Konstrukt, das ein Decorin-Gen und ein Kollagen I ($\alpha 1$)-Gen enthält, ist in [Fig. 19A](#) bis [Fig. 19D](#) (SEQ ID NO: 12) gezeigt. Das dargestellte Konstrukt schließt eine XmnI-Linkersequenz (bp 1–19), DNA, die eine Kollagen-Domäne codiert (bp 20–3190), eine BglII-Linkersequenz (bp 3191–3196), DNA, die eine reife Form von Decorin codiert (bp 3197–4186) und eine PstI-Linkersequenz an. Ein chimäres DNA-Konstrukt, das ein Decorin-Peptidgen und ein Kollagen I ($\alpha 1$)-Gen enthält, ist in den [Fig. 20A](#) bis [Fig. 20C](#) (SEQ ID NO: 13) gezeigt. Das dargestellte Konstrukt schließt eine XmnI-Linkersequenz (bp 1–19), DNA, die eine Kollagen-Domäne codiert (bp 20–3190), eine BglII-Linkersequenz (bp 3191–3196), DNA, die ein Peptid-Fragment von Decorin codiert (bp 3197–3343) und eine PstI-Linkersequenz (bp 3344–3349) ein.

[0145] Die codierende Sequenz für ein EMP kann stromaufwärts und/oder stromabwärts und im Leseraster mit einer codierenden Sequenz für Decorin ligiert sein. Die DNA, welche EMP codiert, kann ein Teil oder ein Fragment von EMP codieren oder das gesamte EMP codieren. Gleichmaßen kann die DNA, die Decorin codiert, ein Fragment davon sein oder das gesamte Gen. Außerdem können zwei oder mehr verschiedene EMPs stromaufwärts und/oder stromabwärts der DNA, die den Decorin-Rest codiert, ligiert sein.

[0146] Jedes der oben beschriebenen chimären DNA-Konstrukte kann in einen geeigneten Klonierungsvektor eingeschlossen sein. [Fig. 21](#) stellt einen pMal-Klonierungsvektor dar, welcher eine Polylinker-Klonierungsstelle enthält. Beispiele derartiger Klonierungsvektoren sind die Plasmide pMal-p2 und pMal-c2 (kommerziell erhältlich von New England Biolabs). Das gewünschte Chimäre DNA-Konstrukt wird in eine Polylinkersequenz des Plasmids eingebaut, welche bestimmte nützliche Restriktionsendonuklease-Schnittstellen enthält, welche in [Fig. 22](#) dargestellt sind (SEQ ID NO: 14). Die pMalp2-Polylinkersequenz hat XmnI-, EcoRI-, BamHI-, HindII-, XbaI-, Sall- und PstI-Restriktionsendonuklease-Schnittstellen, welche in [Fig. 22](#) dargestellt sind. Die Polylinkersequenz wird mit einer geeigneten Restriktionsendonuklease verdaut und das chimäre Konstrukt wird in den Klonierungsvektor durch Ligieren derselben mit DNA-Sequenzen des Plasmids eingebaut. Das chimäre DNA-Konstrukt kann mit dem Plasmid durch Verdauen der Enden des DNA-Konstrukts und dem Plasmid mit dem gleichen Restriktionendonukleasen verbunden werden, um "klebrige Enden" mit 5'-Phosphat- und 3'-Hydroxyl-Gruppen zu erzeugen, die ermöglichen, daß das DNA-Konstrukt sich in dem Klonierungsvektor aneinander lagert. Lücken zwischen dem insertierten DNA-Konstrukt und dem Plasmid werden dann mit einer DNA-Ligase versiegelt. Andere Techniken zum Einbau des DNA-Konstruktes in Plasmid-DNA schließen die

Ligierung über glatte Enden, Poly(dA.dT)-Schwanztechniken und die Verwendung chemisch synthetisierter Linker ein. Ein alternatives Verfahren zum Einschleusen des chimären DNA-Konstruktes in einen Klonierungsvektor ist der Einbau der DNA, welcher das extrazelluläre Matrixprotein codiert, in einen Klonierungsvektor, der bereits ein Gen enthält, welches einen therapeutisch wirksamen Rest codiert.

[0147] Die Klonierungsstellen in der oben angegebenen Polylinker-Schnittstelle ermöglichen der cDNA für chimäres Kollagen I ($\alpha 1$)/BMP-2B Protein, welches in den [Fig. 14A–Fig. 14C](#) (SEQ ID NO: 7) dargestellt ist in die XmnI- und die HindIII-Schnittstellen inseriert zu werden. Die cDNA, welche das Kollagen I (21)(TGF- β_1)-Protein codiert, welches in den [Fig. 16A](#) bis [Fig. 16C](#) dargestellt ist (SEQ ID NO: 9) wird zwischen die XmnI- und die XbaI-Schnittstellen inseriert. cDNA, welche das Kollagen 1 ($\alpha 1$)/Decorin-Protein codiert, welches in den [Fig. 19A](#) bis [Fig. 19D](#) dargestellt ist (SEQ ID NO: 12), wird zwischen die XmnI- und die PstI-Schnittstellen inseriert. Die cDNA, die das Kollagen I ($\alpha 1$)/Decorin-Peptid codiert, ist in den [Fig. 20A](#) bis [Fig. 20C](#) (SEQ ID NO: 13) dargestellt, und ist zwischen die XmnI- und PstI-Schnittstellen inseriert.

[0148] Plasmide, die das chimäre DNA-Konstrukt enthalten, werden durch Standardtechniken, wie Gel-Elektrophorese, identifiziert. Verfahren und Materialien zur Herstellung rekombinanter Vektoren, die Transformation von Wirtszellen mit den Vektoren und die Wirtszellexpression von Polypeptiden sind in Sambrook et al., Molecular Cloning: A Laboratory Manual, oben beschrieben. Allgemein können prokaryontische oder eukaryontische Wirtszellen mit rekombinanten DNA-Plasmiden transformiert werden. Transformierte Wirtszellen können durch Phänotyp-Selektionsgene des Klonierungsvektors lokalisiert werden, welche eine Resistenz gegenüber einem bestimmten Antibiotikum ermöglichen, wenn die Wirtszellen in einem Kulturmedium gezüchtet werden, welches das Antibiotikum enthält.

[0149] Transformierte Wirtszellen werden isoliert und kultiviert, um die Expression des chimären Proteins zu fördern. Das chimäre Protein kann dann aus dem Kulturmedium isoliert werden und mit verschiedenen Verfahren wie Dialyse, Dichtegradientenzentrifugation, Flüssigsäulenchromatographie, isoelektrische Präzipitation, Lösungsmittelfraktionierung und Elektrophorese gereinigt werden. Jedoch ist die Reinigung des chimären Proteins durch Affinitätschromatographie bevorzugt, wodurch das chimäre Protein gereinigt wird, indem es an ein bindendes Protein ligiert wird, und indem es mit einem Liganden oder einem Substrat kontaktiert wird, für das das bindende Protein eine spezifische Affinität besitzt.

[0150] Um eine effizientere Expression von Säuger- oder menschlichen eukaryontischen Genen in Bakterien (Prokaryonten) zu erzielen, kann das Säuger- oder menschliche Gen unter Kontrolle eines bakteriellen Promotors plaziert sein. Ein Proteinfusions- und ein Reinigungssystem wird verwendet, um das chimäre Protein zu gewinnen. Vorzugsweise wird jedes der oben erwähnten chimären DNA-Konstrukte in einen pMal-Vektor an eine Stelle in der Polylinkersequenz des Vektors kloniert. Infolgedessen wird das chimäre DNA-Konstrukt operativ mit dem malE-Gen des pMal-Vektors fusioniert. Das malE-Gen codiert das Maltose-bindende Protein (MBP). [Fig. 23](#) stellt den pMal-Klonierungsvektor dar, der ein BMP/Kollagen-DNA-Konstrukt enthält. Eine Spacersequenz, die 10 Asparagin-Reste codiert, ist zwischen der malE-Sequenz und der Polylinkersequenz lokalisiert. Diese Spacersequenz isoliert MBP vom Protein von Interesse. [Fig. 24](#), [Fig. 25](#) und [Fig. 26](#) stellen pMal-Klonierungsvektoren dar, die DNA enthalten, welche Kollagen-Chimären mit TGF- β_1 , Decorin und einem Decorin-Peptid enthalten. Der pMal-Vektor, welcher irgendeines der chimären DNA-Konstrukte enthält, das mit malE-Gen fusioniert ist, wird in E. coli transformiert.

[0151] Die E. coli werden in einem Medium kultiviert, welches dazu führt, daß die Bakterien Maltose-bindendes Protein fusioniert an das chimäre Protein produzieren. Diese Technik verwendet den P_{tac} -Promotor des pMal-Vektors. MBP enthält eine 26 Aminosäuren lange N-terminale Signalsequenz, welche das chimäre MBP-Protein durch die Cytoplasma-Membran von E. coli leitet. Das Protein kann dann aus dem Periplasma gereinigt werden. Alternativ dazu kann der pMal-c2-Klonierungsvektor in diesem Proteinfusions- und dem Reinigungssystem verwendet werden. Der pMal-c2-Vektor enthält eine exakte Deletion der malE-Signalsequenz, welche zu einer cytoplasmatischen Expression des Fusionsproteins führt. Ein Rohzelleextrakt, welcher das Fusionsprotein enthält wird hergestellt und über eine Säule aus Amyloseharz geschüttet. Da MBP eine Affinität für die Amylose besitzt, bindet es an das Harz. Alternativ dazu kann die Säule jedes Substrat einschließen, für das MBP eine spezifische Affinität besitzt. Unerwünschte Proteine, welche in dem Rohextrakt vorhanden sind, werden durch die Säule durchgewaschen. Das MBP, welches an das chimäre Protein fusioniert ist, wird von der Säule mit einem neutralen Puffer eluiert, der Maltose oder eine andere verdünnte Lösung eines Desorptionsmittels zum Austauschen des Hybridpolypeptids enthält. Das gereinigte chimäre MBP-Protein wird mit einer Protease gespalten, wie beispielsweise Faktor Xa-Protease, um MBP von dem chimären Protein abzuspalten. Das pMal-p2-Plasmid hat eine Sequenz, welche die Erkennungsstelle für den Proteasefaktor Xa codiert, welcher nach der Aminosäuresequenz Isoleucin-Glutaminsäure-Glycin-Arginin der Polylinkersequenz spaltet.

[0152] Das chimäre Protein wird dann von dem abgespaltenen MBP durch Durchleiten der Mischung durch eine Amylosesäule getrennt. Ein alternatives Verfahren zur Abtrennung des MBPs vom chimären Protein läuft mittels Ionenaustauschchromatographie. Dieses System bringt bis zu 100 mg MBP-chimäres Protein pro Liter Kulturflüssigkeit. Siehe P. Riggs, in F. M. Ausebel, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, K. Struhl (Hrsg.) *Current Protocols in Molecular Biology*, Supplement 19 (16.6.1–16.6.10) (1990) Green Associates/Wiley Interscience, New York, New England Biolabs (Kat. # 800-65S 9pMALc2) pMal Protein and purification system (siehe auch europäisches Patent Nr. 286 239), welches ein ähnliches Verfahren zur Herstellung und Reinigung des Proteins, wie Kollagen, offenbart.)

[0153] Andere Protein-Fusions- und -Reinigungssysteme können verwendet werden, um chimäre Proteine herzustellen. Prokaryonten, wie *E. coli*, sind die bevorzugten Wirtszellen zur Expression des chimären Proteins. Jedoch sind Systeme, die eukaryontische Zelllinien benutzen, ebenfalls annehmbar, wie beispielsweise Hefe, menschliche, Maus-, Ratten-, Hamster-, Affen-, Amphibien-, Insekten-, Algen- und Pflanzenzelllinien. Beispielsweise HeLa (menschliches Epithel), 3T3 (Mausfibroblasten), CHO (Ovarzellen des Chinesischen Hamsters) und SP 2 (Mausplasmazellen) sind annehmbare Zelllinien. Die jeweiligen Wirtszellen, die ausgewählt werden, sollten mit dem jeweiligen Klonierungsvektor, der ausgewählt wird, kompatibel sein.

[0154] Ein anderes annehmbares Protein-Expressionssystem ist das Baculovirus-Expressionssystem, das von Invitrogen, San Diego, Californien, hergestellt wird. Baculoviren bilden auffällige Kristalleinschlußkörperchen innerhalb der Kerne der Zellen, die die infizieren. Jedes Kristalleinschlußkörperchen besteht aus zahlreichen Viruspartikeln, die in ein Protein eingehüllt sind, welches als Polyhedrin bezeichnet wird. In dem Baculovirus-Expressionssystem wird das native Gen, welches Polyhedrin codiert, durch ein DNA-Konstrukt substituiert, welches ein Protein oder Peptid mit einer gewünschten Aktivität codiert. Das Virus stellt dann große Mengen des Proteins her, welches durch das fremde DNA-Konstrukt codiert wird. Der bevorzugte Klonierungsvektor zur Verwendung innerhalb dieses Systems ist pBlueBac III (erhältlich von Invitrogen, San Diego, Californien). Das Baculovirussystem verwendet das Autograph californica multiple nukleare Polyhydrose-Virus (ACM-NPV) reguliert Polyhedrin-Promotorsystem, um die Expression fremder Gene zu regulieren. Das chimäre Gen, d. h. das DNA-Konstrukt, welches das chimäre Protein codiert, wird in den pBlueBac III-Vektor direkt stromabwärts des Baculoviruspolyhedrin-Promotors inseriert.

[0155] Der pBlueBac III-Transfervektor enthält einen B-Galactosidase-Reportergen, welches die Identifizierung des rekombinanten Virus ermöglicht. Das B-Galactosidase-Gen wird durch den Baculovirus ETL-Promotor (P_{ETL}) angetrieben, welcher in entgegengesetzter Richtung zum Polyhedrin-Promotor (P_{PH}) und der multiplen Klonierungsstelle des Vektors angeordnet ist. Dadurch co-exprimiert das rekombinante Virus B-Galactosidase und das chimäre Gen.

[0156] *Spodoptera frugiperda* (Sf9)-Insektenzellen werden kann mit der viralen DNA des Wildtyps und dem pBlueBac III-Vektor co-transfiziert, welcher das chimäre Gen enthält. Die Rekombinationssequenzen in dem pBlueBac III-Vektor steuern die Integration des Vektors in das Genom des Wildtyp-Baculovirus. Eine homologe Rekombination tritt auf, welche zum Austausch des nativen Polyhedrin-Gens des Baculovirus durch das DNA-Konstrukt führt, welches das chimäre Protein codiert. Wildtyp-Baculovirus, welches nicht fremde DNA codiert, exprimiert das Polyhedrin-Protein im Kern der infizierten Insektenzellen. Jedoch produzieren die Rekombinanten das Polyhedrin-Protein nicht und produzieren keine viralen Einschlußkörperchen. Statt dessen produzieren die Rekombinanten das chimäre Protein.

[0157] Alternative Insektenwirtszellen zur Verwendung mit diesem Expressionssystem sind die Sf21-Zelllinie, die von *Spodoptera frugiperda* stammt und High Five-Zelllinien, die von *Trichoplusia ni* stammen.

[0158] Andere annehmbare Klonierungsvektoren schließen Phagen, Cosmide oder künstliche Chromosomen ein. Beispielsweise ist der Bakteriophage Lambda ein nützlicher Klonierungsvektor. Dieser Phage kann Stücke fremder DNA mit bis zu ungefähr 20 000 Basenpaaren Länge aufnehmen. Das Lambda-Phagen genom ist ein lineares doppelsträngiges DNA-Molekül mit einzelsträngigen komplementären (kohäsiven) Enden, die miteinander hybridisieren können, wenn sie innerhalb einer infizierten Wirtszelle vorliegen. Die Lambda-DNA wird mit einer Restriktionsendonuklease ausgeschnitten, und die fremde DNA, z. B. die DNA die kloniert werden soll, wird mit den Phagen-DNA-Fragmenten ligiert. Das erhaltene rekombinante Molekül wird dann in infektiöse Phagenpartikel verpackt. Wirtszellen werden mit den Phagenpartikeln infiziert, welche die rekombinante DNA enthalten. Die Phagen-DNA repliziert sich in der Wirtszelle, um viele Kopien der gewünschten DNA-Sequenz herzustellen.

[0159] Cosmide sind Hybrid-Plasmid/Bakteriophagen-Vektoren, die verwendet werden können, um

DNA-Fragmente von ungefähr 40 000 Basenpaaren zu klonieren. Cosmide sind Plasmide, die ein oder mehrere DNA-Sequenzen haben, die als "cos"-Stellen bezeichnet werden, welche aus dem Bacteriophagen Lambda stammen, um Lambda-DNA in infektiöse Phagenpartikel zu verpacken. Zwei Cosmide werden mit der zu klonierenden DNA ligiert. Das erhaltene Molekül wird in infektiöse Lambda-Phagenpartikel verpackt und in Bakterien-Wirtszellen transfiziert. Wenn die Cosmide innerhalb der Wirtszelle sind, verhalten sie sich wie Plasmide und multiplizieren sich unter der Kontrolle eines Replikationsursprungs des Plasmids. Der Replikationsursprung ist eine DNA-Sequenz, die einem Plasmid ermöglicht, sich innerhalb einer Wirtszelle zu multiplizieren.

[0160] Künstliche Chromosomenvektoren der Hefe ähneln Plasmiden, ermöglichen aber den Einbau viel größerer DNA-Sequenzen von ungefähr 400 000 Basenpaaren. Die künstlichen Hefe-Chromosomen enthalten Sequenzen zur Replikation in Hefe. Das künstliche Hefe-Chromosom enthält die DNA, die in Hefezellen zu klonieren und zu transformieren ist, wo sie sich repliziert, wodurch viele Kopien der gewünschten DNA-Sequenz hergestellt werden. Wenn Phagen, Cosmide oder künstliche Hefe-Chromosome als Klonierungsvektoren verwendet werden, kann die Expression des chimären Proteins durch Kultivierung von Wirtszellen, die mit dem Klonierungsvektor transfiziert oder transformiert worden sind, in einem geeigneten Kulturmedium gewonnen werden.

[0161] Chimäre Proteine, die hier offenbart werden, sind zur Verwendung in der Behandlung von Säugern oder anderen Tieren beabsichtigt. Die therapeutisch wirksamen Reste, die oben beschrieben wurden, zum Beispiel osteogene Mittel wie beispielsweise BMPs, TGFs, Decorin und/oder Fragmente davon sind alle so anzuordnen, daß sie von physiologisch aktiven Mitteln für Zwecke dieser Beschreibung fungieren oder davon stammen. Die chimären Proteine und DNA-Konstrukte, die eine Domäne einschließen, welche von einem oder mehreren zellulären physiologisch aktiven Mitteln stammen, können für die therapeutische Behandlung in vivo, für die Forschung in vitro und/oder für diagnostische Zwecke im allgemeinen verwendet werden.

[0162] Wenn sie in vivo verwendet werden, können Formulierungen, die die vorliegenden chimären Proteine enthalten in direktem Kontakt mit lebendem Gewebe plaziert werden, einschließlich des Knochens, um das Wachstum, die Reparatur und/oder den Austausch derartigen Gewebes zu induzieren oder zu verstärken. Dies kann durch die Anwendung des chimären Proteins direkt an die Zielstelle während eines chirurgischen Eingriffs erreicht werden. Es wird in Erwägung gezogen, daß minimal invasive Techniken wie beispielsweise die Endoskopie verwendet werden sollen, um ein chimäres Protein an einem gewünschten Ort zu applizieren. Formulierungen, die chimäre Proteine enthalten, welche hier offenbart sind, können einfach aus einem oder mehreren chimären Proteinen bestehen, oder sie können ein oder mehr pharmazeutisch annehmbare Adjuvantien einschließen.

[0163] Jedes der oben beschriebenen chimären Proteine kann mit einem Implantat in Kontakt gebracht werden, daran angeheftet werden oder sonstwie eingebaut werden, wie beispielsweise in eine Vorrichtung zur Verabreichung von Medikamenten oder eine Prothesenvorrichtung. Chimäre Proteine können durch Liposomen oder andere Membran-bildende Materialien wie Alginsäure-Derivate vor dem Implantieren mikroverkapselt oder makroverkapselt werden und dann in Form eines taschenartigen Implantats implantiert werden. Das chimäre Protein kann in Strukturen in Form von Kugeln, Aggregaten aus Kernmaterial, welche in einem Kontinuum aus Wandmaterial oder kapillarer Formen eingebaut ist, mikroverkapselt werden. Mikroverkapselungstechniken sind im Fachgebiet gut bekannt und sind in der Encyclopedia of Polymer Science and Engineering, Bd. 9, S. 724 et. seq. (1980) beschrieben.

[0164] Chimäre Proteine können ebenfalls auf medizinisch nützliche Materialien geschichtet werden oder darin eingebaut werden, wie beispielsweise Netze, Tupfer, Wundbedeckung- oder Prothesenvorrichtungen wie Stäbchen, Nägel, Knochen-Platten, künstliche Gelenke, künstliche Gliedmaßen oder Knochenverbesserungs-Implantate eingebaut sein. Die Implantate können teilweise aus biokompatiblen Materialien wie Glas, Metall, Keramik, Calciumphosphat oder auf Calciumcarbonat basierenden Materialien hergestellt sein. Implantate mit biokompatiblen Biomaterialien sind im Fachgebiet gut bekannt und sind alle für die Verwendung hiernach geeignet. Implantierte Biomaterialien, die aus natürlichen Quellen stammen, wie beispielsweise Proteinfasern. Polysaccharide und behandelte natürlich gewonnene Gewebe werden in der Encyclopedia of Polymer Science and Engineering, Bd. 2, S. 267 et seq. (1989) beschrieben. Synthetische biokompatible Polymere sind im Fachgebiet gut bekannt und sind ebenfalls geeignete Implantatmaterialien. Beispiele solcher geeigneter synthetischen Polymere schließen Urethane, Olefine, Terephthalate, Acrylate, Polyester und ähnliche ein. Andere annehmbare Implantatmaterialien sind biologisch abbaubare Hydrogele oder Aggregate aus dicht verpackten Partikeln wie Polymethylmethacrylat-Kügelchen, die eine polymerisierte Hydroxyethylmethacrylat-Beschichtung haben. Siehe Encyclopedia of Polymer Science and Engineering, Bd. 2, S. 267 et seq. (1989).

[0165] Das hier beschriebene chimäre Protein stellt einen nützlichen Weg zur Immobilisierung oder Beschichtung physiologisch annehmbarer Mittel oder eines pharmakologisch annehmbaren Vehikels zur Verabreichung des physiologisch wirksamen Mittels an gewünschte Stellen im lebenden Gewebe bereit. Geeignete Vehikel schließen solche ein, die aus biologisch absorbierbaren Polymeren, biologisch kompatiblen nicht-absorbierbaren Polymeren, Lacton-Masse und gebrannten Gips hergestellt sind. Beispiele geeigneter biologisch absorbierbarer und biologisch kompatibler Polymere schließen Homopolymere, Copolymere und Mischungen aus Hydroxysäuren wie Lactide und Glycolide ein, andere absorbierbare Polymere, die alleine oder in Kombination mit Hydroxysäuren verwendet werden können, schließen Dioxanone, Carbonate wie Trimethylencarbonat, Lactone wie Caprolacton, Polyoxyalkylene und Oxylate ein. Siehe Encyclopedia of Polymer Science and Engineering, Bd. 2, S. 230, et seq. (1989).

[0166] Diese Vehikel können in Form von Kügelchen, Partikel, Masse, Beschichtungen oder schichtbildenden Trägern vorliegen. Diffusionssysteme in denen ein Kern des chimären Proteins durch eine poröse Membranschicht umhüllt ist, sind andere annehmbare Vehikel.

[0167] In einem weiteren Aspekt kann die Menge des Transports des Aminosäureanalogons in eine Zielzelle durch Kontrollieren der Tonizität des Wachstumsmediums reguliert werden. Eine hypertones Wachstumsmedium erhöht die Aufnahme von trans-4-Hydroxyprolin in *E. coli*, wie in [Fig. 2A](#) dargestellt ist. Alle bekannten Verfahren zur Erhöhung der Osmolalität von Wachstumsmedium sind für die Verwendung hier geeignet, einschließlich der Zugabe von Salzen, wie Natriumchlorid, KCl, MgCl₂ und ähnliche, und von Zuckern, wie Saccharose, Glucose, Maltose etc. und Polymeren wie Polyethylenglykol (PEG), Dextran, Cellulose, etc. und anderen Aminosäuren wie Glycin. Die Erhöhung der Osmolalität von Wachstumsmedium führt zu einer größeren intrazellulären Konzentration an Säureanaloga und einem höheren Ausmaß der Komplexierung von Aminosäure-Analoga mit tRNA. Infolgedessen erzielen Proteine, die durch die Zelle hergestellt werden, eine höhere Einbaurrate von Aminosäure-Analoga. [Fig. 12](#) stellt den Prozentsatz des Einbaus von Prolin und Hydroxyprolin in MBP unter isotonischen und hypertonen Medien-Bedingungen im Vergleich mit Prolin im ursprünglichem MBP dar. Das heißt, daß die Manipulierung der Osmolalität zusätzlich zum Einstellen der Konzentration der Aminosäure-Analoga im Wachstumsmedium einen zweiseitigen Einsatz ermöglicht, um deren Aufnahme in prokaryontische Zellen und eukaryontische Zellen zu regulieren, wie oben beschrieben wurde, und infolgedessen den Einbau in die Zielpolypeptide.

[0168] Jedes Wachstumsmedium kann hier verwendet werden, einschließlich kommerziell verfügbaren Wachstumsmediums wie M9-Minimalmedium (erhältlich von Gibco Life Technologies, Inc.), LB-Medium, NZ-CYM-Medium, "terrific broth", SOB-Medium und anderer, die im Fachgebiet gut bekannt sind.

[0169] Kollagen für verschiedene Gewebe kann verschiedene Mengen an trans-4-Hydroxyprolin enthalten. Beispielsweise enthalten die Gewebe, die eine größere Festigkeit benötigen, wie beispielsweise Knochen, eine größere Anzahl an trans-4-Hydroxyprolin-Resten als Kollagen in Geweben, die weniger Stärke benötigen, zum Beispiel Haut. Das vorliegende System stellt ein Verfahren zum Einstellen der Menge an trans-4-Hydroxyprolin in Kollagen, Kollagen-Fragmenten, Kollagen-artigen Peptiden und chimären Peptiden mit einer Kollagen-Domäne, einem Fragment einer Kollagen-Domäne oder einer Domäne eines Kollagen-artigen Peptids, das mit einer physiologisch wirksamen Domäne fusioniert ist, da durch Erhöhung oder Senkung der Konzentration an trans-4-Hydroxyprolin im Wachstumsmedium die Menge an trans-4-Hydroxyprolin, welche in derartige Polypeptide eingebaut wird, erhöht oder entsprechend gesenkt wird. Das Kollagen, Kollagen-Fragmente, Kollagen-artige Peptide und die oben erwähnten chimären Peptide können mit Hilfe vorherbestimmter Mengen an trans-4-Hydroxyprolin exprimiert werden. Auf diese Weise können die physikalischen Eigenschaften einer extrazellulären Matrix auf Grundlage der Anforderung für die endgültige Verwendung eingestellt werden. Ohne die Absicht an eine bestimmte Theorie gebunden zu sein, wird davon ausgegangen, daß der Einbau von trans-4-Hydroxyprolin in die EMP-Reste hier eine Grundlage der Selbstaggregation, die hier beschrieben ist, bereitstellt.

[0170] Nach einem weiteren Aspekt spiegelt die Kombination des Einbaus von trans-4-Hydroxyprolin in Kollagen und Fragmente davon unter Verwendung von hyperosmotischen Medien und Genen, die verändert worden sind, so daß die Codon-Verwendung stärker die wieder angibt, die in *E. coli* gefunden wird, als die Aminosäuresequenz, die im nativen menschlichen Kollagen gefunden wird, überraschend die Produktion menschlichen Kollagens und Fragmenten davon durch *E. coli* wieder, welche in der Lage waren, sich selbst zu aggregieren.

[0171] Die menschliche Kollagen-Typ I (α_1)-Gen-Sequenz ([Fig. 27A-Fig. 27E](#)) (SEQ ID NO: 15) enthält eine große Anzahl an Glycin- und Prolin-Codons (347 Glycin und 240 Prolin-Codons), welche in einer stark repeti-

tiven Weise angeordnet sind. Tabelle I unten ist eine tabellenartige Darstellung der Codon-Frequenz für menschliches Typ I (α_1)-Kollagen-Gen. Von besonderer Bedeutung ist, daß der GGA-Glycin-Codon 64-mal auftritt und der CCC-Codon für Prolin 93-mal auftritt. Diese beiden Codons werden als seltene Codons in E. coli angesehen. Siehe P. M. Sharp und W.-H. Li, Nucleic Acids Res. 14: 7737–7749, 1986. Diese und ähnliche Erwägungen anderer menschlicher Kollagen-Gene, die hier gezeigt werden, sind für die Schwierigkeit, menschliche Kollagen-Gene in E. coli zu exprimieren, verantwortlich.

Tabelle I

Codon	Anzahl	Prozentsatz	Codon	Anzahl	Prozentsatz
TTT-Phe	1	0,09	TCT-Ser	18	1,70
TTC-Phe	14	1,32	TCC-Ser	4	0,37
TTA-Leu	0	0,00	TCA-Ser	2	0,18
TTG-Leu	3	0,28	TCG-Ser	0	0,00
CTT-Leu	4	0,37	CCT-Pro	141	13,33
CTC-Leu	7	0,66	CCC-Pro	93	8,79
CTA-Leu	0	0,00	CCA-Pro	6	0,56
CTG-Leu	7	0,66	CCG-Pro	0	0,00
ATT-Ile	6	0,56	ACT-Thr	11	1,04
ATC-Ile	0	0,00	ACC-Thr	4	0,37
ATA-Ile	1	0,09	ACA-Thr	2	0,18
ATG-Met	7	0,66	ACG-Thr	0	0,00
GTT-Val	10	0,94	GCT-Ala	93	8,79
GTC-Val	5	0,47	GCC-Ala	24	2,27
GTA-Val	0	0,00	GCA-Ala	6	0,56
GTG-Val	5	0,47	GCG-Ala	0	0,00

Tabelle I (Fortsetzung)

Codon	Anzahl	Prozentsatz	Codon	Anzahl	Prozentsatz
TAT-Tyr	2	0,18	TGT-Cys	0	0,00
TAC-Tyr	2	0,18	TGC-Cys	0	0,00
TAA-***	0	0,00	TGA-***	0	0,00
TAG-***	0	0,00	TGG-Trp	0	0,00
CAT-His	0	0,00	CGT-Arg	26	2,45
CAC-His	3	0,28	CGC-Arg	6	0,56
CAA-Gln	13	1,22	CGA-Arg	11	1,04
CAG-Gln	17	1,60	CGG-Arg	1	0,09
AAT-Asn	6	0,56	AGT-Ser	4	0,37
AAC-Asn	5	0,47	AGC-Ser	11	1,04
AAA-Lys	19	1,79	AGA-Arg	9	0,85
AAG-Lys	19	1,79	AGG-Arg	0	0,00
GAT-Asp	23	2,17	GGT-Gly	174	16,46
GAC-Asp	11	1,04	GGC-Gly	97	9,17
GAA-Glu	24	2,27	GGA-Gly	64	6,05
GAG-Glu	25	2,36	GGG-Gly	11	1,04

[0172] In einem ersten Schritt wird die Sequenz des heterologen Kollagen-Gens geändert, um die Codon-Neigung in *E. coli*, die in den Codon-Verwendungstabellen (siehe z. B. Ausubel et al. (1995), *Current Protocols in Molecular Biology*, John Wiley & Sons, New York, New York; Wada et al., 1992, oben) widerzuspiegeln. Seltene *E. coli*-Codons (siehe P. M. Sharp und W.-H. Li, *Nucleic Acids Res.* 14: 7737–7749, 1986) werden vermieden. Zum Zweiten werden einzigartige Restriktionsenzym-Schnittstellen ausgewählt, die ungefähr alle 120 bis 150 Basenpaare in der Sequenz vorkommen. In einigen Fällen setzt dies die Änderung der Nukleotidsequenz voraus, aber es ändert die Aminosäuresequenz nicht. Drittens werden Oligos aus ungefähr 80 Nukleotiden so synthetisiert, daß wenn zwei derartige Oligos miteinander verbunden werden und durch eine DNA-Polymerase verlängert werden, sie einen ungefähr 120–150 Basenpaare langen Abschnitt des Gens rekonstruieren ([Fig. 28](#)). Der Abschnitt des Gens, welcher den ganz am Aminoterminal liegenden Anteil des Proteins codiert, hat ein Methionin(ATG)-Startcodon am 5'-Ende und eine einzigartige Restriktionsschnittstelle, der ein Stopp(TAAT)-Signal am 3'-Ende folgt. Die verbliebenen Abschnitte haben einzigartige Restriktionsschnittstellen an den 5'-Enden und einzigartige Restriktionsschnittstellen, gefolgt durch ein TAAT-Stopp-Signal am 3'-Ende. Das Gen wird durch sequentielle Zufügung jedes Abschnitts zum vorhergehenden 5'-Abschnitt zusammengebaut. Auf diese Weise kann jeder sukzessive größere Abschnitt unabhängig konstruiert und exprimiert werden. [Fig. 28](#) ist eine schematische Darstellung der Konstruktion des menschlichen Kollagen-Gens, ausgehend von synthetischen Oligos.

[0173] Ein Fragment der menschlichen Typ 1 $\alpha 1$ -Kollagen-Kette, welche an den C-Terminus der Glutathion-S-Transferase (GST-D4, [Fig. 29](#)) (SEQ ID NO: 18) fusioniert ist, wurde hergestellt und hinsichtlich der Expression in dem *E. coli*-Stamm JM109 (F⁻) unter Bedingungen eines hyperosmotischen Schocks getestet. Das Kollagen-Fragment schloß die C-terminalen 193 Aminosäuren der dreifach-helikalen Region ein und die 26 Aminosäuren des C-terminalen Telopeptids ein. [Fig. 29](#) ist ein Schema der Aminosäuresequenz von GST-ColIECol (SEQ ID NO: 17) und GST-D4 (SEQ ID NO: 18) Fusionsproteinen. ColIECol umfaßt 17 Aminosäuren des N-terminalen Telopeptids, 338 sich wiederholende Gly-X-Y Tripeptide und das 26 Aminosäure lange C-terminale Telopeptid. Es gibt ein einzigartiges Methionin an der Verbindung zwischen GST und D4, gefolgt von 64 Gly-X-Y-Wiederholungen und dem 26 Aminosäure langen Telopeptid. Der Rest (Phe199) in dem C-terminalen Telopeptid von D4, wo Pepsin spaltet, ist angegeben. Das Gen wurde für das Kollagen-Fragment synthetisiert, von dem synthetische Oligonukleotide entworfen wurden, um die optimale Verwendung durch *E. coli* widerzuspiegeln. [Fig. 30](#) ist eine Tabelle, die das Auftreten der vier Prolin- und vier Glycin-Codons im menschlichen Typ I $\alpha 1$ -Gen (HCol) und dem Typ 1 $\alpha 1$ -Gen mit optimierter *E. coli*-Codonverwendung (ColIECol) darstellt. Die Verwendung der verbliebenen Codons in ColIECol wurde ebenfalls für die *E. coli*-Expression nach Warda et al., oben, optimiert. Wie in Wada et al. beschrieben wird, sind die Eigenschaften der Codon-Verwendung eines breiteren Bereichs an Organismen wie auch Viren und Organellen, die Häufigkeit (pro Tausend) der Codon-Verwendung in jedem Organismus, für die mehr als 20 Gene verfügbar sind auf Grundlage der Summierung der Anzahl der Codon-Verbindungen zu berechnen. Das Protein GST-D4 wurde effizient in JM109 (F⁻) in Minimalmedium, das kein Prolin enthält, das aber mit Hyp und NaCl supplementiert ist (siehe [Fig. 31](#) und [Fig. 32](#)) exprimiert. Die Expression war von der Induktion mit Isopropyl-1-thio- β -galactopyranosid (IPTG), trans-4-Hydroxyprolin und NaCl abhängig. Bei einer festgelegten NaCl-Konzentration von 500 mM war die Expression bei trans-4-Hydroxyprolin-Konzentrationen unter ~20 mM minimal, während das Expressionsniveau bei trans-4-Hydroxyprolin-Konzentrationen über 40 mM ein Plateau erreichte. Siehe [Fig. 31](#), welche ein Gel darstellt, das die Expression und Abhängigkeit der Expression von GST-D4 von Hydroxyprolin zeigt. Die Konzentration von Hydroxyprolin ist über jeder Spur angegeben. Das Osmolyt (NaCl) wurde zu 500 mM in jede Kultur gegeben und jede wurde mit 1,5 mM IPTG induziert. Der Pfeil gibt die Position von GST-D4 an. Gleichermaßen führten bei einer festgelegten trans-4-Hydroxyprolin-Konzentration von 40 mM NaCl Konzentrationen unter 300 mM zu einer geringen Proteinanhäufung und die Expression sank bei über 700 bis 800 mM NaCl. Siehe [Fig. 32](#), welche ein Gel zeigt, daß die Expression von GST-D4 in hyperosmotischem Medium zeigt. Die Spuren 2 und 3 sind nicht-induzierte bzw. induzierte Proben, jeweils ohne zugegebenes Osmolyt. Die Identität und Menge des Osmolyts ist über jeder der anderen Spuren angegeben. trans-4-Hydroxyprolin wurde zu 40 mM in jeder Kultur gegeben, und die Kulturen, ausgenommen der in Spur 1, wurden mit 1,5 mM IPTG induziert. Der Pfeil gibt die Position von GST-D4 an.

[0174] Sowohl Saccharose oder KCl kann NaCl als Osmolyt substituieren (siehe [Fig. 32](#)). So war die durch den osmotischen Schock vermittelte intrazelluläre Anhäufung von trans-4-Hydroxyprolin eher eine kritische Determinante der Expression als die präzise chemische Identität des Osmolyts. Trotz der großen Anzahl an Prolinen (66) in GST-D4, seiner Größe (46 kDa) und der nicht-optimalen Wachstumsbedingungen wurde es zu ungefähr 10% des gesamten zellulären Proteins exprimiert. Exprimierte Proteine von geringerer als der vollständigen Länge, welche ein Hinweis auf eine abgebrochene Transkription, Translation oder mRNA Instabilität sind, wurden nicht nachgewiesen.

[0175] Das Gen für Protein D4 enthält 52 Prolin-Codons. In Expressionsexperimenten, die in den [Fig. 31](#) und [Fig. 32](#) widergespiegelt sind, wurde erwartet, daß trans-4-Hydroxyprolin an jedem dieser Codons inseriert werden würde, was zu einem Protein führen würde, in dem trans-4-Hydroxyprolin alle Proline substituiert worden ist. Um das zu bestätigen, wurde GST-D4 mit BrCN in 0,1 N HCl an den Methioninen innerhalb von GST und an den einzigartigem Methionin am N-Terminus des D4 gespalten, und D4 wurde durch Umkehrphasen-HPLC gereinigt. Unbehandeltes GST-D4 wurde in 0,1 M HCl in einer Rundbodenflasche unter Rühren gelöst. Nach der Zugabe eines 2- bis 10-fachen molaren Überschusses an reinem, kristallinem BrCN, wurde die Flasche geleert und mit Stickstoff gefüllt. Die Spaltung wurde für 24 Stunden ablaufen lassen, worauf das Lösungsmittel im Vakuum entfernt wurde. Der Rest wurde in 0,1% Trifluoressigsäure (TFA) gelöst und durch Umkehrphasen-HPLC unter Verwendung einer Vydac C4 RP-HPLC-Säule (10 × 250 mm, 5 µ, 300 Å) auf einem BioCad Sprint-System (Perceptive Biosystems, Framingham, MA) gereinigt. D4 wurde mit einem Gradienten aus 15 zu 40% Acetonitril/0,1% TFA über einen Zeitraum von 45 Minuten eluiert. D4 eluierte als einzelner Gipfel bei 25% Acetonitril/0,1% TFA. Standard-BrCN-Spaltbedingungen/70% Ameisensäure führten zu einer umfassenden Formylierung von D4, vermutlich an den Hydroxyl-Gruppen der trans-4-Hydroxyprolin-Reste. Die Formylierung von BrCN/Ameisensäure-gespaltenen Proteinen ist zuvor beschrieben worden (Beavis et al., Anal. Chem., 62, 1836 (1990)). Die Aminosäure-Analyse wurde auf einem Beckman-Ionenaustausch-Instrument mit Post-Säulen-Derivatisierung durchgeführt. Die N-terminale Sequenzierung wurde auf einem Applied Biosystems-Sequenziergerät durchgeführt, das mit einem on-line HPLC-System ausgerüstet ist. Elektrospray-Massenspektren wurden mit einem VG Biotech BIO-Q-Quadropol-Analysegerät von M-Scan Inc. (West Chester, PA) gewonnen. Für CD-Temperaturschmelzen wurde die Temperatur in 0,5°C-Anstiegen von 4°C auf 85°C mit einer 4-minütigen Äquilibration zwischen den Schritten erhöht. Die Daten wurden bei 221,5 nm aufgezeichnet. Die thermische Umwandlung wurde unter Verwendung des Programms ThermoDyne (MORE) berechnet. Die Elektrospray-Massenspektroskopie dieses Proteins ergab ein einzelnes molekulares Ion, das einer Masse von 20,807 Da entspricht. Diese Masse liegt innerhalb von 0,05% dessen, was für D4 erwartet würde, falls es 100% trans-4-Hydroxyprolin anstelle von Prolin enthält. Prolin wurde in der Aminosäure-Analyse des gereinigten D4 nicht entdeckt, das wiederum mit einer kompletten Substituierung von trans-4-Hydroxyprolin für Prolin übereinstimmt. Um ferner zu bestätigen, daß die trans-4-Hydroxyprolin-Substituierung nur an den Prolin-Codons aufgetreten ist, wurden die 13 N-terminalen Aminosäuren von D4 wie oben sequenziert. Die ersten 13 Codons von D4 geben die Proteinsequenz H₂N-Gly-Pro-Pro-Gly-Leu-Ala-Gly-Pro-Pro-Gly-Glu-Ser-Gly (SEQ ID NO: 41) an. Die Sequenz, die gefunden wurde, war H₂N-Gly-Hyp-Hyp-Gly-Leu-Ala-Gly-Hyp-Hyp-Gly-Glu-Der-Gly (SEQ ID NO: 42) siehe [Fig. 69](#). Zusammengefaßt zeigen diese Ergebnisse, daß trans-4-Hydroxyprolin (Hyp) nur an den Prolin-Codons inseriert war, und daß die Zuverlässigkeit der Translationsmaschinerie von E. coli in sonstiger Weise weder durch die hohen intrazellulären Konzentrationen von trans-4-Hydroxyprolin oder durch hyperosmotische Kulturbedingungen geändert war.

[0176] Um zu bestimmen, ob D4, welches trans-4-Hydroxyprolin sowohl an den X- und Y-Positionen enthält, homotrimere Helices bildet und um die Stabilität des nativen Kollagens damit zu vergleichen, wurde folgendes bemerkt: In Phosphatpuffer mit neutralen pH weist D4 ein zirkuläres Dichroismus(CD)-Spektrum auf, das charakteristisch für eine Tripelhelix ist (siehe [Fig. 33](#) und Bhatnagar et al., Circular Dichroism and the Conformational Analysis of Biomolecules, G. D. Fasman, Hrsg. Plenum Press, New York (1966, S. 183). [Fig. 33](#) stellt die zirkulären Dichroismus-Spektren von nativem und wärme-denaturiertem D4 in neutralem Phosphatpuffer dar. HPLC-gereinigtes D4 wurde in 0,1 M Natriumphosphat, pH 7,0, auf eine Endkonzentration von 1 mg/ml gelöst ($E^{280} = 3628 \text{ M}^{-1} \cdot \text{cm}^{-1}$). Die Lösung wurde bei 4°C für 2 Tage inkubiert, um zu ermöglichen, daß sich vor der Analyse Tripelhelices bilden. Die Spektren wurden auf einem Aviv-Modell 62DS-Spektropolarimeter (Yale University, Molecular Biophysics und Biochemistry Department) gewonnen. Eine 1 mm-Weglänge Quarz-Suprasil-Fluorimeterzelle wurde verwendet. Nach einer 10-minütigen Inkubationsdauer bei 4°C wurden Standard-Wellenspektren von 260 bis 190 nm unter Verwendung von 10-sekündigen Akquisitions-Zeiträumen und 0,5 nm Scan-Schritten aufgezeichnet. Dieses Spektrum wird durch eine negative Elliptizität bei 198 nm und positive Elliptizität bei 221 nm gekennzeichnet. Die Größenordnungen beider dieser Absorptionen waren in neutralem pH-Puffer größer, wenn es mit sauren Bedingungen verglichen wurde. Eine vergleichbare Abhängigkeit der Stabilität vom pH-Wert ist für Kollagen-artige Dreifach-Helices bemerkt worden. Siehe z. B. Venugopal et al., Biochemistry 33, 7948 (1994). Die Erwärmung auf 85°C für 5 Minuten vor dem Gewinnen des CD-Spektrums senkte das Ausmaß der Absorption bei 198 nm und verhinderte die Absorption bei 221 nm ([Fig. 33](#)). Dieses Verhalten ist ebenfalls typisch für die Tripelhelix-Struktur von Kollagen. Siehe R. S. Bhatnagar et al., Circular Dichroism and the Conformational Analysis of Biomolecules, G. D. Gasman Hrsg. oben. Ein thermisches Schmelzprofil von D4, das wie oben angegeben in Phosphatpuffer durchgeführt wurde, ergab eine Schmelztemperatur von 29°C. Ein Fragment der C-terminalen Region der Kollagen-Kette von Typ α1 vom Rind, welche mit der Länge von D4 vergleichbar ist, bildet homotrimere Helices mit einer Schmelztemperatur von 26°C. (Siehe A. Rossi et al., Biochemistry 35, 6048 (1996)).

[0177] Die Resistenz gegenüber einem Pepsinverdau ist ein zweiter häufig verwendeter Hinweis auf die Dreifachhelix-Struktur. Bei 4°C wird der größte Teil von D4 schnell durch Pepsin in ein Protein von etwas geringerem Molekulargewicht verdaut. [Fig. 34](#) ist ein Gel, das das Ergebnis des Verdaus von D4 mit Rinderpepsin zeigt. Gereinigtes D4 wurde in 0,1 M Natriumphosphat, pH 7,0 auf 1,6 µg/µl gelöst und für 7 Tage bei 4°C inkubiert. Aliquots (10 µl) wurden in 1,5 ml Zentrifugenröhrchen plaziert und mit Wasser und mit 1 M Essigsäure-Lösungen auf ein Endvolumen von 25 µl auf 20 mM endgültige Essigsäure-Konzentration eingestellt. Jedes Röhrchen wurde dann für 20 Minuten bei der angegebenen Temperatur inkubiert und Pepsin (0,5 µl einer 0,25 µg/µl-Lösung) wurde zu jedem Röhrchen hinzugegeben und der Verdau wurde für 45 Minuten ablaufen gelassen. Nach dem Verdau wurden Proben mit Beladungspuffer abgeschreckt und mittels SDS-PAGE analysiert. Jedoch ist das anfängliche Pepsin-Spaltprodukt gegenüber einem weiteren Verdau bis zu ungefähr 30°C resistent. Die aminoterminal Sequenzierung des ursprünglichen Pepsin-Spaltprodukts, wie oben zeigte daß der N-Terminus mit dem des D4 vollständiger Länge identisch war. Die obige Massenspektrometrie-Analyse des Verdauprodukts ergab ein Eltern-Ion mit einem Molekulargewicht, das mit der Spaltung des C-terminalen Telopeptids auf der N-terminalen Seite von Phe119 übereinstimmt (siehe [Fig. 29](#)), was vermuten läßt, daß dieser Teil des Proteins entweder kugelförmig ist oder von schlecht definierter Struktur und durch Pepsin leicht gespalten wird, wobei die tripelhelikale Region gegenüber einem Verdau resistent ist. So wurden trotz einer umfassenden trans-4-Hydroxyprolin für Prolin-Substituierung sowohl an den X- und Y-Positionen von D4 Tripelhelices gebildet, die eine ähnliche Stabilität wie Fragmente von Rinderkollagen vergleichbarer Größe haben, welche Hyp im normalen Prozentbereich haben und nur an der Y-Position.

[0178] Die Kollagen-Kette vollständiger Länge des menschlichen Typ I α1, obwohl sie mehr als viermal größer ist als D4, wurde ebenfalls als N-terminale Fusion mit GST (GST-ColECol, [Fig. 29](#)) in JM108 (F⁻) in Hyp/NaCl-Medium exprimiert. [Fig. 35](#) ist ein Gel, das die Expression von GST-HCol und GST-ColECol darstellt. trans-4-Hydroxyprolin wurde zu 40 mM zugegeben und NaCl zu 500 mM. Die Expression wurde mit 1,5 mM IPTG induziert. Der Pfeil gibt die Position von GST-ColECol an. In den Verfahren, die zu den Gelen führen, die in den [Fig. 31](#), [Fig. 32](#) und [Fig. 35](#) gezeigt sind, wurden 5 ml-Kulturen von JM109 (F⁻), welche das Expressionsplasmid beherbergen in LB-Medium, welches 100 µg/ml Ampicillin enthält, über Nacht gezüchtet. Die Kulturen wurden zentrifugiert und die Zellpellets wurden zweimal mit 5 ml M9/Amp-Medium (siehe J. Sambrook, E. F. Fritsch, T. Maniatis, Molecular Cloning: A Laboratory Manual (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1989)) gewaschen, welches 0,5 Glucose und 100 µg/ml aller Aminosäuren ausgenommen Glycin und Alanin ergänzt war, welche zu 200 µg/ml vorlagen und kein Prolin enthielt. Die Zellen wurden endgültig in 5 ml den oben erwähnten Mediums resuspendiert. Nach einer Inkubation bei 37°C für 30 min wurden Hydroxyprolin, Osmolyt oder IPTG, wie angegeben zugegeben. Nach 4 Stunden wurden Aliquots der Kulturen mittels SDS-PAGE analysiert.

[0179] Wie bei D4 wurde das Gen für Protein ColECol aus synthetischen Oligonukleotiden konstruiert, welche so entworfen wurden, um die Codon-Verwendung in hoch-exprimierenden E. coli-Genen zu emittieren. Im Gegensatz zu GST-ColECol, konnte eine Expression der GST-menschlichen Typ I α1-Kollagen-Genfusion (pHCol), welche mit GST-ColECol in der codierten Aminosäuresequenz identisch ist, welche hier nicht die menschliche Codon-Verteilung enthält, in Coomassie-Blau gefärbten SDS-PAGE-Gels von Gesamtzelllysaten induzierter JM109 (F⁻)/pHCol-Kulturen ([Fig. 35](#)) nicht nachgewiesen werden. Das Gen für das Kollagen-Polypeptid vom Typ I α1 wurde durch die Polymerasekettenreaktion des Gens aus mRNA kloniert, die aus menschlichen Vorhautzellen (HS27, ATCC 1634) mit Primern, die nach der veröffentlichten Gensequenz (GenBank Z74615) entworfen wurden isoliert wurde. Der 5'-Primer fügte eine flankierende EcoRI-Erkennungsstelle hinzu und der 3'-Primer, eine flankierende HindIII-Erkennungsstelle. Das Gen wurde in die EcoRI/HindIII-Schnittstelle des Plasmids pBSKS⁺ (Stratagene, La Jolla, CA) kloniert, vier Mutationen wurden unter Verwendung des ExSite-Mutagenese-Kits korrigiert (Stratagene, La Jolla, CA), die Sequenz wurde durch Dideoxy-Sequenzierung bestätigt, und schließlich wurde das EcoRI/XhoI-Fragment in das Plasmid pGEX-4T.1 (Pharmacia, Piscataway, NJ) subkloniert. Das GST-HCol-Gen ist Expressions-kompetent, da ein Protein des gleichen Molekulargewichts wie GST-ColECol nachgewiesen wird, wenn Immunblots von Gesamtzelllysaten mit einem Anti-Typ-I-Kollagen-Antikörper untersucht werden. So sind Sequenzen oder strukturelle Unterschiede zwischen den Genen für ColECol und HCol kritische Determinanten der Expressionseffizienz in E. coli. Dies liegt wahrscheinlich an der Codon-Verteilung in diesen Genen und schließlich der Unterschiede in den tRNA-Isoakzeptor-Niveaus in E. coli verglichen mit Menschen.

[0180] GST-ColECol, GST-D4 und GST-HCol reichern sich im Medium für den hyperosmotischen Schock nicht an, wenn Prolin für Hydroxyprolin substituiert ist oder in angereichertem Medium. Eine mögliche Erklärung liegt darin, daß die trans-4-Hydroxyprolin-haltigen Proteine gegenüber einem Abbau resistent sein können, da sie sich in Protease-resistente Dreifach-Helices falten, während die Prolin-haltigen Proteine diese Struktur nicht annehmen. Die größere Anzahl an Codons, die für E. coli nicht optimal sind, welche im menschlichen

Gen gefunden werden, und die Instabilität von Prolin-haltigem Kollagen in *E. coli* kann teilweise erklären, warum die Expression von menschlichem Kollagen in *E. coli* zuvor nicht beschrieben worden ist.

[0181] Wie oben diskutiert worden ist, werden Kollagen-mimetische Polypeptide, d. h. künstlich entworfene Polypeptide, die mit Kollagen einige Zusammensetzungs- und Strukturmerkmale gemeinsam haben, hierin ebenfalls bereitgestellt. Solche Kollagen-mimetischen Polypeptide können ebenfalls dazu veranlaßt werden, Aminosäure-Analoga einzubauen, wie oben beschrieben wurde. GST-CM4 besteht aus Glutathion-S-Transferase, welche an 30 Wiederholungen der Gly-X-Y-Sequenz fusioniert ist. Die sich wiederholenden Gly-X-Y-Sektion macht die sich wiederholende Gly-X-Y-Einheit des menschlichen nach und wird als Kollagenmimetikum 4 oder hier als CM4 bezeichnet. So wurde gezeigt, daß die Hydroxyprolin-Einbautechnik ebenfalls mit einer Protein und einer DNA-Sequenz funktioniert, die analog ist zu der, die im menschlichen Kollagen gefunden wird. Die Analyse der Aminosäure von gereinigtem CM4-Protein, das in dem *E. coli*-Stamm JM109 (F⁻) unter Hydroxyprolin-Einbaubedingungen exprimiert wird, verglichen mit der Analyse des gleichen Proteins, das unter Prolin-Einbaubedingungen exprimiert wurde zeigt, daß die Techniken hierin zu einer im wesentlichen vollständigen Substituierung mit Hydroxyprolin für Prolin führen. Die Aminosäureanalyse wurde anhand von CM4-Protein durchgeführt, welches vom GST abgespalten und gereinigt wurde. Dies entfernt jegliche möglichen Unklarheiten, die mit dem Fusionsprotein verbunden sind.

[0182] Die Expression in Medien, die mindestens ungefähr 200 mM NaCl enthalten, ist bevorzugt, um signifikante Mengen eines Proteins anzuhäufen, welches Hydroxyprolin enthält. Eine Konzentration von ungefähr 400–500 mM NaCl scheint optimal zu sein. Entweder KCl, Saccharose oder Kombinationen daraus können zur Substituierung von oder mit NaCl verwendet werden. Jedoch führt die Expression in Medien ohne zugegebenes Osmolyt (d. h. unter Bedingungen, die stärker die von Deming et al., *In Vivo Incorporation of Prolin Analogs into Artificial Protein*, Poly. Mater. Sci. Engin. Proceed., oben, nachmachen) nicht zu einer signifikanten Expression von Hydroxyprolin-haltigen Proteinen in JM109 (F⁻). Dies wird in [Fig. 36](#) dargestellt, welche ein Scan eines SDS-PAGE-Gels ist, welches die Expression von GST-CM4 in Medien mit oder ohne 500 mM NaCl und die entweder Prolin oder Hydroxyprolin enthalten, zeigt. Das SDS-PAGE-Gel stellt Proben 5 Stunden nach der Induktion von GST-CM4 dar, welches in JM109 (F⁻) exprimiert wurde. Gleiche Mengen, basierend auf der OD von 600 nm, jeder Kultur wurden in jede Spur geladen. Die Gele wurden mit Coomassie Blue gefärbt, entfärbt und auf einem PDI 420 oe-Scanner gescannt. Spur 1: 2,5 mM Prolin/0 mM NaCl. Spur 2: 2,5 mM Prolin/500 mM NaCl. Spur 3: 80 mM Hydroxyprolin/0 mM NaCl. Spur 4: 80 mM Hydroxyprolin/500 mM NaCl. Spur 5: Molekulargewichtsmarker. Der untere Pfeil gibt die Migrationsposition von Prolin-haltigem GST-CM4 in Spuren 1 und 2 wider. Der obere Pfeil gibt die Migrationsposition von Hydroxyprolin-haltigem GST-CM4 in Spuren 3 und 4 an. Anzumerken ist, das GST-CM4, welches in Gegenwart von Hydroxyprolin exprimiert wird, bei einem höheren augenscheinlichen Molekulargewicht läuft (vergleiche Spuren 1 und 4). Dies ist zu erwarten, da Hydroxyprolin ein höheres Molekulargewicht hat als Prolin. Wenn alle Proline in GST-CM4 durch Hydroxyprolin substituiert wurden, beträgt der Molekulargewichtsanstieg 671 Da (+2%). Zu bemerken ist ebenfalls, daß das Protein, das in Gegenwart von Prolin exprimiert wird, sich in Kulturen unabhängig von der NaCl-Konzentration anhäuft (vergleiche Spuren 1 und 2). Im Vergleich dazu tritt eine signifikante Expression in Gegenwart von Hydroxyprolin nur in der Kultur auf, die 500 mM NaCl enthält (vergleiche Spuren 3 und 4). [Fig. 37](#) stellt ferner die Abhängigkeit der Expression von der NaCl-Konzentration dar, indem gezeigt wird, daß eine signifikante Expression von GST-CM4 nur bei einer NaCl-Konzentration von größer als 200 mM auftritt. Das SDS-PAGE-Gel spiegelt 6 Stunden nach Induktion Proben von GST-CM4 wider, welche in JM109 (F⁻) bei verschiedenen Konzentrationen an NaCl exprimiert wurden. Alle Kulturen enthielten 80 mM Hydroxyprolin. Spur 1: 500 mM NaCl, nicht induziert. Spuren 2–6: 500 mM, 400 mM, 300 mM, 200 mM bzw. 100 mM NaCl. Alle wurden mit 1,5 mM IPTG induziert. Spur 7: Molekulargewichtsmarker. Der Pfeil zeigt die Migrationsposition von Hydroxyprolin-haltigem GST-CM4 an. [Fig. 38](#) ist ein Scan eines SDS-PAGE-Gels der Expression von GST-CM4 in entweder 400 mM NaCl oder 800 mM Saccharose. Das SDS-PAGE-Gel zeigt Proben von GST-CM4, welche in JM109 (F⁻) exprimiert worden sind, 4 Stunden nach der Induktion. Alle Kulturen enthielten 80 mM Hydroxyprolin und alle, außer, die in Spur 2 elektrophoretisch aufgetrennt wurde, enthielten 400 mM NaCl. Spur 2 zeigt die Expression in Saccharose statt in NaCl. Spur 1: Molekülsmarker. Spur 2: 800 mM Saccharose (kein NaCl). Spuren 3 bis 9: 0 mM, 0,028 mM, 0,1 mM, 0,4 mM, 0,8 mM, 1,25 mM, 2,5 mM Prolin. Der obere Pfeil zeigt die Migrationsposition von Hydroxyprolin-haltigem GST-CM4 an und der untere Pfeil zeigt die Migrationsposition von Prolin-haltigem GST-CM4 an. Die Expression ist in beiden Fällen erkennbar (vergleiche Spuren 2 und 3).

[0183] Wenn die Expression von GST-CM4, wie in Beispiel 17 unten beschrieben, bei verschiedenen Verhältnissen von Hydroxyprolin und Prolin durchgeführt wird, scheint das exprimierte Protein verschiedene Mengen an Hydroxyprolin zu enthalten. So ist, wenn nur Hydroxyprolin während der Expression vorhanden ist, ein einziges exprimiertes Protein des erwarteten Molekulargewichts auf einem SDS-PAGE-Gel sichtbar ([Fig. 38](#), Spur 3). Wenn mehr als ungefähr 1 mM Prolin vorhanden ist, ist wieder ein einzelnes exprimiertes Protein sicht-

bar, aber bei einem niedrigeren offensichtlichen Molekulargewicht als für ein Protein erwartet werde, das nur Prolin enthält ([Fig. 38](#), Spuren 7–9). Wenn geringere Prolin-Mengen während der Expression verwendet werden, sind Arten mit offensichtlichen Molekulargewichten, die zwischen diesen Extremen liegen, sichtbar. Dieses Phänomen, welches als "Schmier" oder "Leiter" aus Proteinen, die zwischen den zwei Molekulargewichtsextremen auf einem SDS-PAGE-Gel laufen, sichtbar ist, wird in Spuren 3–9 der [Fig. 38](#) dargestellt. Spuren 3–9 auf diesem Gel sind Proteine der Expression einer festgelegten Konzentration an 80 mM Hydroxyprolin und 400 mM NaCl. Jedoch steigt beim Voranschreiten von Spur 3 bis 9 die Prolin-Konzentration von keiner (Spur 3) bis 2,5 mM (Spur 9) und die Expression ändert sich von einem Protein mit höherem Molekulargewicht (Hydroxyprolin-haltiges GST-CM4) zu einem niedrigeren Molekulargewicht (Prolin-haltiges GST-CM4). Bei Prolin-Konzentrationen von 0,025 mM und 0,1 mM sind Arten von dazwischenliegenden Molekulargewicht erkennbar (Spuren 4 und 5). Dies stellt klar dar, daß der prozentuale Einbau von Hydroxyprolin in ein exprimiertes Protein durch Expression in verschiedenen Verhältnissen des Analogons zur Aminosäure kontrolliert werden kann.

[0184] Das Hungern nach Prolin vor dem Eingau von Hydroxyprolin ist eine wichtige hier verwendete Technik. Sie stellt sicher, daß kein restliches Prolin während der Expression vorhanden ist, welches mit Hydroxyprolin wetteifern könnte. Dies ermöglicht eine im wesentlichen 100%ige Substituierung durch das Analogon. Wie in [Fig. 38](#) gezeigt ist, ermöglichen die Hungerbedingungen die Expression bei präzise kontrollierten Verhältnissen von Prolin und Hydroxyprolin. Die Menge an Hydroxyprolin gegenüber Prolin, welches in das rekombinante Protein eingebaut wird, kann dadurch kontrolliert werden. So können bestimmte Eigenschaften des rekombinanten Proteins, die von der relativen Menge des Analogons, das eingebaut wird, abhängen, durch die vorliegende Methode maßgeschneidert werden, um Polypeptide mit einzigartigen und günstigen Eigenschaften herzustellen.

[0185] Menschliches Kollagen, Kollagen-Fragmente, Kollagen-artige Peptide (Kollagen-Mimetika) und die oben erwähnten chimären Polypeptide, die durch rekombinante Verfahren hergestellt werden, haben merkliche Vorteile gegenüber Kollagen und dessen Derivaten, welche aus nicht-menschlichen Tieren gewonnen werden. Falls ein menschliches Gen verwendet wird, wird Kollagen im Zusammenhang eines medizinischen Implantats nicht als Xenograft wirken. Außerdem kann entgegen dem natürlich auftretenden Kollagen das Ausmaß der Prolin-Hydroxylierung vorherbestimmt werden. Dieser nicht zuvor gemachte Grad an Kontrolle ermöglicht eine detaillierte Untersuchung der Beteiligung von trans-4-Hydroxyprolin an der Stabilisierung der Dreifachhelix, der Fibrillenbildung und der biologischen Aktivität. Zusätzlich wird der Entwurf medizinischer Implantate, welche auf der gewünschten Stärke der Kollagen-Fibrillen beruhen, ermöglicht.

[0186] Die folgenden Beispiele sind zum Zwecke der Darstellung enthalten und sollten nicht als Beschränkungen hiervon verstanden werden.

Beispiel 1

Transmembrantransport

[0187] Eine 5 ml-Kultur des E. coli-Stamms DH5 α (supE44 Δ lacU169 (ϕ 80lacZ Δ M15) hsdR17 recA1 endA1 gyrA96 thi-1 relA1), welcher ein Plasmid enthält, das eine Resistenz gegenüber Ampicillin verleiht (pMal-c2, [Fig. 1](#)) wurde in Luria-Nährflüssigkeit bis zur Konfluenz gezüchtet (ungefähr 16 Stunden nach der Inokulierung). Diese Zellen wurden verwendet, um eine 1 l-Schüttelflasche zu inokulieren, die 500 ml M9-Minimalmedium enthält (M9-Salze, 2% Glucose, 0,01 mg/ml Thiamin, 100 μ g/ml Ampicillin, welches mit allen Aminosäuren zu 20 μ g/ml ergänzt ist), welches bis zu einer optischen Dichte von AU₆₀₀ von 1,0 gezüchtet wurde (18–20 Stunden). Die Kultur wurde in der Hälfte geteilt, die Zellen wurden durch Zentrifugieren geerntet. Die Zellen einer Kultur wurden in 250 ml M9-Medium resuspendiert und die anderen in 250 ml M9-Medium, das 0,5 M NaCl enthält. Die Kulturen wurden in einem Luftschüttler für 20 Minuten bei 37°C (225 Upm) äquilibriert und in zehn 25 ml-Aliquots geteilt. Die Kulturen wurden auf den Schüttler zurückgegeben und 125 μ l an 1 M Hydroxyprolin in destilliertem H₂O wurde in jedes Röhrchen gegeben. Nach 2, 4, 8, 12 und 20 Minuten wurden 4 Kulturröhrchen (2 isotone, 4 hypertone) auf 1 μ m Polycarbonat-Filter vakuumfiltriert, welche direkt in 2 ml Mikrofugenröhrchen gegeben wurden, die 1,2 ml 0,2 M NaOH/2% SDS in destilliertem Wasser enthalten. Nach Lyse über Nacht wurden die Filter vorsichtig aus den Röhrchen entfernt, und der Überstandspuffer wurde hinsichtlich des Hydroxyprolins gemäß dem Verfahren von Grant, Journal of Clinical Pathology, 17: 685 (1964) untersucht. Die intrazelluläre Konzentration an trans-4-Hydroxyprolin gegen die Zeit ist graphisch in [Fig. 2](#) dargestellt.

Beispiel 2

Auswirkungen der Salzkonzentration auf den Transmembrantransport

[0188] Um die Auswirkungen der Salzkonzentration auf den Transmembrantransport zu bestimmen, wurde ein ähnlicher Ansatz wie in Beispiel 1 verwendet. Eine 5 ml-Kultur des *E. coli*-Stamms DH5 α (supE44 Δ lacU169 (ϕ 80lacZ Δ M15) hsdR17 recA1 ental gyrA96 thi-1 relA1), welcher ein Plasmid enthält, das eine Resistenz gegenüber Ampicillin verleiht (pMal-c2, [Fig. 1](#)) wurde in Luria-Nährflüssigkeit bis zur Konfluenz gezüchtet (ungefähr 16 Stunden nach der Inokulierung). Diese Zellen wurden verwendet, um eine 1 l-Schüttelflasche zu inokulieren, die 500 ml des M9-Minimalmediums enthält (M9-Salze, 2% Glucose, 0,01 mg/ml Thiamin, 100 μ g/ml Ampicillin ergänzt mit allen Aminosäuren zu 20 μ g/ml), welche dann bis zu einer AU_{600} von 0,6 gezüchtet wurden. Die Kultur wurde in drei gleiche Teile geteilt, die Zellen in jedem wurden mittels Zentrifugieren eingesammelt und in 150 ml M9-Medium, 150 ml M9-Medium, das 0,5 M NaCl enthält und 150 ml M9-Medium, das 1,0 M NaCl enthält, resuspendiert. Die Kulturen wurden für 20 Minuten auf einem Schüttler bei 37°C äquilibriert (225 Upm) und dann in sechs 25 ml-Aliquots aufgeteilt. Die Kulturen wurden auf den Schüttler zurückgegeben und 125 μ l an 1 M Hydroxyprolin in destilliertem H₂O wurden in jedes Röhrchen gegeben. Nach 5 und 15 Minuten wurden 9 Kulturröhrchen (3 isotone, 3 $\times$ 0,5 M NaCl und 3 $\times$ 1,0 M NaCl) auf 1 μ m Polycarbonat-Filter vakuumfiltriert, die sofort in 2 ml Mikrofugenröhrchen gegeben wurden, die 1,2 ml 0,2 M NaOH/2% SDS in Destillation H₂O enthalten. Nach einer Lyse über Nacht wurden die Filter aus den Röhrchen entfernt und der Puffer im Überstand wurde hinsichtlich des Hydroxyprolins nach dem Verfahren von Grant, oben, untersucht.

Beispiel 2A

Auswirkung der Salzkonzentration auf den Transmembrantransport

[0189] Um die Auswirkungen von Salzkonzentration auf den Transmembrantransport zu bestimmen wurde ein Ansatz der Beispiel 1 ähnelt, verwendet. Eine gesättigte Kultur von JM109 (F⁻), welche das Plasmid pD4 ([Fig. 48](#)) beherbergen, die in Luria-Nährflüssigkeit (LB) gezüchtet wurden, die 100 μ g/ml Ampicillin (Amp) enthält, wurden verwendet um 20 ml-Kulturen aus LB/Amp bis zu einer OD bei 600 nm von 0,1 AU zu inokulieren. Die Kulturen wurden unter Schütteln bei 37°C bis zu einer OD von 600 nm zwischen 0,7 und 1,0 AU gezüchtet. Die Zellen wurden durch Zentrifugieren eingesammelt und mit 10 ml M9-Medium gewaschen. Jedes Zellpellet wurde in 20 ml M9/Amp-Medium resuspendiert, das mit 0,5% Glucose und 100 μ g/ml aller Aminosäuren, ausgenommen Prolin, ergänzt war. Die Kulturen wurden bei 37°C für 30 Minuten gezüchtet, um endogenes Prolin zu entfernen. Nach dem Auswachsen wurde NaCl in den angegebenen Konzentrationen hinzugegeben, Hyp wurde zu 40 mM zugegeben, und IPTG in einer Menge von 1,5 mM. Nach 3 Stunden bei 37°C wurden Zellen aus drei 5 ml-Aliquots jeder Kultur separat auf Polycarbonat-Filtern eingesammelt und zweimal mit 5 ml M9-Medium gewaschen, das 0,5% Glucose und die geeignete Konzentration an NaCl enthält. Die Zellen wurden in 1 ml 70% Ethanol durch Vortexen für 30 min bei Raumtemperatur lysiert. Die Zelllyse-Überstände wurden bis zur Trockenheit gebracht, in 100 μ l 2,5 N NaOH resuspendiert und auf Hyp durch das Verfahren von Neuman und Logan, R. E. Neuman und M. A. Logan, Journal of Biological Chemistry, 184: 299 (1950) untersucht. Das Gesamtprotein wurde durch den BCA-Kit (Pierce, Rockford II) nach der Zelllyse durch drei Beschallungs/Einfrier-Auftau-Zyklen bestimmt. Die Daten sind Mittelwerte $\pm$ Standardabweichung aus drei-getrennten Experimenten. Die intrazelluläre Konzentration an trans-4-Hydroxyprolin gegenüber der NaCl-Konzentration ist graphisch in [Fig. 2A](#) dargestellt.

Beispiel 3

Bestimmung der Prolin-Hungerbedingungen in *E. coli*

[0190] Der Prolin-auxotrophe *E. coli*-Stamm NM519(pro⁻), welcher das Plasmid pMal-c2 einschließt, das eine Ampicillin-Resistenz verleiht, wurde in M9-Minimalmedium gezüchtet (M9-Salze, 2% Glucose, 0,01 mg/ml Thiamin, 100 μ g/ml Ampicillin, ergänzt mit allen Aminosäuren zu 20 μ g/ml, ausgenommen Prolin, welches zu 12,5 mg/l supplementiert wurde) bis zu einer konstanten AU_{600} von 0,53 AU (17 Stunden nach Inokulierung). Hydroxyprolin wurde zu 0,08 M zugegeben und Hydroxyprolin-abhängiges Wachstum wurde durch Zunahme der OD_{600} auf 0,61 AU über einen einstündigen Zeitraum nachgewiesen.

Beispiel 4

Hydroxyprolin-Einbau in Protein in *E. coli* unter Prolin-Hungerbedingungen

[0191] Das Plasmid pMal-c2 (kommerziell erhältlich von New England Biolabs), welches DNA enthält, die das Maltose-bindende Protein (MBP) codiert, wurde verwendet, um den Prolin-auxotrophen *E. coli*-Stamm NM519(pro⁻) zu transformieren. Zwei 1 l-Kulturen der transformierten NM519(pro⁻) in M9-Minimalmedium (M9-Salze, 2% Glucose, 0,01 mg/ml Thiamin, 100 µg/ml Ampicillin ergänzt mit allen Aminosäuren zu 20 µg/ml, ausgenommen Prolin, welches zu 12,5 mg/l supplementiert wurde) wurden bis zu einer AU₆₀₀ von 0,53 gezüchtet (ungefähr 17 Stunden nach Inokulierung). Die Zellen wurden durch Zentrifugieren geerntet, das Medium in einer Kultur wurde durch ein gleiches Volumen an M9-Medium ersetzt, das 0,08 M Hydroxyprolin enthält und das Medium in der zweiten Kultur wurde durch ein gleiches Volumen an M9-Medium ersetzt, das 0,08 M Hydroxyprolin und 0,5 M NaCl enthält. Nach einer Äquilibrierung für eine Stunde wurden die Kulturen mit 1 mM Isopropyl-β-D-thiogalactopyranosid induziert. Nach dem Züchten für zusätzliche 3,25 Stunden wurden die Zellen durch Zentrifugieren geerntet, in 10 ml 10 mM Tris-HCl (pH 8), 1 mM EDTA, 100 mM NaCl (TEN-Puffer) resuspendiert und durch Auftauen und Beschallen lysiert. MBP wurde durch Passagieren der Lysate über 4 ml Amyloseharz-Spin-Säulen gereinigt, durch Waschen der Säulen mit 10 ml TEN-Puffer, gefolgt von der Eluierung des gebundenen MBPs mit 2 ml TEN-Puffer, enthaltend 10 mM Maltose. Die eluierten Proben wurden unter Stickstoff in Ampullen mit dem gleichen Volumen konzentrierter HCl (11,7 M) versiegelt und für 12 Stunden bei 120°C hydrolysiert. Nach der Klärung mit Aktivkohle, wurde der Hydroxyprolin-Gehalt in den Proben mittels HPLC und das Verfahren von Grant, oben, bestimmt. Der prozentuale Einbau von trans-4-Hydroxyprolin verglichen mit Prolin in MBP ist graphisch in [Fig. 12](#) gezeigt.

Beispiel 5

Hydroxyprolin-Einbau in Protein in *S. cerevisiae* über Integrationsvektoren unter Prolin-Hungerbedingungen

[0192] Das in Beispiel 4 oben beschriebene Verfahren wird in Hefe unter Verwendung eines integrierenden Vektors durchgeführt, welcher den Prolin-Biosyntheseweg stört. Ein Gen, welches menschliches Typ I (α₁)-Kollagen codiert, wird in einen speziellen Shuttle-Vektor hinter dem induzierbaren GAL10-Promotor inseriert. Diese Promotor/Genkassette wird von 5'- und 3'-terminalen Sequenzen flankiert, die aus einem *S. cerevisiae*-Prolin-Synthetase-Gen stammen. Das Plasmid wird durch Restriktionsverdau sowohl an den 5'- und 3'-terminalen Bereichen linearisiert und verwendet, um einen Prolinprototrophen *S. cerevisiae*-Stamm zu transformieren. Die Transformationsmischung wird auf Selektionsmedium ausplattiert und die Transformanten werden ausgewählt. Durch homologe Rekombination und Gen-Zerstörung bildet das Konstrukt gleichzeitig eine stabile Integration und wandelt den *S. cerevisiae*-Stamm in einen Prolin-auxotrophen um. Eine einzelne Transformante wird ausgewählt und bei 30°C in YPD-Medium bis zu einer OD₆₀₀ von 2 AU gezüchtet. Die Kultur wird zentrifugiert und die Zellen werden in Hefe-Dropout-Medium suspendiert, das mit allen Aminosäuren, ausgenommen Prolin, ergänzt ist, und bis zu einer konstanten OD₆₀₀ bezüchtet, welche Prolin-Hungerbedingungen anzeigt. 0,08 M L-Hydroxyprolin und 2% (G/V) Galactose werden dann zugegeben. Die Kulturen werden für zusätzliche 6 bis 48 Stunden gezüchtet. Die Zellen werden durch Zentrifugieren (5000 Upm, 10 Minuten) geerntet und durch mechanischen Aufschluß lysiert. Hydroxyprolin-haltiges menschliches Typ I (α₁)-Kollagen wird mittels Ammoniumsulfat-Fraktionierung und Säulenchromatographie gereinigt.

Beispiel 6

Hydroxyprolin-Einbau in Protein in *S. cerevisiae* durch nicht-integrierende Vektoren unter Prolin-Hungerbedingungen

[0193] Das oben in Beispiel 4 beschriebene Verfahren wird in Hefe-Prolin Auxotrophen unter Verwendung eines nicht-integrierenden Vektors durchgeführt. Ein Gen, das menschliches Typ I (α₁)-Kollagen codiert, wird nach dem induzierbaren GAL10-Promotor in den YEp24-Shuttle-Vektor inseriert, welcher den selektierbaren Ura⁺-Marker enthält. Das daraus hervorgehende Plasmid wird in Prolin-auxotrophe *S. cerevisiae* durch Sphäroblasten-Transformierung transformiert. Die Transformationsmischung wird auf Selektionsmedium ausplattiert und die Transformanten werden ausgewählt. Eine einzelne Transformante wird bei 30°C in YPD-Medium bis zu einer OD₆₀₀ von 2 AU gezüchtet. Die Kultur wird zentrifugiert und die Zellen werden in Hefe-Dropout-Medium resuspendiert, das mit allen Aminosäuren ergänzt ist, ausgenommen Prolin, und bis einer konstanten OD₆₀₀ gezüchtet wird, was Prolin-Hungerbedingungen anzeigt. Dann werden 0,08 M L-Hydroxyprolin und 2% (G/V) Galactose zugegeben. Die Kulturen werden für zusätzliche 6 bis 48 Stunden gezüchtet. Die Zellen werden mittels Zentrifugation (5000 Upm, 10 Minuten) geerntet und durch mechanisches Aufschließen ly-

siert. Hydroxyprolin-haltiges menschliches Typ I (α_1)-Kollagen wurde mittels Ammoniumsulfat-Fraktionierung und Säulenchromatographie gereinigt.

Beispiel 7

Hydroxyprolin-Einbau in Protein in einem Baculovirus-Expressionssystem

[0194] Ein Gen, das menschliches Typ I (α_1)-Kollagen codiert, wird in den pBacPAK8-Baculovirus-Expressionsvektor hinter dem AcMNPV-Polyhedron-Promotor inseriert. Dieses Konstrukt wird zusammen mit linearisierter AcMNPV-DNA durch Standard-Calciumphosphat-Co-Präzipitierung in SF9 co-transfiziert. Transfektanten werden für 4 Tage bei 37°C in TNM-FH-Medium kultiviert, das mit 10% FBS ergänzt ist. Das Medium wird geerntet und rekombinante Viruspartikel werden mittels Plaque-Assay isoliert. Rekombinante Viren werden verwendet, um 1 l SF9-Zellen zu infizieren, die in Grace's-Medium ohne Prolin wachsen, welches mit 10% FBS und 0,08 M Hydroxyprolin ergänzt ist. Nach dem Wachstum bei 27°C für 2 bis 10 Tage werden die Zellen durch Zentrifugieren geerntet und durch mechanisches Aufschließen lysiert.

[0195] Hydroxyprolin-haltiges menschliches Typ I (α_1)-Kollagen wird durch Ammoniumsulfat-Fraktionierung und Säulenchromatographie gereinigt.

Beispiel 8

Hydroxyprolin-Einbau in menschliches Kollagen-Protein in Escherichia coli unter Prolin-Hungerbedingungen

[0196] Ein Plasmid (pHuCol, [Fig. 4](#)), welches die Gensequenz von menschlichem Typ I (α_1)-Kollagen ([Fig. 3A](#) und [Fig. 3B](#)) (SEQ ID NO: 1) codiert, wird hinter dem Isopropyl- β -D-thiogalactopyranosid(IPTG)-induzierbaren Tac-Promotor platziert, der ebenfalls β -Lactamase codiert, und in dem Prolin-auxotrophen Escherichia coli-Stamm NM519(pro⁻) durch Standard-Hitzeschock-Transformierung transformiert. Die Transformationskulturen werden auf Luria-Nährflüssigkeit (LB) ausplattiert, die 100 μ g/ml Ampicillin enthält, und nach dem Wachstum über Nacht wird eine einzige Ampicillin-resistente Kolonie verwendet, um 5 ml LB zu inokulieren, das 100 μ g/ml Ampicillin enthält. Nach dem Wachstum für 10 bis 16 Stunden unter Schütteln (225 Upm) bei 37°C wird diese Kultur verwendet, um 1 l M9-Minimalmedium (M9-Salze, 2% Glucose, 0,01 mg/ml Thiamin, 100 μ g/ml Ampicillin, ergänzt mit allen Aminosäuren zu 20 μ g/ml, ausgenommen Prolin, welches zu 12,5 mg/l ergänzt ist) in einer 1,5 l-Schüttelflasche zu inokulieren. Nach dem Wachstum bei 37°C, 225 Upm für 15 bis 20 Stunden nach der Inokulierung ist der optische Durchmesser bei 600 nm bei ungefähr 0,5 OD/ml konstant. Die Zellen werden durch Zentrifugieren (5000 Upm, 5 Minuten) geerntet, und das Medium dekantiert und die Zellen in 1 l M9-Minimalmedium resuspendiert, das 100 μ g/ml Ampicillin, 0,08 M L-Hydroxyprolin und 0,5 M NaCl enthält. Nach dem Wachstum für 1 Stunde bei 37°C, 225 Upm, wird IPTG zu 1 mM zugegeben und die Kulturen werden für zusätzliche 5 bis 15 Stunden wachsen gelassen. Die Zellen werden durch Zentrifugieren (5000 Upm, 10 Minuten) geerntet und durch mechanisches Aufschließen lysiert. Das Hydroxyprolin-haltige Kollagen wird mittels Ammoniumsulfat-Fraktionierung und Säulenchromatographie gereinigt.

Beispiel 9

Hydroxyprolin-Einbau in Fragmente menschlichen Kollagen-Proteins in Escherichia coli unter Prolin-Hungerbedingungen

[0197] Ein Plasmid (pHuCol-FI, [Fig. 6](#)), das die Gensequenz der ersten 80 Aminosäuren vom menschlichen Typ I (α_1)-Kollagen ([Fig. 5](#)) (SEQ ID NO: 2) codiert, welches hinter den Isopropyl- β -D-thiogalactopyranosid(IPTG)-induzierbaren Tac-Promotor platziert ist und ebenfalls β -Lactamase codiert, wird in den Prolin-auxotrophen Escherichia coli-Stamm NM519(pro⁻) durch Standard-Hitzeschockverfahren transformiert. Die Transformationskulturen werden auf Luria-Nährflüssigkeit (LB) ausplattiert, die 100 μ g/ml Ampicillin enthält und nach dem Wachstum über Nacht wird eine einzelne Ampicillin-resistente Kolonie verwendet, um 5 ml LB zu inokulieren, das 100 μ g/ml Ampicillin enthält. Nach dem Wachstum für 10 bis 16 Stunden unter Schütteln (bei 225 Upm) bei 37°C wird diese Kultur verwendet, um 1 l M9-Minimalmedium (M9-Salze, 2% Glucose, 0,01 mg/ml Thiamin, 100 μ g/ml Ampicillin, ergänzt mit allen Aminosäuren zu 20 μ g/ml, ausgenommen Prolin, welches zu 12,5 mg/l ergänzt ist) in einer 1,5 l-Schüttelflasche zu inokulieren. Nach dem Wachstum bei 37°C, 225 Upm, für 15 bis 20 Stunden nach der Inokulierung, ist die optische Dichte bei 600 nm bei ungefähr 0,5 OD/ml konstant. Die Zellen werden durch Zentrifugieren (5000 Upm, 5 Minuten) geerntet, das Medium dekantiert und die Zellen werden in 1 l M9-Minimalmedium resuspendiert, das 100 μ g/ml Ampicillin, 0,08 M L-Hydroxyprolin und 0,5 M NaCl enthält. Nach dem Wachstum für 1 Stunde bei 37°C, 225 Upm, wird IPTG in einer Menge von 1

mM zugegeben und die Kulturen werden für zusätzliche 5 bis 15 Stunden wachsen gelassen. Die Zellen werden durch Zentrifugieren (5000 Upm, 10 Minuten) geerntet und mittels mechanischem Aufschließen lysiert. Das Hydroxyprolin-haltige Kollagen-Fragment wird durch Ammoniumsulfat-Fraktionierung und Säulenchromatographie gereinigt.

Beispiel 10

Konstruktion und Expression in *E. coli* des menschlichen Kollagen-Typ I (α_1)-Gens mit optimierter *E. coli*-Codon-Verwendung

A. Konstruktion des Gens:

[0198] Die Nukleotidsequenz der helikalen Region von menschlichem Kollagen Typ I (α_1)-Gen flankiert von 17 Aminosäuren der Amino-terminalen extrahelikalen und 26 Aminosäuren der C-terminalen extrahelikalen Region ist in **Fig. 27** (SEQ ID NO: 15) gezeigt. Eine tabellarische Auflistung der Codonhäufigkeit dieses Gens ist in Tabelle 1 angegeben. Die Gensequenz, die in **Fig. 27** gezeigt ist, wurde zuerst geändert, um die Neigung für die *E. coli*-Codons widerzuspiegeln. Ein Startmethionin wurde am 5'-Ende des Gens inseriert und eine TAAT-Stoppssequenz am 3'-Ende. Die einzigartigen Restriktionsschnittstellen wurden identifiziert oder ungefähr alle 150 Basenpaare erzeugt. Das daraus resultierende Gen (HuCol^{Ec}, [Fig. 39A–Fig. 39E](#)) (SEQ ID NO: 20) hat die Codonverwendung, die in Tabelle II angegeben ist, die unten gezeigt ist. Andere Sequenzen, die sich dem *E. coli*-Codon Bias nähern, sind ebenfalls annehmbar.

Tabelle II

Codon	Anzahl	Prozentsatz	Codon	Anzahl	Prozentsatz
TTT-Phe	6	0,56	TCT-Ser	3	0,28
TTC-Phe	9	0,86	TCC-Ser	3	0,28
TTA-Leu	0	0,00	TCA-Ser	0	0,00
TTG-Leu	0	0,00	TCG-Ser	0	0,00
CTT-Leu	0	0,00	CCT-Pro	13	1,22
CTC-Leu	1	0,08	CCC-Pro	12	1,13
CTA-Leu	1	0,08	CCA-Pro	29	2,74
CTG-Leu	19	1,79	CCG-Pro	186	17,58
ATT-Ile	3	0,28	ACT-Thr	2	0,18
ATC-Ile	4	0,37	ACC-Thr	11	1,08
ATA-Ile	0	0,00	ACA-Thr	0	0,00
ATG-Met	8	0,75	ACG-Thr	4	0,37
GTT-Val	3	0,28	GCT-Ala	10	0,94
GTC-Val	5	0,47	GCC-Ala	24	2,26
GTA-Val	0	0,00	GCA-Ala	8	0,75
GTG-Val	12	1,13	GCG-Ala	80	7,56

Tabelle II (Fortsetzung)

Codon	Anzahl	Prozentsatz	Codon	Anzahl	Prozentsatz
TAT-Tyr	2	0,18	TGT-Cys	0	0,00
TAC-Tyr	2	0,18	TGC-Cys	0	0,00
TAA-***	0	0,00	TGA-***	0	0,00
TAG-***	0	0,00	TGG-Trp	0	0,00
CAT-His	0	0,00	CGT-Arg	26	2,45
CAC-His	3	0,28	CGC-Arg	26	2,45
CAA-Gln	5	0,47	CGA-Arg	0	0,00
CAG-Gln	25	2,36	CGG-Arg	1	0,09
AAT-Asn	0	0,00	AGT-Ser	1	0,09
AAC-Asn	11	1,03	AGC-Ser	32	3,02
AAA-Lys	38	3,59	AGA-Arg	0	0,00
AAG-Lys	0	0,00	AGG-Arg	0	0,00
GAT-Asp	20	1,89	GGT-Gly	148	13,98
GAC-Asp	14	1,32	GGC-Gly	178	16,82
GAA-Glu	40	3,78	GGA-Gly	9	0,85
GAG-Glu	9	0,85	GGG-Gly	12	1,13

[0199] Oligos aus ungefähr 80 Nukleotiden wurden auf einem Beckman Oligo 1000 DNA-Synthesegerät synthetisiert, gespalten und mit wäßriger NH_4OH entschützt und durch Elektrophorese in einem 7 M-Harnstoff/12% Polyacrylamid-Gel gereinigt. Jeder Satz an Oligos wurde so entworfen, daß er eine EcoRI Restriktionsenzymstelle am 5'-Ende hat, eine einzigartige Restriktionsschnittstelle nahe dem 3'-Ende, gefolgt von einer TAAT-Stopp-Sequenz und einer HindIII-Restriktionsenzymschnittstelle ganz am 3'-Ende. Die ersten vier Oligos, die die ersten 81 Aminosäuren des menschlichen Kollagen Typ I (α_1)-Gens umfassen, sind in [Fig. 40](#) angegeben, welche die Sequenz die Restriktionskarten der synthetischen Oligos zeigt, die verwendet wurden, um die ersten 243 Basenpaare des menschlichen Typ I (α_1)-Kollagen-Gens mit optimierter E. coli-Codonverwendung zu konstruieren. Oligos N1-1 (SEQ ID NO: 21) und N1-2 (SEQ ID NO: 22) wurden entworfen, um einen Start-Methionin-(ATG)-Codon am 5'-Ende des Gens zu inserieren.

[0200] In einem Fall wurden die Oligos N1-1 und N1-2 (1 μg jeweils) in 20 μl T7-DNA-Polymerasepuffer (40 mM Tris-HCl (pH 8,0), 5 mM MgCl_2 , 5 mM Dithiothreitol, 50 mM NaCl, 0,05 mg/ml Rinderserumalbumin) durch Erwärmen bei 90°C für 5 Minuten angelagert, gefolgt von der langsamen Abkühlung auf Raumtemperatur. Nach einer kurzen Zentrifugierung bei 14 000 Upm wurden 10 Einheiten T7-DNA-Polymerase und 2 μl einer Lösung aller vier dNTPs (jeweils 2,5 mM dART, dGTP, dCTP, dTTP) zu den aneinandergelagerten Oligos zugegeben. Die Verlängerungsreaktionen wurden bei 37°C für 30 Minuten inkubiert und dann für 10 Minuten auf 70°C erwärmt. Nach dem Abkühlen auf Raumtemperatur wurde HindII-Puffer (5 μl einer 10-fach-Konzentration), 20 μl H_2O und 10 Einheiten HindIII-Restriktionsenzym hinzugegeben und die Röhrchen wurden bei 37°C für 10 Stunden inkubiert. HindIII-Puffer (2 μl eine 10-fach-Konzentration), 13,5 μl , 0,5 M Tris-HCl (pH 7,5), 1,8 μl 1% Triton X100, 5,6 μl H_2O und 20 U EcoRI wurden zu jedem Röhrchen zugegeben und die Inkubation wurde für 2 Stunden bei 37°C fortgesetzt. Die Verdau wurden einmal mit einem gleichen Volumen an Phenol extrahiert, einmal mit Phenol/Chloroform/Isoamylalkohol und einmal mit Chloroform/Isoamylalkohol. Nach einer Ethanol-Präzipitierung wurde das Pellet in 10 μl TE-Puffer (10 mM Tris-HCl (pH 8,0), 1 mM EDTA) resuspendiert. Resuspendierte Pellets (4 μl) wurden über Nacht bei 16°C mit Agarosegel gereinigt im EcoRI/HindIII-verdauten pBSKS⁺-Vektor (1 μg) unter Verwendung von T4-DNA-Ligase (100 Einheiten) ligiert. Eine Hälfte der Transformationsmischung wurde durch einen Wärmeschock in DH5 α -Zellen transformiert und 100 μl der 1,0 ml-Transformationsmischung wurden auf Luria-Nährflüssigkeit (LB)-Agarplatten ausplattiert, die 70 $\mu\text{g}/\text{ml}$ Ampicillin enthalten. Die Platten wurden über Nacht bei 37°C inkubiert. Ampicillin-resistente Kolonien (6–12) wurden herausgepickt und über Nacht in LB-Medium, enthaltend 70 mg/ml Ampicillin gezüchtet. Die Plasmid-DNA wurde aus jeder Kultur durch Wizard-Minipreps (Promega Corporation, Madison WI) isoliert und auf das Vorhandensein der ungefähr 120 Basenpaare großen Insertion durch einen Verdau mit EcoRI und HindIII und das Laufenlassen der Produkte des Verdau auf Agarose-Elektrophesegelel durchmustert. Klone mit Insertionen wurden durch Standard-Didesoxy-Terminierungs-DNA-Sequenzierung bestätigt. Der korrekte Klon wurde

pBSN1-1 ([Fig. 41](#)) bezeichnet und das Kollagenfragment hat die Nukleotidsequenz, die in [Fig. 42](#) (SEQ ID NO: 25) angegeben ist.

[0201] Oligos N1-3 (SEQ ID NO: 23) und N1-4 (SEQ ID NO: 24) ([Fig. 40](#)) wurden synthetisiert, gereinigt, aneinandergelagert, verlängert und nach dem gleichen Verfahren, das oben für die Oligos N1-1 und N1-2 angegeben ist, in pBSKS⁺ kloniert. Das erhaltene Plasmid wurde als pBSN1-2A bezeichnet. Um gemeinsam die Sektionen des Kollagen-Gens von pBSN1-1- und pBSN1-2A zu klonieren, wurde das Plasmid pBSN1-1 (1 µg) für 2 Stunden bei 37°C mit RsrII und HindIII verdaut. Der verdaute Vektor wurde mittels Agarosegelelektrophorese gereinigt. Plasmid pBSN1-2A (3 µg) wurde für 2 Stunden bei 37°C mit RsrII/HindIII verdaut und die Insertion wurde mittels Agarosegelelektrophorese gereinigt. RsrII/HindIII-verdautes pBSN1-1 wurde mit diesem Insert über Nacht bei 16°C durch T4-DNA-Ligase ligiert. Eine Hälfte der Ligationsmischung wurde in DH5α-Zellen transformiert und 1/10 des Transformationsgemisches wurde auf LB-Agarplatten, die 70 µg/ml Ampicillin enthalten, ausplattiert. Nach einer Inkubation über Nacht bei 37°C wurden Ampicillin-resistente Klone herausgepickt und auf das Vorhandensein der inserierten DNA, wie oben beschrieben, durchmustert. Klone wurden mittels Didesoxy-Terminierungssequenzierung überprüft. Der korrekte Klon wurde als pBSN1-2 ([Fig. 43](#)) bezeichnet und das Kollagen-Fragment hat die in [Fig. 44](#) angegebene Sequenz.

[0202] In ähnlicher Weise wurde der Rest des Kollagengens konstruiert, so daß die endgültige DNA-Sequenz die ist, die in [Fig. 39A–Fig. 39E](#) (SEQ ID NO: 19) angegeben ist.

B) Expression des Gens in E. coli:

[0203] Nach der Konstruktion des gesamten menschlichen Kollagen-Typ I (α_1)-Gens mit der Codonverwendung, die für E. coli optimiert worden ist, wurde das klonierte Gen in E. coli exprimiert. Ein Plasmid (pHuCol^{Ec}, [Fig. 45](#)), das das gesamte synthetische Kollagen-Gen ([Fig. 39A–Fig. 39E](#)) codiert, wurde hinter den Isopropyl-β-D-thiogalactopyranosid-(IPTG)-induzierbaren Tac-Promotor plaziert, das ebenfalls β-Lactamase codiert, und in den Escherichia coli-Stamm DH5α (supE44 ΔlacU169 (φ80lacZ ΔM15) hsdR17 recA1 endA1 gyrA96 thi-1 relA1) transformiert, indem eine Standard-Wärmeschock-Transformation durchgeführt wurde. Transformationskulturen werden auf Luria-Nährflüssigkeit (LB) ausplattiert, 19 100 µg/ml Ampicillin enthält und nach einem Wachstum über Nacht wird eine einzige Ampicillin-resistente Kolonie verwendet, um 10 ml LB, das 100 µg/ml Ampicillin enthält, zu inokulieren. Nach dem Wachstum für 10 bis 16 Stunden unter Schütteln (225 Upm) bei 37°C wird diese Kultur verwendet, um 1 l LB, der 100 µg/ml Ampicillin enthält in einer 1,5 l-Schüttelflasche zu inokulieren. Nach dem Wachstum bei 37°C, bei 225 Upm, für 2 Stunden nach der Inokulierung ist die optische Dichte bei 600 nm ungefähr 0,5 OD/ml. IPTG wird zu 1 mM hinzugegeben und die Kultur wird für zusätzliche 5 bis 10 Stunden wachsen gelassen. Die Zellen werden durch Zentrifugieren (5000 Upm, 10 Minuten) geerntet und mittels mechanischem Aufschluß lysiert. Rekombinantes menschliches Kollagen wird mittels Ammoniumsulfatfraktionierung und Säulenchromatographie gereinigt. Die Ausbeute beträgt typischerweise 15 bis 25 mg/l Kultur.

Beispiel 11

Expression eines 81 Aminosäuren-Fragments von menschlichem Kollagen-Typ I (α_1) mit optimierter E. coli-Codonverwendung in E. coli

[0204] Ein Plasmid (pTrcN1-2, [Fig. 46](#)) welches die Gensequenz der ersten 81 Aminosäuren des menschlichen Typ I (α_1)-Kollagens mit optimierter E. coli-Codonverwendung codiert, welches in einer Phase mit einem 6-Histidin-Anhang am 5'-Ende des Gens kloniert ist und hinter den Isopropyl-β-D-thiogalactopyranosid-(IPTG)-induzierbaren Trc-Promotor plaziert ist, das auch β-Lactamase codiert, wurde durch Subklonieren des EcoRI/HindIII-Insertions aus pBSN1-2 in die EcoRI/HindIII-Schnittstelle des Plasmid pTrcB (Invitrogen, San Diego, CA) konstruiert. Plasmid pTrcN1-2 wurde in den Escherichia coli-Stamm DH5α transformiert (supE44 ΔlacU169 (φ80lacZ ΔM15) hsdR17 recA1 endA1 gyrA96 thi-1 relA1), indem eine Standard-Wärmeschock-Transformation durchgeführt wurde. Die Transformationskulturen wurden auf Luria-Nährflüssigkeit (LB) ausplattiert, die 100 µg/ml Ampicillin enthält und nach dem Wachstum über Nacht wurde eine einzige Ampicillin-resistente Kolonie verwendet, um 5 ml LB zu inokulieren, das 100 µg/ml Ampicillin enthält. Nach dem Wachstum für 10–16 Stunden unter Schütteln (225 Upm) bei 37°C wurde diese Kultur verwendet, um 50 ml LB, das 100 µg/ml Ampicillin enthält, in einer 250 ml-Schüttelflasche zu inokulieren. Nach dem Wachstum bei 37°C, 225 Upm, für 2 Stunden nach der Inokulierung, betrug die optische Dichte bei 600 nm ungefähr 0,5 OD/ml. IPTG wurde zu 1 mM zugegeben und die Kultur wurde für zusätzliche 5 bis 10 Stunden wachsen gelassen. Die Zellen wurden durch Zentrifugieren (5000 Upm, 10 Minuten) geerntet und bei –20°C gelagert. Die 6 Histin-Anhang-Kollagenfragmentfusion wurde auf Nickelharzsäulen gereinigt. Zellpellets wurden in 10 ml 6

M Guanidinhydrochlorid/20 mM Natriumphosphat/500 mM NaCl (pH 7,8) resuspendiert und in zwei 5 ml-Batches an das Nickelharz gebunden. Die Säulen wurden zweimal mit 4 ml Bindungspuffer (8 M Harnstoff/20 mM Natriumphosphat/500 mM NaCl (pH 7,8)) gewaschen, zweimal mit Waschpuffer 1 (8 M Harnstoff/20 mM Natriumphosphat/500 mM NaCl (pH 6,0)) und zweimal mit Waschpuffer 2 (8 M Harnstoff/20 mM Natriumphosphat (500 mM NaCl (pH 5,3))). Die 6 Histidin-Anhang-Kollagenfragmentfusion wurde von der Säule mit 5 ml Elutionspuffer (8 M Harnstoff/20 mM Natriumphosphat/500 mM NaCl (pH 4,0)) in 1 ml Fraktionen eluiert. Die Fraktionen wurden mittels Gel-Elektrophorese auf Protein untersucht und Fusionen-haltige Fraktionen wurden aufkonzentriert und bei -20°C gelagert. Die Ausbeute betrug typischerweise 15 bis 25 mg/l Kultur.

[0205] Das Kollagen wird vom 6 Histidin-Anhang durch Enterokinase abgespalten. Fusionen-haltige Fraktionen werden gegen Spaltpuffer (50 mM Tris-HCl, pH 8,0/5 mM CaCl_2) dialysiert. Nach Zugabe von Enterokinase zu 1 μg Enzym je 100 μg Fusion wird die Lösung bei 37°C für 4 bis 10 Stunden inkubiert. Das Voranschreiten der Spaltung wird durch Gel-Elektrophorese überwacht. Der abgespaltene 6 Histidin-Anhang kann von dem Kollagen-Fragment durch Durchleiten über eine Nickelharzsäule, wie oben angegeben, abgetrennt werden.

Beispiel 12

Expression in *E. coli* von Fragmenten des menschlichen Kollagen-Typ I (α_1) mit optimierter *E. coli*-Codonverwendung

[0206] Ein Plasmid (pN1-3, [Fig. 47](#)), das das Gen für die Amino-terminalen 120 Aminosäuren des menschlichen Kollagen-Typ I (α_1) mit optimierter *E. coli*-Codonverwendung codiert, welches hinter dem Isopropyl- β -D-thiogalactopyranosid(IPTG)-induzierbaren Tac-Promotor plaziert ist, und ebenfalls β -Lactamase codiert, wird in *Escherichia coli*-Stamm DH5 α (supE44 Δ lacU169 (ϕ 80lacZ Δ M15) hsdR17 recA1 endA1 gyrA96 thi-1 relA1) durch Standardwärmeschocktransformation transformiert. Die Transformationskulturen werden auf Luria-Nährflüssigkeit (LB) ausplattiert, die 100 $\mu\text{g}/\text{ml}$ Ampicillin enthalten und nach einem Wachstum über Nacht wird eine einzelne Ampicillin-resistente Kolonie verwendet, um 10 ml LB zu inokulieren, das 100 $\mu\text{g}/\text{ml}$ Ampicillin enthält. Nach dem Wachstum für 10 bis 16 Stunden unter Schütteln (225 Upm) bei 37°C , wird diese Kultur verwendet, um 1 l LB zu inokulieren, der 100 $\mu\text{g}/\text{ml}$ Ampicillin in 1,5 l-Schüttelflasche enthält. Nach dem Wachstum bei 37°C , 225 Upm, für 2 Stunden nach der Inokulierung, wird die optische Dichte bei 600 nm ungefähr 0,5 OD/ml betragen. IPTG wird zu 1 mM zugegeben, und die Kultur wird für weitere 5 bis 10 Stunden wachsen gelassen. Die Zellen werden durch Zentrifugieren (5000 Upm, 10 Minuten) geerntet und mittels mechanischem Aufschließen lysiert. Rekombinantes menschliches Kollagen wird durch Ammoniumsulfatfraktionierung und Säulenchromatographie gereinigt. Die Ausbeute beträgt typischerweise 15 bis 25 mg/l der Kultur.

Beispiel 13

Expression eines C-terminalen Fragments des menschlichen Kollagen-Typ I (α_1) mit optimierter *E. coli*-Codonverwendung in *E. coli*

[0207] Ein Plasmid (pD4, [Fig. 48](#)), welches das Gen für die Carboxy-terminalen 219 Aminosäuren des menschlichen Kollagen-Typ I (α_1) mit optimierter *E. coli*-Codonverwendung codiert, welches hinter dem Isopropyl- β -D-thiogalactopyranosid(IPTG)-induzierbaren Tac-Promotor plaziert ist und ebenfalls β -Lactamase codiert, wird in *Escherichia coli*-Stamm DH5 α (supE44 Δ lacU169 (ϕ 80lacZ Δ M15) hsdR17 recA1 endA1 gyrA96 thi-1 relA1) durch Standard-Wärmeschocktransformation transformiert. Die Transformationskulturen werden auf Luria-Nährflüssigkeit (LB) ausplattiert, welche 100 $\mu\text{g}/\text{ml}$ Ampicillin enthalten und nach dem Wachstum über Nacht wird eine einzelne Ampicillin-resistente Kolonie verwendet, um 10 ml LB zu inokulieren, das 100 $\mu\text{g}/\text{ml}$ Ampicillin enthält. Nach dem Wachstum für 10 bis 16 Stunden unter Schütteln (225 Upm) bei 37°C wird diese Kultur verwendet, um 1 l LB, welches 100 $\mu\text{g}/\text{ml}$ Ampicillin in einer 1,5 l-Schüttelflasche enthält zu inokulieren. Nach dem Wachstum bei 37°C , 225 Upm, für 2 Stunden nach der Inokulierung beträgt die optische Dichte bei 600 nm ungefähr 0,5 OD/ml. IPTG wird zu 1 mM zugegeben und die Kultur wird für zusätzliche 5 bis 10 Stunden wachsen gelassen. Die Zellen werden durch Zentrifugieren (5000 Upm, 10 Minuten) geerntet, durch mechanisches Aufschließen lysiert. Rekombinantes menschliches Kollagenfragment wird durch Ammoniumsulfatfraktionierung und Säulenchromatographie gereinigt. Die Ausbeute beträgt typischerweise 15 bis 25 mg/l Kultur.

Beispiel 14

Konstruktion und Expression des menschlichen Kollagen-Typ I (α_2)-Gens mit optimierter E. coli-Codon-Verwendung in Escherichia coli

A) Konstruktion des Gens:

[0208] Die Nukleotidsequenz des helikalen Bereichs des menschlichen Kollagen-Typ I (α_2)-Gens, welches von 11 Aminosäuren der Amino-terminalen extrahelikalen 12 Aminosäure des C-terminalen extrahelikalen Bereichs flankiert wird, ist in den [Fig. 49A–Fig. 49E](#) (SEQ ID NO: 29) gezeigt. Eine tabellarische Auflistung der Codonhäufigkeit dieses Gens ist in Tabelle III unten gezeigt. Die Gensequenz, die in den [Fig. 49A](#) bis [Fig. 49E](#) gezeigt ist, wurde zuerst geändert, um die Neigung des E. coli-Codons widerzuspiegeln. Ein Initiator-Methionin wurde am 5'-Ende des Gens inseriert und eine TAAT-Stoppsequenz am 3'-Ende. Einzigartige Restriktionschnittstellen werden identifiziert oder ungefähr alle 150 Basenpaare erzeugt. Das daraus hervorgehende Gen (HuCol(α_2)^{Ec}, [Fig. 50A](#) bis [Fig. 50E](#)) (SEQ ID NO: 31) hat die Codonverwendung, die in der Tabelle IV unten angegeben ist. Andere Sequenzen, die sich der E. coli-Codonneigung annähern, sind ebenfalls annehmbar.

Tabelle III

Codon	Anzahl	Prozentsatz	Codon	Anzahl	Prozentsatz
TTT-Phe	3	0,28	TCT-Ser	11	1,06
TTC-Phe	10	0,96	TCC-Ser	4	0,38
TTA-Leu	1	0,09	TCA-Ser	1	0,09
TTG-Leu	2	0,19	TCG-Ser	1	0,09
CTT-Leu	16	1,54	CCT-Pro	125	12,06
CTC-Leu	9	0,86	CCC-Pro	42	4,05
CTA-Leu	2	0,19	CCA-Pro	30	2,89
CTG-Leu	5	0,48	CCG-Pro	3	0,28
ATT-Ile	14	1,35	ACT-Thr	14	1,35
ATC-Ile	3	0,28	ACC-Thr	0	0,00
ATA-Ile	1	0,09	ACA-Thr	3	0,28
ATG-Met	5	0,48	ACG-Thr	1	0,09
GTT-Val	20	1,93	GCT-Ala	82	7,91
GTC-Val	5	0,48	GCC-Ala	17	1,64
GTA-Val	3	0,28	GCA-Ala	9	0,86
GTG-Val	10	0,96	GCG-Ala	0	0,00

Tabelle III (Fortsetzung)

Codon	Anzahl	Prozentsatz	Codon	Anzahl	Prozentsatz
TAT-Tyr	2	0,19	TGT-Cys	0	0,00
TAC-Tyr	3	0,28	TGC-Cys	0	0,00
TAA-***	0	0,00	TGA-***	0	0,00
TAG-***	0	0,00	TGG-Trp	0	0,00
CAT-His	7	0,67	CGT-Arg	17	1,64
CAC-His	6	0,57	CGC-Arg	6	0,57
CAA-Gln	13	1,25	CGA-Arg	6	0,57
CAG-Gln	9	0,86	CGG-Arg	4	0,38
AAT-Asn	10	0,96	AGT-Ser	11	1,06
AAC-Asn	14	1,35	AGC-Ser	4	0,38
AAA-Lys	15	1,44	AGA-Arg	16	1,54
AAG-Lys	16	1,54	AGG-Arg	6	0,57
GAT-Asp	20	1,93	GGT-Gly	179	17,27
GAC-Asp	5	0,48	GGC-Gly	74	7,14
GAA-Glu	29	2,79	GGA-Gly	80	7,72
GAG-Glu	16	1,54	GGG-Gly	16	1,54

Tabelle IV

Codon	Anzahl	Prozentsatz	Codon	Anzahl	Prozentsatz
TTT-Phe	5	0,48	TCT-Ser	7	0,67
TTC-Phe	7	0,67	TCC-Ser	12	1,15
TTA-Leu	0	0,00	TCA-Ser	0	0,00
TTG-Leu	0	0,00	TCG-Ser	0	0,00
CTT-Leu	1	0,09	CCT-Pro	10	0,96
CTC-Leu	1	0,09	CCC-Pro	0	0,00
CTA-Leu	0	0,00	CCA-Pro	15	1,44
CTG-Leu	32	3,07	CCG-Pro	177	17,00
ATT-Ile	11	1,05	ACT-Thr	3	0,28
ATC-Ile	7	0,67	ACC-Thr	6	0,57
ATA-Ile	0	0,00	ACA-Thr	0	0,00
ATG-Met	6	0,57	ACG-Thr	10	0,96
GTT-Val	18	1,72	GCT-Ala	30	2,88
GTC-Val	7	0,67	GCC-Ala	21	2,01
GTA-Val	9	0,85	GCA-Ala	20	1,92
GTG-Val	6	0,57	GCG-Ala	38	3,65

Tabelle IV (Fortsetzung)

Codon	Anzahl	Prozentsatz	Codon	Anzahl	Prozentsatz
TAT-Tyr	3	0,28	TGT-Cys	0	0,00
TAC-Tyr	2	0,19	TGC-Cys	0	0,00
TAA-***	0	0,00	TGA-***	0	0,00
TAG-***	0	0,00	TGG-Trp	0	0,00
CAT-His	2	0,19	CGT-Arg	37	3,55
CAC-His	11	1,05	CGC-Arg	18	1,72
CAA-Gln	7	0,67	CGA-Arg	0	0,00
CAG-Gln	15	1,44	CGG-Arg	0	0,00
AAT-Asn	6	0,57	AGT-Ser	0	0,00
AAC-Asn	18	1,72	AGC-Ser	13	1,24
AAA-Lys	25	2,40	AGA-Arg	0	0,00
AAG-Lys	6	0,57	AGG-Arg	0	0,00
GAT-Asp	11	1,05	GGT-Gly	209	20,07
GAC-Asp	13	1,24	GGC-Gly	141	13,54
GAA-Glu	33	3,17	GGA-Gly	0	0,00
GAG-Glu	12	1,15	GGG-Gly	0	0,00

[0209] Oligos aus ungefähr 80 Nukleotiden werden auf einem Beckman Oligo 1000 DNA-Synthesegerät synthetisiert, gespalten und mit wäßriger NH_4OH entschützt und mittels Elektrophorese in einem 7 M-Harnstoff/12% Polyacrylamid-Gel gereinigt. Jeder Satz an Oligos wird so entworfen, daß er eine EcoRI-Restriktionsenzymstelle am 5'-Ende hat, eine einzigartige Restriktionsschnittstelle nahe dem 3'-Ende; gefolgt von einer TAAT-Stopp-Sequenz und einer HindIII-Restriktionsenzymstelle ganz am 3'-Ende. Oligos N1-1 (α_2) und N1-2 (α_2) werden entworfen, um an Initiationsmethionin(ATG)-Codon am 5'-Ende des Gens einzufügen.

[0210] In einem Fall wurden die Oligos N1-2 (α_2) und N1-2 (α_2) (jeweils 1 μg) ([Fig. 51](#) stellt die Sequenz und Restriktionskarten der synthetischen Oligos dar, die verwendet wurden, um die ersten 240 Basenpaare des menschlichen Typ I (α_2)-Kollagen-Gens mit optimierter E. coli-Codonverwendung zu konstruieren) in 20 μl T7-DNA-Polymerasepuffer (40 mM Tris-HCl (pH 8,0), 5 mM MgCl_2 , 5 mM Dithiothreitol, 50 mM NaCl, 0,05 mg/ml Rinderserumalbumin) aneinandergelagert, indem sie bei 90°C für 5 Minuten erwärmt werden, gefolgt vom langsamen Abkühlen auf Raumtemperatur. Nach einem kurzen Zentrifugationsschritt bei 14 000 Upm werden 10 Einheiten T7-DNA-Polymerase und 2 μl einer Lösung aus allen vier dNTPs (jeweils 2,5 mM dATP, dGTP, dCTP, dTTP) zu den aneinandergelagerten Oligos zugegeben. Die Verlängerungsreaktionen werden bei 37°C für 30 Minuten inkubiert und dann für 10 Minuten auf 70°C erwärmt. Nach dem Abkühlen auf Raumtemperatur wird HindIII-Puffer (5 μl einer 10-fach-Konzentration), 20 μl H_2O und 10 Einheiten HindIII-Restriktionsenzym hinzugegeben und die Röhrchen werden bei 37°C für 10 bis 16 Stunden inkubiert. HindIII-Puffer (2 μl einer 10-fach-Konzentration), 13,5 μl , 0,5 M Tris-HCl (pH 7,5), 1,8 μl von 1% Triton X100, 5,6 μl H_2O und 20 U EcoRI werden zu jedem Röhrchen gegeben und die Inkubation wird für 2 Stunden bei 37°C fortgesetzt. Die Verdau werden einmal mit einem gleichen Volumen an Phenol extrahiert, dann einmal mit Phenol/Chloroform/Isoamylalkohol und einmal mit Chloroform/Isoamylalkohol. Nach der Ethanol-Präzipitierung wird das Pellet in 10 μl TE-Puffer (10 mM Tris-HCl (pH 8,0), 1 mM EDTA) resuspendiert. Das resuspendierte Pellet (4 μl) wird über Nacht bei 16°C mit Aragoosegel gereinigtem EcoRI/HindIII-verdauten pBSKS⁺-Vektor (1 μg) unter Verwendung von T4-DNA-Ligase (100 Einheiten) ligiert. Eine Hälfte der Transformationsmischung wird durch einen Wärmeschock in DH5 α -Zellen transformiert und 100 μl der 1,0 ml-Transformationsmischung wird auf Luria-Nährflüssigkeit(LB)-Agarplatten, die 70 $\mu\text{g/ml}$ Ampicillin enthalten, ausplattiert.

[0211] Die Platten werden über Nacht bei 37°C inkubiert. Ampicillin-resistente Kolonien (6–12) werden herausgepickt und über Nacht in LB-Medium gezüchtet, welches 70 mg/ml Ampicillin enthält. Die Plasmid-DNA wird aus jeder Kultur durch Wizard-Minipreps (Promega Corporation, Madison WI) isoliert und auf das Vorhandensein des ungefähr 120 Basenpaare langen Insertionen durch einen Verdau mit EcoRI und HindIII und das Laufenlassen der Produkte des Verdau auf Agarose-Elektrophoresegelen durchmustert. Klone mit Insertionen wurden durch Standard-Didesoxy-Terminations-DNA-Sequenzierung überprüft. Der korrekte Klon wird als

pBSN1-1 (α_2) ([Fig. 52](#)) bezeichnet.

[0212] Oligos N1-3 (α_2) und N-1-4 (α_2) werden synthetisiert, gereinigt, aneinandergelagert, verlängert und in pBSKS⁺ unter Verfolgung des gleichen Verfahrens, das oben für die Oligos N1-1 (α_2) und N1-2 (α_2) angegeben wurde, kloniert. Das erhaltene Plasmid wird als pBSN1-2A bezeichnet. Um die Abschnitte des Kollagen-Gens von pBSN1-1 (α_2) (1 μ g) zusammenzuklonieren, wird selbiges für 2 Stunden bei 37°C BsrFI und HindIII verdaut. Der verdauten Vektor wird durch Agarosegelelektrophorese gereinigt. Plasmid pBSN1-2 (α_2) (3 μ g) wird für 2 Stunden bei 37°C mit BsrFI und HindIII verdaut und das Insert durch Agarosegelelektrophorese gereinigt. BsrFI/HindIII-verdautes pBSN1-1 wird mit dieser Insertion über Nacht bei 16°C mittels T4-DNA-Ligase ligiert. Eine Hälfte des Ligationsgemischs wird in DH5 α -Zellen transformiert und ein Zehntel der Transformationsmischung wird auf LB-Agarplatten, die 70 μ g/ml Ampicillin enthalten, ausplattiert. Nach einer Inkubation über Nacht bei 37°C werden Ampicillin-resistente Klone herausgepickt und auf das Vorhandensein eines DNA-Insertions wie oben beschrieben, durchmustert. Klone werden mittels Didesoxy-Terminationssequenzierung kontrolliert. Der korrekte Klon wird als pBSN1-2 (α_2) ([Fig. 53](#)) bezeichnet und das Kollagenfragment hat die Sequenz, die in [Fig. 54](#) (SEQ ID NO: 37) angegeben ist.

[0213] In ähnlicher Weise wird der Rest des Kollagen-Gens konstruiert, so daß die endgültige DNA-Sequenz die ist, die in den [Fig. 50A](#) bis [Fig. 50E](#) (SEQ ID NO: 31) angegeben ist.

B) Expression des Gens in E. coli:

[0214] Nach der Konstruktion des gesamten menschlichen Kollagen-Typ I (α_2)-Gens unter Codonverwendung, die für E. coli optimiert ist, wird das klonierte Gen in E. coli exprimiert. Das Plasmid (pHuCol(α_2)^{Ec}, ([Fig. 55](#)), welches das gesamte synthetische Kollagen-Gen ([Fig. 50A–Fig. 50E](#)) codiert, das hinter den Isopropyl- β -D-thiogalactopyranosid-(IPTG)-induzierbaren Tac-Promotor plziert ist und ebenfalls β -Lactamase codiert in Escherichia coli-Stamm DH5 α (supE44 Δ lacU169 (ϕ 80lacZ Δ M15) hsdR17 recA1 endA1 gyrA96 thi-1 relA1) durch Standard-Wärmeschocktransformation transformiert. Die Transformationskulturen werden auf Luria-Nährflüssigkeit (LB), die 100 μ g/ml Ampicillin enthalten, ausplattiert und nach dem Wachstum über Nacht wird eine einzige Ampicillin-resistente Kolonie verwendet, um 10 ml LB zu inokulieren, welches 100 μ g/ml Ampicillin enthält und nach Über-Nacht-Wachstum wird eine einzige Ampicillin-resistente Kolonie verwendet, um 10 ml LB, das 100 μ g/ml Ampicillin enthält, zu inokulieren. Nach dem Wachstum für 10 bis 16 Stunden unter Rühren (225 Upm) bei 37°C wird diese Kultur verwendet, um 1 l LB, welche 100 μ g/ml Ampicillin enthält, in einer 1,5 l-Schüttelflasche zu inokulieren. Nach dem Wachstum bei 37°C, 225 Upm, bei 2 Stunden nach der Inokulierung, ist die optische Dichte bei 600 nm ungefähr 0,5 OD/ml. IPTG wird zu 1 mM zugegeben und die Kultur wird für zusätzliche 5 bis 10 Stunden wachsen gelassen. Die Zellen werden durch Zentrifugieren (5000 Upm, 10 Minuten) geerntet und durch mechanisches Aufschließen lysiert. Rekombinantes menschliche Kollagen wird durch Ammoniumsulfatfraktionierung und Säulenchromatographie gereinigt. Die Ausbeute beträgt typischerweise 15 bis 25 mg/l Kultur

Beispiel 14A

Alternative Konstruktion und Expression des menschlichen Kollagen-Typ I (α_2)-Gens mit optimierter E. coli-Codonverwendung in E. coli

A) Konstruktion des Gens:

[0215] Die Nukleotidsequenz des helikalen Bereich des menschlichen Kollagen-Typ I (α_2)-Gens, das von 112 Aminosäuren des Amino-terminalen extrahelikalen und 12 Aminosäuren des C-terminalen extrahelikalen Bereichs flankiert ist, wird in den [Fig. 49A](#) bis [Fig. 49E](#) (SEQ ID NO: 29) gezeigt. Eine tabellarische Auflistung der Codonhäufigkeit dieses Gens ist in Tabelle III angegeben. Die Gensequenz, die in den [Fig. 49A](#) bis [Fig. 49E](#) gezeigt ist, wurde zuerst geändert, um eine Neigung in Richtung der E. coli-Codons widerzuspiegeln. Ein Initiator-methionin wurde am 5'-Ende des Gens inseriert und eine TAAT-Stoppsequenz am 3'-Ende. Einzigartige Restriktionsschnittstellen wurden identifiziert oder an geeigneten Positionen in dem Gen erzeugt (ungefähr alle 150 Basenpaare). Das daraus resultierende Gen (HuCol(α_2)^{Ec}) hat die Codonverwendung, die in Tabelle IV wiedergegeben ist. Andere Sequenzen, die sich der Neigung der E. coli-Codons annähern, sind ebenfalls annehmbar.

[0216] Oligonukleotide wurden auf einem Beckman Oligo 1000 DNA-Synthesegerät synthetisiert, gespalten und mit wäßriger NH₄OH entschützt und mittels Elektrophorese in 7 M Harnstoff/12 Polyacrylamid-Gelen gereinigt. Gereinigte Oligos (32,5 pmol) wurden in 20 μ l Ligationspuffer (Boehringer Mannheim, Kat. Nr. 1635379)

und Erwärmen auf 95°C, gefolgt von einer langsamen Abkühlung auf 20°C über 45 Minuten aneinandergelagert. Die aneinandergelagerten Oligonukleotide wurden für 5 Minuten bei Raumtemperatur mit verdautem Vektor (1 µg) unter Verwendung von T4-DNA-Ligase (5 Einheiten) ligiert. Eine Hälfte des Transformationsgemischs wurde durch Wärmeschock in DH5α-Zellen transformiert und 100 µl der 1,0 ml-Transformationsmischung werden auf Luria-Nährflüssigkeit (LB)-Agarplatten ausplattiert, die 70 µg/ml Ampicillin enthalten. Die Platten wurden über Nacht bei 37°C inkubiert. Die Ampicillin-resistenten Kolonien (6–12) wurden herausgepickt und über Nacht in LB-Medium gezüchtet, das 70 µg/ml Ampicillin enthält. Plasmid-DNA wurde aus jeder Kultur durch QIAprep Miniprep (Qiagen, Valencia, CA) isoliert und auf das Vorhandensein einer Insertion durch Verdau mit den flankierenden Restriktionsenzymen durchmustert und durch Laufenlassen der verdauten Produkte auf Agaroseelektrophoresegelels. Die Klone mit Insertionen wurden durch Standard-Didesoxy-Terminations-DNA-Sequenzierung überprüft. Um die Abschnitte des Kollagen-Gens zusammenzuklonieren, wurde ein Insert, das die flankierenden Bereiche des Gens abdeckt in den Vektor, welcher den benachbarten Gen-Anteil enthält, ligiert. Insertionen wurden aus Plasmiden isoliert und die Vektoren wurden durch Doppelverdau für 2 Stunden bei 37°C mit den geeigneten Restriktionsenzymen geschnitten. Der verdauten Vektor und das Insert wurden durch Agarosegelelektrophorese gereinigt. Insert und Vektor wurden 5 Minuten bei Raumtemperatur nach dem Verfahren in dem Rapid-DNA-Ligation-Kit (Boehringer Mannheim) ligiert. Eine Hälfte des Ligationsgemischs wurde in DH5α-Zellen transformiert und 1/10 des Transformationsgemischs wurde auf LB-Agarplatten, die 70 µg/ml Ampicillin enthalten, ausplattiert. Nach einer Inkubation über Nacht bei 37°C wurden Ampicillin-resistente Kolonien herausgepickt und auf das Vorhandensein von Insert-DNAs, wie oben beschrieben, durchmustert. Klone wurden durch Didesoxy-Terminationssequenzierung überprüft.

[0217] In ähnlicher Weise wurde der Rest des Kollagen-Gens hergestellt, so daß die endgültige DNA-Sequenz die ist, die in den [Fig. 50A](#) bis [Fig. 50E](#) (SEQ ID NO: 31) ist.

B) Expression des Gens in E. coli:

[0218] Nach der Konstruktion des gesamten menschlichen Kollagen-Typ I (α_2)-Gens unter Codonverwendung, die für E. coli optimiert ist, wird das klonierte Gen in E. coli exprimiert. Das Plasmid (pHuCol(α_2)^{Ec}, ([Fig. 55](#)), welches das gesamte synthetische Kollagen-Gen ([Fig. 50A–Fig. 50E](#)) codiert, das hinter dem Isopropyl-β-D-thiogalactopyranosid-(IPTG)-induzierbaren Tac-Promotor plziert ist und ebenfalls β-Lactamase codiert in den Escherichia coli-Stamm DH5α (supE44 ΔlacU169 (φ80lacZ ΔM15) hsdR17 recA1 endA1 gyrA96 thi-1 relA1) durch Standard-Wärmeschocktransformation transformiert. Die Transformationskulturen werden auf Luria-Nährflüssigkeit (LB), die 100 µg/ml Ampicillin enthalten, ausplattiert und nach dem Wachstum über Nacht wird eine einzige Ampicillin-resistente Kolonie verwendet, um 10 ml LB zu inokulieren, welches 100 µg/ml Ampicillin enthält und nach Über-Nacht-Wachstum wird eine einzige Ampicillin-resistente Kolonie verwendet, um 10 ml LB, das 100 µg/ml Ampicillin enthält, zu inokulieren. Nach dem Wachstum für 10 bis 16 Stunden unter Rühren (225 Upm) bei 37°C wird diese Kultur verwendet, um 1 l LB, welche 100 µg/ml Ampicillin enthält, in einer 1,5 l-Schüttelflasche zu inokulieren. Nach dem Wachstum bei 37°C, 225 Upm, bei 2 Stunden nach der Inokulierung, ist die optische Dichte bei 600 nm ungefähr 0,5 OD/ml. IPTG wird zu 1 mM zugegeben und die Kultur wird für zusätzliche 5 bis 10 Stunden wachsen gelassen. Die Zellen werden durch Zentrifugieren (5000 Upm, 10 Minuten) geerntet und durch mechanisches Aufschließen lysiert. Rekombinantes menschliche Kollagen wird durch Ammoniumsulfatfraktionierung und Säulenchromatographie gereinigt. Die Ausbeute beträgt typischerweise 15 bis 25 mg/l Kultur

Beispiel 15

Expression von menschlichem Kollagen-Typ I (α_2) mit optimierter E. coli-Verwendung in E. coli

[0219] Ein Plasmid (pN1-2, [Fig. 56](#)), welches das Gen für die Amino-terminalen 80 Aminosäuren des menschlichen Kollagen-Typ I (α_2) (SEQ ID NO: 31, [Fig. 54](#)) mit optimierter E. coli-Codonverwendung codiert, ist hinter dem Isopropyl-β-D-thiogalactopyranosid-(IPTG)-induzierbaren Tac-Promotor plziert und codiert ebenfalls β-Lactamase und wird in Escherichia coli-Stamm DH5α (supE44 ΔlacU169 (φ80lacZ ΔM15) hsdR17 recA1 endA1 gyrA96 thi-1 relA1) durch Standard-Wärmeschocktransformation transformiert. Die Transformationskulturen werden auf Luria-Nährflüssigkeit (LB), die 100 µg/ml Ampicillin enthalten, ausplattiert und nach dem Wachstum über Nacht wird eine einzige Ampicillin-resistente Kolonie verwendet, um 10 ml LB zu inokulieren, welches 100 µg/ml Ampicillin enthält und nach Über-Nacht-Wachstum wird eine einzige Ampicillin-resistente Kolonie verwendet, um 10 ml LB, das 100 µg/ml Ampicillin enthält, zu inokulieren. Nach dem Wachstum für 10 bis 16 Stunden unter Rühren (225 Upm) bei 37°C wird diese Kultur verwendet, um 1 l LB, welche 100 µg/ml Ampicillin enthält, in einer 1,5 l-Schüttelflasche zu inokulieren. Nach dem Wachstum bei 37°C, 225 Upm, bei 2 Stunden nach der Inokulierung, ist die optische Dichte bei 600 nm ungefähr 0,5 OD/ml. IPTG wird

zu 1 mM zugegeben und die Kultur wird für zusätzliche 5 bis 10 Stunden wachsen gelassen. Die Zellen werden durch Zentrifugieren (5000 Upm, 10 Minuten) geerntet und durch mechanisches Aufschließen lysiert. Rekombinantes menschliche Kollagen wird durch Ammoniumsulfatfraktionierung und Säulenchromatographie gereinigt. Die Ausbeute beträgt typischerweise 15 bis 25 mg/l Kultur

Beispiel 16

Hydroxyprolin-Einbau in Proteine in *E. coli* unter Hungerbedingungen für Prolin

[0220] 7 Plasmide pGEX-4T.1 ([Fig. 73](#)), pTrc-TGF ([Fig. 74](#)), pMal-C2 ([Fig. 1](#)), pTrc-FN ([Fig. 75](#)), pTrc-Fn-TGF ([Fig. 76](#)), pTrc-FN-Bmp ([Fig. 77](#)) und pGEX-HuColl^{Ec}, die jeweils getrennt Gene enthalten, die die folgenden Proteine codieren: Glutathion-S-Transferase (GST), das reife menschliche TGF- β_1 -Polypeptid (TGF- β_1), Mannose-bindendes Protein (MBP), ein 70 kDa-Fragment von menschlichem Fibronectin (FN), eine Fusion aus FN und TGF- β_1 (FN-TGF- β_1), eine Fusion von FN und menschlichem Knochenmorphogenese-Protein 2A (FN-BMP-2A) und eine Fusion aus GST und Kollagen (GST-Coll), wurden jeweils individuell verwendet, um den Prolin-auxotrophen *E. coli*-Stamm JM109 (F⁻) zu transformieren. Transformationskulturen wurden auf LB-Agar, der 100 µg/ml Ampicillin enthält, ausplattiert. Nach einer Inkubation über Nacht bei 37°C wurde eine einzige Kolonie von einer frischen Transformationsplatte verwendet, um 5 ml LB-Medium zu inokulieren, das 400 mg Ampicillin enthält. Nach Über-Nacht-Wachstum bei 37°C wurde diese Kultur zentrifugiert, der Überstand verworfen und die Zellpellets wurden zweimal mit 5 ml M9-Medium (1 × M9-Salze, 0,5% Glucose, 1 mM MgCl₂, 0,01% Thiamin, 200 µg/ml Glycin, 200 µg/ml Alanin 100 µg/ml der anderen Aminosäuren ausgenommen Prolin und 400 µg/ml Ampicillin) gewaschen. Die Zellen wurden schließlich in 500 ml M9-Medium resuspendiert. Nach einer Inkubation unter Schütteln bei 37°C für 30 Minuten wurde trans-4-Hydroxyprolin zu 40 mM zugegeben, NaCl zu 0,5 M und Isopropyl- β -D-thiogalactopyranosid zu 1,5 mM. In einigen der Kulturen wurde eine dieser Zugaben nicht durchgeführt, wie sie an den Markierungen für die Spuren der Gele angegeben ist. Nach der Zugabe wurde die Inkubation unter Schütteln bei 37°C fortgesetzt. Nach 4 Stunden wurden die Kulturen zentrifugiert, die Überstände verworfen und die Zellpellets in SDS-PAGE-Probenpuffer resuspendiert (300 mM Tris (pH 6,8)/0,5% SDS/10% Glycerin/0,4 M β -Mercaptoethanol/0,2% Bromphenolblau) auf 15 OD 600 nm AU/ml, in ein kochendes Wasserbad für 5 Minuten gestellt und in denaturierenden Polyacrylamidgelen elektrophoretisch aufgetrennt. Die Proteine in den Gelen wurden durch Färben mit Coomassie Blue R250 visualisiert. Die Ergebnisse der Gele werden in den Scans, die in [Fig. 57](#) bis [Fig. 59](#) gezeigt sind, wiedergegeben. Die Scans beziehen sich auf GST, TGF- β_1 , MBP, FN, FN-TGF- β_1 und FN-BMP-2A ([Fig. 57](#) und [Fig. 58](#)), welche drei Spuren zeigen, die jedes Peptid betreffen, d. h. eine Spur, die +NaCl/+Hyp anzeigt, worin NaCl (hyperosmotisch) und trans-4-Hydroxyprolin vorhanden sind; eine Spur, die -NaCl angibt, worin trans-4-Hydroxyprolin vorhanden ist, aber NaCl nicht; und eine Spur, die -Hyp anzeigt, welche +NaCl ist, aber ohne trans-4-Hydroxyprolin. Die Sternchen auf den Scans markieren die Proteinbanden, die dem entsprechenden Zielprotein entsprechen. Die Fälle, in denen Zielprotein exprimiert war, schließen alle +NaCl ein in Verbindung mit +Hyp, wodurch die +NaCl- und +Hyp-Abhängigkeit gezeigt wird.

[0221] Der Scan, der in [Fig. 59](#) gezeigt ist, welcher GST-Kollagen betrifft, zeigt vier Spuren, die GST-Coll betreffen, d. h. eine Spur die +Hyp/+NaCl/-IPTG zeigt, wobei trans-4-Hydroxyprolin und NaCl vorhanden sind, aber IPTG (der Protein-Expressionsinduktor) nicht, und da es dort keinen Induktor gibt, gibt es keine Zielprotein-Bande; eine Spur, die +NaCl/+IPTG/-Hyp angibt, worin NaCl und IPTG vorhanden sind, aber trans-4-Hydroxyprolin nicht, und da trans-4-Hydroxyprolin nicht vorhanden ist, keine Zielprotein-Bande vorhanden ist; eine Spur, die +NaCl/+Pro/+IPTG anzeigt, worin NaCl, Prolin und IPTG vorhanden sind, aber da das Zielprotein nicht stabil ist, wenn es Prolin enthält, gibt es keine Zielprotein-Bande; und eine Spur, die als +IPTG/+NaCl/+Hyp bezeichnet wird, worin IPTG, NaCl und trans-4-Hydroxyprolin vorhanden sind, und da dieses Protein durch die Anwesenheit von trans-4-Hydroxyprolin stabilisiert wird, ist eine durch einen Stern markierte Proteinbande sichtbar.

Beispiel 17

Hydroxyprolin-Einbau in ein Kollagen-artiges Peptid in *E. coli*

[0222] Ein Plasmid (pGST-CM4, [Fig. 60](#)), das das Gen für Kollagen-Mimetikum 4 (CM 4, [Fig. 61](#)) (SEQ ID NO: 39) enthält, welches genetisch mit dem 3'-Ende des Gens für *S. japonicum*-Glutathion-S-Transferase verbunden ist, wurde verwendet, um durch Elektroporierung den Prolin-auxotrophen *E. coli*-Stamm JM109 (F⁻) zu transformieren. Transformationskulturen wurden auf LB-Agar ausplattiert, der 100 µg/ml Ampicillin enthält. Nach einer Inkubation über Nacht bei 37°C wurde eine einzelne Kolonie einer frischen Transformationsplatte verwendet, um 5 ml LB-Medium zu inokulieren, welche 100 µg/ml Ampicillin enthält. Nach dem Wachstum über

Nacht bei 37°C wurden 500 µl dieser Kultur zentrifugiert, der Überstand verworfen, und das Zellpellet wurde einmal mit 500 µl M9-Medium (1 × M9-Salze, 0,5% Glucose, 1 mM MgCl₂, 0,1% Thiamin, 20 µg/ml Glycin, 200 µg/ml Alanin, 100 µg/ml der anderen Aminosäuren ausgenommen Prolin und 400 µg/ml Ampicillin) gewaschen. Die Zellen wurden schließlich in 5 ml M9-Medium suspendiert, das 10 µg/ml Prolin enthält und 2 ml hiervon wurde verwendet, um 30 ml M9-Medium zu inokulieren, das 10 µg/ml Prolin enthält. Nach der Inkubation unter Schütteln bei 37°C für 8 Stunden wurde die Kultur zentrifugiert und das Zellpellet wurde einmal mit M9-Medium gewaschen, das 5 µg/ml Prolin enthält. Das Pellet wurde in 15 ml M9-Medium resuspendiert, das 5 µg/ml Prolin enthält und diese Kultur wurde verwendet, um 1 l M9-Medium zu inokulieren, das 5 µg/ml Prolin enthält.

[0223] Diese Kultur wurde für 18 Stunden bei 37°C nach hungernd Prolin gezüchtet. Zu diesem Zeitpunkt wurde die Kultur zentrifugiert, die Zellen wurden einmal mit M9-Medium (ohne Prolin) gewaschen und die Zellen wurden in 1 l M9-Medium resuspendiert, das 80 mM Hydroxyprolin, 0,5 M NaCl und 1,5 mM Isopropyl-β-D-thiogalactopyranosid enthält, resuspendiert. Die Inkubation wurde bei 37°C unter Schütteln für 22 Stunden fortgesetzt. Die Kulturen wurden zentrifugiert und die Zellpellets wurden bei -20°C gelagert bis sie weiterverarbeitet wurden.

Beispiel 18

Prolin-Einbau in ein Kollagen-artiges Peptid in E. coli

[0224] Ein Plasmid (pGST-CM4, [Fig. 60](#)), das das Gen für Kollagen-Mimetikum 4 (CM 4, [Fig. 61](#)) (SEQ ID NO: 39) enthält, welches genetisch mit dem 3'-Ende des Gens für S. japonicum-Glutathion-S-Transferase verbunden ist, wurde verwendet, um durch Elektroporierung den Prolin-auxotrophen E. coli-Stamm JM109 (F⁻) zu transformieren. Transformationskulturen wurden auf LB-Agar ausplattiert, der 100 µg/ml Ampicillin enthält. Nach einer Inkubation über Nacht bei 37°C wurde eine einzelne Kolonie einer frischen Transformationsplatte verwendet, um 5 ml LB-Medium zu inokulieren, welche 100 µg/ml Ampicillin enthält. Nach dem Wachstum über Nacht bei 37°C wurden 500 µl dieser Kultur zentrifugiert, der Überstand verworfen, und das Zellpellet wurde einmal mit 500 µl M9-Medium (1 × M9-Salze, 0,5% Glucose, 1 mM MgCl₂, 0,1% Thiamin, 20 µg/ml Glycin, 200 µg/ml Alanin, 100 µg/ml der anderen Aminosäuren ausgenommen Prolin und 400 µg/ml Ampicillin) gewaschen. Die Zellen wurden schließlich in 5 ml M9-Medium suspendiert, das 10 µg/ml Prolin enthält und 2 ml hiervon wurden verwendet, um 30 ml M9-Medium zu inokulieren, das 10 µg/ml Prolin enthält. Nach der Inkubation unter Schütteln bei 37°C für 8 Stunden wurde die Kultur zentrifugiert und das Zellpellet wurde einmal mit M9-Medium gewaschen, das 5 µg/ml Prolin enthält. Das Pellet wurde in 15 ml M9-Medium resuspendiert, das 5 µg/ml Prolin enthält und diese Kultur wurde verwendet, um 1 l M9-Medium zu inokulieren, das 5 µg/ml Prolin enthält. Diese Kultur wurde für 18 Stunden bei 37°C nach Prolin hungernd gezüchtet. Zu diesem Zeitpunkt wurde die Kultur zentrifugiert, die Zellen wurden einmal mit M9-Medium (ohne Prolin) gewaschen und die Zellen wurden in 1 l M9-Medium resuspendiert, das 80 mM Hydroxyprolin, 0,5 M NaCl und 1,5 mM Isopropyl-β-D-thiogalactopyranosid enthält, resuspendiert. Die Inkubation wurde bei 37°C unter Schütteln für 22 Stunden fortgesetzt. Die Kulturen wurden zentrifugiert und die Zellpellets wurden bei -20°C gelagert bis sie weiterverarbeitet wurden.

Beispiel 19

Reinigung von Hydroxyprolin-haltigem Kollagen-artigen Peptid aus E. coli

[0225] Das Zellpellet aus 1 l Fermentationskultur, welche wie in Beispiel 17 oben beschrieben hergestellt wurde, wurde in 20 ml Dulbecco's Phosphat-gepufferter Salzlösung (pH 7,1) (PBS), welche 1 mM EDTA, 100 µM PMSF, 0,5 µg/ml E64 und 0,7 µg/ml Pepstatin (Resuspensionspuffer) enthält, resuspendiert. Die Zellen wurden lysiert, indem sie zweimal durch eine French-Press passiert wurden. Nach der Lyse wurde die Suspension für 30 Minuten bei 30 000 xg zentrifugiert. Der Überstand wurde verworfen und das Pellet wurde einmal mit 5 ml Resuspensionspuffer, welcher 1 M Harnstoff enthält und mit 0,5% Triton X100 gewaschen, gefolgt von einem Waschschritt mit 7 ml mit Resuspensionspuffer ohne Harnstoff oder Triton X100. Das Pellet wurde schließlich in 5 ml 6 M Guanidinhydrochlorids in Dulbeccos-Phosphat-gepufferter Salzlösung (pH 7,1) resuspendiert, die 1 mM EDTA und 2 mM β-Mercaptoethanol enthält und für 3 × 60 Sekunden auf Eis mit Ultraschall behandelt (Mikrotip, Stärke = 3,5, Heat Systems XL-2020 model sonicator). Die beschallte Suspension wurde bei 4°C für 18 Stunden inkubiert und dann bei 14 000 Upm in einer Mikrozentrifuge zentrifugiert. Der Überstand (6 ml) wurde gegen 4 × 4 l destillierten Wassers bei 4°C dialysiert (10 000 MWCO). Die Inhalte der Dialyseröhrchen wurden in eine 150 ml-Rundbodenflasche überführt und bis zur Trockenheit lyophilisiert. Der Rest (~30 mg) wurde in 3 ml 70%iger Ameisensäure gelöst und 40 mg Cyanogenbromid wurde zugegeben. Die Flasche wurde einmal mit Stickstoff ausgespült, geleert und es wurde ihr ermöglicht für 18 Stunden bei Raumtemperatur zu Rühren. Der Inhalt der Flasche wurde im Vakuum bei Raumtemperatur bis zur Trockenheit gebracht,

der Rest wurde in 5 ml destilliertem Wasser resuspendiert und wieder bis zur Trockenheit verdampft. Dies wurde zweimal wiederholt. Der Rest wurde schließlich in 2 ml 0,2%iger Trifluoressigsäure (TFA) gelöst. Das Trifluoressigsäure-lösliche Material wurde in 100 µl-Aliquots auf eine Roros R2-Säule (4,6 mm × 100 mm) aufgetragen, welche bei 5 ml/min mit einem Ausgangspuffer von 98% 0,1% Trifluoressigsäure in Wasser/2% 0,1% TFA in Acetonitril läuft. Das Hydroxyprolin-haltige Protein wurde mit einem Gradienten aus 2% zu 0,1%igem TFA/Acetonitril bis 40% zu 0,1% TFA/Acetonitril über 25 Säulenvolumina eluiert ([Fig. 62A](#)). Das Kollagen-Mimetikum eluierte zwischen 18 und 23%/0,1% TFA/Acetonitril. [Fig. 62A](#) ist ein Chromatogramm der Eluierung des Hydroxyprolin-haltigen CM4 von einer Poros RP2-Säule (erhältlich von Perseptive Biosystems, Framingham, MA). Der Pfeil zeigt den Peak an, welcher das Hydroxyprolin-haltige CM4 enthält. Die Fraktionen wurden mittels SDS-PAGE untersucht und die Kollagen-Mimetikum-haltigen Fraktionen wurden gepoolt und lyophilisiert. Lyophilisiertes Material wurde bei -20°C gelagert.

Beispiel 20

Reinigung von Prolin-haltigem Kollagen-artigen Peptid aus E. coli

[0226] Das Zellpellet aus 500 ml Fermentationskultur, welches wie oben in Beispiel 18 beschrieben hergestellt wurde, wurde in 20 ml Dulbeccos-Phosphat-gepufferter Salzlösung (pH 7,1) (PBS) resuspendiert, welche 10 mM EDTA, 100 µM PMSF, 0,5 µg/ml E64 und 0,06 µg/ml Aprotinin enthält. Lysozym (2 mg) wurde zu der Suspension zugegeben, die bei 4°C für 60 Minuten inkubiert wurde. Die Suspension wurde für 5 × 60 Sekunden beschallt (Mikrotip, Power = 3,5, Heat Systems XL-2020 model Sonicator). Die beschallte Suspension wurde bei 20 000 xg für 15 Minuten zentrifugiert. Der Überstand wurde auf 1% Triton X100 eingestellt und für 30 Minuten bei Raumtemperatur mit 7 ml Glutathionsepharose 4B prääquilibriert in PBS inkubiert. Die Suspension wurde für 3 Minuten bei 500 Upm zentrifugiert. Der Überstand wurde dekantiert und das Harz wurde dreimal mit 8 ml PBS gewaschen. Gebundene Proteine wurden mit 3 Aliquots (2 ml jeweils, 10 Minuten leichtes Schwenken bei Raumtemperatur) aus 10 mM Glutathion 50 mM Tris (pH 8,0) eluiert. Die Eluate wurden kombiniert und dialysiert (10 000 MWCO), gegen 3 × 4 l destillierten Wassers bei 4°C. Die Inhalte der Dialyseröhrchen wurden in eine 150 ml-Rundbodenflasche überführt und bis zur Trockenheit lyophilisiert. Der Rest wurde in 3 ml 70%iger Ameisensäure gelöst und 4 mg Cyanogenbromid wurde zugegeben. Die Flasche wurde einmal mit Stickstoff ausgespült, geleert und es wurde ihr ermöglicht für 18 Stunden bei Raumtemperatur zu Rühren. Der Inhalt der Flasche wurde im Vakuum bei Raumtemperatur bis zur Trockenheit gebracht, der Rest wurde in 5 ml destilliertem Wasser resuspendiert und wieder bis zur Trockenheit verdampft. Dies wurde zweimal wiederholt. Der Rest wurde schließlich in 2 ml 0,2%iger Trifluoressigsäure (TFA) gelöst.

[0227] Das Trifluoressigsäure-lösliche Material wurde in 100 µl-Aliquots auf eine Roros R2-Säule (4,6 mm × 100 mm) aufgetragen, welche bei 5 ml/min mit einem Ausgangspuffer von 98% 0,1% Trifluoressigsäure in Wasser/2% 0,1% TFA in Acetonitril läuft. Das Hydroxyprolin-haltige Protein wurde mit einem Gradienten aus 2% 0,1%igem TFA/Acetonitril zu 40% 0,1% TFA/Acetonitril über 25 Säulenvolumina eluiert ([Fig. 62A](#)). Das Kollagen-Mimetikum eluierte zwischen 18 und 23% 0,1% TFA/Acetonitril. [Fig. 62A](#) ist ein Chromatogramm der Eluierung des Hydroxyprolin-haltigen CM4 von einer Poros RP2-Säule (erhältlich von Perseptive Biosystems, Framingham, MA). Der Pfeil zeigt den Peak an, welcher das Hydroxyprolin-haltige CM4 enthält. Die Fraktionen wurden aus SDS-PAGE untersucht und die Kollagen-Mimetikum-haltigen Fraktionen wurden gepoolt und lyophilisiert. Lyophilisiertes Material wurde bei -20°C gelagert.

Beispiel 21

Aminosäureanalyse des Hydroxyprolin-haltigen Kollagenmimetikums und des Protein-haltigen Kollagenmimetikums

[0228] Ungefähr 30 µg des gereinigten Hydroxyprolin-haltigen Kollagenmimetikums und des Prolin-haltigen Kollagenmimetikums, die hergestellt wurden wie in den Beispielen 19 bzw. 20 beschrieben, wurden in 250 µl 6 N Salzsäure in Glasampullen gelöst. Die Ampullen wurden zweimal mit Stickstoff gespült, unter Vakuum versiegelt und bei 110°C für 23 Stunden inkubiert. Nach der Hydrolyse wurden die Proben aus den Ampullen entfernt und im Vakuum bis zur Trockenheit gebracht. Die Proben wurden in 15 µl 0,1 N Salzsäure gelöst und eine Aminosäureanalyse auf einem Hewlett Packard AminoQuant 1090 Aminosäure-Analysegerät unter Verwendung von Standard-OPA und FMOC-Derivatisierungsschemie unterworfen. Beispiele der Ergebnisse der Aminosäureanalyse, die den Bereich der Chromatogramme darstellen, wo die sekundären Aminosäuren (Prolin und Hydroxyprolin) eluieren, sind in den [Fig. 63A](#) bis [Fig. 63D](#) gezeigt. Diese Figuren zeigen auch Chromatogramme von Prolin- und Hydroxyprolin-Aminosäurestandards. Genauer gesagt zeigt [Fig. 63A](#) ein Chromatogramm eines Prolin-Aminosäurestandards (250 pmol). * zeigt einen kontaminierenden Peak an; [Fig. 63B](#) zeigt

ein Chromatogramm eine Hydroxyprolin-Aminosäurestandards (250 Pool). * zeigt einen kontaminierenden Peak an. [Fig. 63C](#) stellt ein Aminosäure-Analysechromatogramm der Hydrolyse von Prolin-haltigem CM4 dar. Nur der Bereich des Chromatogramms, in dem Prolin und Hydroxyprolin eluieren, ist gezeigt. * zeigt einen kontaminierenden Peak an. [Fig. 63D](#) stellt ein Aminosäure-Analysechromatogramm der Hydrolyse von Hydroxyprolin-haltigem CM4 dar. Nur der Bereich des Chromatogramms, wo Prolin und Hydroxyprolin eluieren, ist gezeigt. Das * zeigt einen kontaminierenden Peak an.

Beispiel 22

Bestimmung der Prolin-Hungerbedingungen in *E. coli* (Stamm JM109 (F⁻))

[0229] Ein Plasmid (pGST-CM4, [Fig. 60](#)), das das Gen für Kollagen-Mimetikum 4 (CM4, [Fig. 61](#)) enthält, das genetisch mit dem 3'-Ende des Gens für *S. japonicum*-Glutathion-S-Transferase verbunden ist, wurde verwendet, um durch Elektroporation den Prolin-auxotrophen *E. coli*-Stamm JM106 (F⁻) zu transformieren. Die Transformationskulturen wurden auf LB-Agar ausplattiert, welcher 100 µg/ml Ampicillin enthält. Nach einer Inkubation über Nacht bei 37°C wird eine einzelne Kolonie von einer frischen Transformationsplatte verwendet, um 2 ml M9-Medium zu inokulieren (1 × M9-Salze, 0,5% Glucose, 1 mM MgCl₂, 0,01% Thiamin, 200 µg/ml Glycin, 200 µg/ml Alanin, 100 µg/ml der anderen Aminosäuren ausgenommen Prolin und 200 µg/ml Carbenicillin) und welches 20 µg/ml Prolin enthält. Nach dem Wachstum bei 37°C unter Schütteln für 8 Stunden wurden 1,5 ml verwendet, um 27 ml M9-Medium zu inokulieren, das 45 µg/ml Prolin enthält. Nach der Inkubation bei 37°C unter Schütteln für 7 Stunden wurde die Kultur zentrifugiert, das Zellpellet wurde mit 7 ml M9-Medium ohne Prolin gewaschen und schließlich in 17 ml M9-Medium ohne Prolin resuspendiert. Diese Kultur wurde verwendet, um vier 35 ml-Kulturen M9-Medium zu inokulieren, welche 4 µg/ml Prolin enthalten bei einer OD₆₀₀ von 0,028. Kulturen wurden unter Schütteln bei 37°C inkubiert und die OD₆₀₀ wurde überwacht. Nach 13,5 Stunden Wachstum hatte die OD₆₀₀ ein Plateau erreicht. Zu diesem Zeitpunkt wurde eine Kultur mit 15 µg/ml Prolin supplementiert, eine mit 15 µg/ml Hydroxyprolin, eine mit 15 µg/ml aller Aminosäuren ausgenommen Prolin und Hydroxyprolin, und eine Kultur mit gar nichts. Die Inkubation wurde fortgesetzt und die OD₆₀₀ für insgesamt 24 Stunden überwacht. [Fig. 64](#) ist ein Graph der OD₆₀₀ gegenüber der Zeit für Kulturen von JM109 (F⁻), welche bis zum Plateau gezüchtet wurden und dann mit verschiedenen Aminosäuren supplementiert wurden. Der Punkt, an dem die Kulturen ergänzt wurden, ist mit einem Pfeil angegeben. Das Hungern nach Prolin ist erkennbar, da nur die Kultur, die mit Prolin supplementiert wurde, über das Plateau hinaus wuchs.

Beispiel 23

Hydroxyprolin-Einbau in Typ I (α₁)-Kollagen in *E. coli*

[0230] Ein Plasmid (pHuCol(α₁)^{Ec}, [Fig. 65](#)), welches das Gen für Typ I (α₁)-Kollagen mit optimierter *E. coli*-Codonverwendung ([Fig. 39A–Fig. 39E](#)) (SEQ ID NO: 19) unter Kontrolle des Tac-Promotors enthält und das Gen für Chloramphenicol-Resistenz enthält, wurde verwendet, um durch Elektroporation den Prolin-auxotrophen *E. coli*-Stamm JM109 (F⁻) zu transformieren. Die Transformationskulturen wurden auf LB-Agar ausplattiert, welcher 20 µg/ml Chloramphenicol enthält.

[0231] Nach der Inkubation über Nacht bei 37°C wurde eine einzelne Kolonie einer frischen Transformationsplatte verwendet, um 100 ml LB-Medium zu inokulieren, das 20 µg/ml Chloramphenicol enthält. Diese Kultur wurde bis zu einer OD₆₀₀ nm von 0,5 gezüchtet und 100 µl-Aliquots wurden in 1,5 ml-Röhrchen überführt. Die Röhrchen wurden bei -80°C gelagert. Zur Expression wurde ein Röhrchen auf Eis aufgetaut und verwendet, um 25 ml LB-Medium zu inokulieren, das 20 µg/ml Chloramphenicol enthält. Nach dem Wachstum über Nacht bei 37°C wurde ein 4 ml-Aliquot entfernt, zentrifugiert, und das Zellpellet wurde einmal mit 1 ml 2 × YT-Medium gewaschen, das 20 µg/ml Chloramphenicol enthält, und die gewaschenen Zellen wurden verwendet, um 1 l 2 × YT-Medium zu inokulieren, das 20 µg/ml Chloramphenicol enthält. Diese Kultur wurde bei 37°C bis zu einer OD₆₀₀ nm von 0,8 gezüchtet. Die Kultur wurde zentrifugiert und das Zellpellet wurde einmal mit 100 ml M9-Medium gewaschen (1 × M9-Salze, 0,5% Glucose, 1 mM MgCl₂, 0,01% Thiamin, 200 µg/ml Glycin, 200 µg/ml Alanin, 100 µg/ml der anderen Aminosäuren ausgenommen Prolin und 20 µg/ml Chloramphenicol). Die Zellen wurden in 910 ml M9-Medium (1 × M9-Salze, 0,5% Glucose, 1 mM MgCl₂, 0,01% Thiamin, 200 µg/ml Glycin, 200 µg/ml Alanin, 100 µg/ml der anderen Aminosäuren ausgenommen Prolin und 20 µg/ml Chloramphenicol) resuspendiert und für 30 Minuten bei 37°C wachsen gelassen. NaCl (80 ml, 5 M), Hydroxyprolin (7,5 ml, 2 M) und IPTG (500 µl 1 M) wurde zugegeben und das Wachstum wurde für 3 Stunden fortgesetzt. Die Zellen wurden durch Zentrifugieren geerntet und bei -20°C gelagert.

Beispiel 24

Hydroxyprolin-Einbau in Typ I (α_2)-Kollagen in *E. coli*

[0232] Ein Plasmid (pHuCol(α_2)^{Ec}, [Fig. 66](#)), welches das Gen für Typ I (α_2)-Kollagen mit optimierter *E. coli*-Codonverwendung ([Fig. 50A–Fig. 50E](#)) (SEQ ID NO: 31) unter Kontrolle des Tac-Promotors enthält, und das das Gen für Chloramphenicolresistenz enthält, wurde verwendet, um durch Elektroporation den Prolin-auxotrophen *E. coli*-Stamm JM109 (F⁻) zu transformieren. Die Transformationskulturen wurden auf LB-Agar ausplattiert, welcher 20 µg/ml Chloramphenicol enthält. Nach der Inkubation über Nacht bei 37°C wurde eine einzelne Kolonie einer frischen Transformationsplatte verwendet, um 100 ml LB-Medium zu inokulieren, das 20 µg/ml Chloramphenicol enthält. Diese Kultur wurde bis zu einer OD₆₀₀ nm von 0,5 gezüchtet und 100 µl-Aliquots wurden in 1,5 ml-Röhrchen überführt. Die Röhrchen wurden bei –80°C gelagert. Zur Expression wurde ein Röhrchen auf Eis aufgetaut und verwendet, um 25 ml LB-Medium zu inokulieren, das 20 µg/ml Chloramphenicol enthält. Nach dem Wachstum über Nacht bei 37°C wurde ein 4 ml-Aliquot entfernt, zentrifugiert, und das Zellpellet wurde einmal mit 1 ml 2 × YT-Medium gewaschen, das 20 µg/ml Chloramphenicol enthält, und die gewaschenen Zellen wurden verwendet, um 1 l 2 × YT-Medium zu inokulieren, das 20 µg/ml Chloramphenicol enthält. Diese Kultur wurde bei 37°C bis zu einer OD₆₀₀ nm von 0,8 gezüchtet. Die Kultur wurde zentrifugiert und das Zellpellet wurde einmal mit 100 ml M9-Medium gewaschen (1 × M9-Salze, 0,5% Glucose, 1 mM MgCl₂, 0,01% Thiamin, 200 µg/ml Glycin, 200 µg/ml Alanin, 100 µg/ml der anderen Aminosäuren ausgenommen Prolin und 20 µg/ml Chloramphenicol). Die Zellen wurden in 910 ml M9-Medium (1 × M9-Salze, 0,5% Glucose, 1 mM MgCl₂, 0,01% Thiamin, 200 µg/ml Glycin, 200 µg/ml Alanin, 100 µg/ml der anderen Aminosäuren ausgenommen Prolin und 20 µg/ml Chloramphenicol) resuspendiert und für 30 Minuten bei 37°C wachsen gelassen. NaCl (80 ml, 5 M), Hydroxyprolin (7,5 ml, 2 M) und IPTG (500 µl 1 M) wurde zugegeben und das Wachstum wurde für 3 Stunden fortgesetzt. Die Zellen wurden durch Zentrifugieren geerntet und bei –20°C gelagert.

Beispiel 25

Hydroxyprolin-Einbau in ein C-terminales Fragment vom Typ I (α_1)-Kollagen in *E. coli*

[0233] Ein Plasmid (pD4- α_1 , [Fig. 67](#)), welches das Gen für die Carboxy-terminalen 219 Aminosäuren des menschlichen Typ I (α_1)-Kollagens mit optimierter *E. coli*-Codonverwendung codiert, welches an das 3'-Ende des Gens für Glutathion-S-Transferase fusioniert ist und unter Kontrolle des Tac-Promotors ist, und das ein Gen für Ampicillin-Resistenz enthält, wurde verwendet, um durch Elektroporation den Prolin-auxotrophen *E. coli*-Stamm JM109 (F⁻) zu transformieren. Die Transformationskulturen wurden auf LB-Agar ausplattiert, welcher 100 µg/ml Ampicillin enthält. Nach der Inkubation über Nacht bei 37°C wurde eine einzelne Kolonie einer frischen Transformationsplatte verwendet, um 100 ml LB-Medium zu inokulieren, das 100 µg/ml Ampicillin enthält. Diese Kultur wurde bis zu einer OD₆₀₀ nm von 0,5 gezüchtet und 100 µl-Aliquots wurden in 1,5 ml-Röhrchen überführt. Die Röhrchen wurden bei –80°C gelagert. Zur Expression wurde ein Röhrchen auf Eis aufgetaut und verwendet, um 25 ml LB-Medium zu inokulieren, das 400 µg/ml Ampicillin enthält. Nach dem Wachstum über Nacht bei 37°C wurde ein 4 ml-Aliquot entfernt, zentrifugiert, und das Zellpellet wurde einmal mit 1 ml 2 × YT-Medium gewaschen, das 400 µg/ml Ampicillin enthält, und die gewaschenen Zellen wurden verwendet, um 1 l 2 × YT-Medium zu inokulieren, das 400 µg/ml Ampicillin enthält. Diese Kultur wurde bei 37°C bis zu einer OD₆₀₀ nm von 0,8 gezüchtet. Die Kultur wurde zentrifugiert und das Zellpellet wurde einmal mit 100 ml M9-Medium gewaschen (1 × M9-Salze, 0,5% Glucose, 1 mM MgCl₂, 0,01% Thiamin, 200 µg/ml Glycin, 200 µg/ml Alanin, 100 µg/ml der anderen Aminosäuren ausgenommen Prolin und 400 µg/ml Ampicillin). Die Zellen wurden in 910 ml M9-Medium (1 × M9-Salze, 0,5% Glucose, 1 mM MgCl₂, 0,01% Thiamin, 200 µg/ml Glycin, 200 µg/ml Alanin, 100 µg/ml der anderen Aminosäuren ausgenommen Prolin und 400 µg/ml Ampicillin) resuspendiert und für 30 Minuten bei 37°C wachsen gelassen. NaCl (80 ml, 5 M), Hydroxyprolin (7,5 ml, 2 M) und IPTG (500 µl 1 M) wurde zugegeben und das Wachstum wurde für 3 Stunden fortgesetzt. Die Zellen wurden durch Zentrifugieren geerntet und bei –20°C gelagert.

Beispiel 26

Hydroxyprolin-Einbau in ein C-terminales Fragment von Typ I (α_1)-Kollagen in *E. coli*

[0234] Ein Plasmid (pD4- α_2 , [Fig. 68](#)), welches das Gen für die Carboxy-terminalen 219 Aminosäuren des menschlichen Typ I (α_2)-Kollagens mit optimierter *E. coli*-Codonverwendung codiert, welches an das 3'-Ende des Gens für Glutathion-S-Transferase fusioniert ist und unter Kontrolle des Tac-Promotors ist, und das das Gen für Ampicillin-Resistenz enthält, wurde verwendet, um durch Elektroporation den Prolin-auxotrophen *E.*

coli-Stamm JM109 (F⁻) zu transformieren. Die Transformationskulturen wurden auf LB-Agar ausplattiert, welcher 100 µg/ml Ampicillin enthält. Nach der Inkubation über Nacht bei 37°C wurde eine einzelne Kolonie einer frischen Transformationsplatte verwendet, um 100 ml LB-Medium zu inokulieren, das 100 µg/ml Ampicillin enthält. Diese Kultur wurde bis zu einer OD_{600 nm} von 0,5 gezüchtet und 100 µl-Aliquots wurden in 1,5 ml-Röhrchen überführt. Die Röhrchen wurden bei -80°C gelagert. Zur Expression wurde ein Röhrchen auf Eis aufgetaut und verwendet, um 25 ml LB-Medium zu inokulieren, das 400 µg/ml Ampicillin enthält. Nach dem Wachstum über Nacht bei 37°C wurde ein 4 ml-Aliquot entfernt, zentrifugiert, und das Zellpellet wurde einmal mit 1 ml 2 × YT-Medium gewaschen, das 400 µg/ml Ampicillin enthält, und die gewaschenen Zellen wurden verwendet, um 1 l 2 × YT-Medium zu inokulieren, das 400 µg/ml Ampicillin enthält. Diese Kultur wurde bei 37°C bis zu einer OD_{600 nm} von 0,8 gezüchtet. Die Kultur wurde zentrifugiert und das Zellpellet wurde einmal mit 100 ml M9-Medium gewaschen (1 × M9-Salze, 0,5% Glucose, 1 mM MgCl₂, 0,01% Thiamin, 200 µg/ml Glycin, 200 µg/ml Alanin, 100 µg/ml der anderen Aminosäuren ausgenommen Prolin und 400 µg/ml Ampicillin). Die Zellen wurden in 910 ml M9-Medium (1 × M9-Salze, 0,5% Glucose, 1 mM MgCl₂, 0,01% Thiamin, 200 µg/ml Glycin, 200 µg/ml Alanin, 100 µg/ml der anderen Aminosäuren ausgenommen Prolin und 400 µg/ml Ampicillin) resuspendiert und für 30 Minuten bei 37°C wachsen gelassen. NaCl (80 ml, 5 M), Hydroxyprolin (7,5 ml, 2 M) und IPTG (500 µl 1 M) wurde zugegeben und das Wachstum wurde für 3 Stunden fortgesetzt. Die Zellen wurden durch Zentrifugieren geerntet und bei -20°C gelagert.

Beispiel 27

Reinigung von Hydroxyprolin-haltigem C-terminalen Fragment vom Typ I (α₁)-Kollagen

[0235] Zellmasse, die aus einer 1 l-Kultur geerntet wurde, welche wie in Beispiel 25 gezüchtet wurde, wurde in 30 ml Lysepuffer resuspendiert (2 M Harnstoff, 137 mM NaCl, 2,7 mM KCl, 4,3 mM Na₂HPO₄, 1,4 mM KH₂PO₄, 10 mM EDTA, 10 mM βME, 0,1% Triton X-100, pH 7,4) bei 4°C. Lysozym (Hühnereiweiß) wurde zu 100 µg/l zugegeben und die Lösung wurde bei 4°C für 30 Minuten inkubiert. Die Lösung wurde zweimal durch eine Zellaufschluß-Pressen (SLM Instruments, Rochester, NY) passiert und dann bei 30 000 xg für 30 Minuten zentrifugiert. Das Pellet wurde in 30 ml 50 mM Tris-HCl, pH 7,6, resuspendiert, bei 30 000 xg für 30 Minuten zentrifugiert, und das Pellet wurde in 25 ml Lösungspuffer (8 M Harnstoff, 137 mM NaCl, 2,7 mM KCl, 4,3 mM Na₂HPO₄, 1 mM KH₂PO₄, 5 mM βME) aufgelöst. Die Lösung wurde bei 30 000 xg für 30 Minuten zentrifugiert und der Überstand wurde gegen zwei Wechsel aus 4 l destillierten Wassers bei 4°C dialysiert. Nach der Dialyse wurde die gesamte Mischung lyophilisiert. Der lyophilisierte Feststoff wurde in 0,1 M HCl in einer Flasche unter Rühren gelöst. Nach der Zugabe eines 5-fachen Überschusses aus kristallinem BrCN wurde die Flasche geleert und mit Stickstoff gefüllt. Die Spaltung wurde für 24 Stunden ablaufen gelassen, worauf das Lösungsmittel im Vakuum entfernt wurde. Der Rest wurde in 0,1% Trifluoressigsäure (TFA) gelöst und mittels Umkehrphasen-HPLC unter Verwendung einer Vydac C4 RP-HPLC-Säule (10 × 250 mm, 5 µ, 300 Å) auf einem BioCad Sprint System (Perceptive Biosystems, Framingham, MA) gereinigt. Hydroxyprolin-haltiges D4-Protein wurde mit einem Gradienten auf 15 bis 40% Acetonitril/0,1% TFA über einen 45-minütigen Zeitraum eluiert. Protein D4-αα eluierte bei 26% Acetonitril/0,1% TFA.

Beispiel 28

Reinigung eines Hydroxyprolin-haltigen C-terminalen Fragments vom Typ I (α₂)-Kollagen

[0236] Zellmasse, die aus 1 l Kultur stammte, welche wie in Beispiel 26 gezüchtet wurde, wurde in 30 ml Lysepuffer resuspendiert (2 M Harnstoff, 137 mM NaCl, 2,7 mM KCl, 4,3 mM Na₂HPO₄, 1,4 mM KH₂PO₄, 10 mM EDTA, 10 mM βME, 0,1% Triton X-100, pH 7,4) bei 4°C. Lysozym (Hühnereiweiß) wurde zu 100 µg/l zugegeben und die Lösung wurde bei 4°C für 30 Minuten inkubiert. Die Lösung wurde zweimal durch eine Zellaufschluß-Pressen (SLM Instruments, Rochester, NY) passiert und dann bei 30 000 xg für 30 Minuten zentrifugiert. Das Pellet wurde in 30 ml 50 mM Tris-HCl, pH 7,6, resuspendiert, bei 30 000 xg für 30 Minuten zentrifugiert, und das Pellet wurde in 25 ml Lösungspuffer (8 M Harnstoff, 137 mM NaCl, 2,7 mM KCl, 4,3 mM Na₂HPO₄, 1 mM KH₂PO₄, 5 mM βME) aufgelöst. Die Lösung wurde bei 30 000 xg für 30 Minuten zentrifugiert und der Überstand wurde gegen zwei Wechsel aus 4 l destillierten Wassers bei 4°C dialysiert. Nach der Dialyse wurde die gesamte Mischung lyophilisiert. Der lyophilisierte Feststoff wurde in 0,1 M HCl in einer Flasche unter Rühren gelöst. Nach der Zugabe eines 5-fachen Überschusses aus kristallinem BrCN wurde die Flasche geleert und mit Stickstoff gefüllt. Die Spaltung wurde für 24 Stunden ablaufen gelassen, worauf das Lösungsmittel im Vakuum entfernt wurde. Der Rest wurde in 0,1% Trifluoressigsäure (TFA) gelöst und mittels Umkehrphasen-HPLC unter Verwendung einer Vydac C4 RP-HPLC-Säule (10 × 250 mm, 5 µ, 300 Å) auf einem BioCad Sprint System (Perceptive Biosystems, Framingham, MA) gereinigt. Hydroxyprolin-haltiges D4-Protein wurde mit einem Gradienten auf 15 bis 40% Acetonitril/0,1% TFA über einen 45-minütigen Zeitraum eluiert. Protein

D4- α eluierte bei 25% Acetonitril/0,1% TFA.

Beispiel 29

Aminosäure-Zusammensetzungsanalyse eines Hydroxyprolin-haltigen C-terminalen Fragments vom Typ I (α_1)-Kollagen

[0237] Protein D4- α (10 μ g), welches in Beispiel 27 gereinigt wurde, wurde bis zur Trockenheit in vacuo in einem 1,5 ml-Mikrozentrifugenröhrchen gebracht. Eine Probe wurde einer Aminosäureanalyse im W. M. Keck-Foundation Biotechnology Resource Laboratory (New Haven, CT) auf einem Applied Biosystems-Sequenziergerät, das mit einem On-line-HPLC-System ausgerüstet ist, unterworfen. Die experimentell bestimmte Sequenz der ersten 13 Aminosäuren (SEQ ID NO: 41) und die Sequenz die aufgrund der DNA-Sequenz vorhergesagt wurde (SEQ ID NO: 42) ist in [Fig. 69](#) gezeigt. Eine Probe des Proteins D4- α 1 wurde einer Massenspektrometrieanalyse auf einem VG Biotech BIO-Q-Quadrupole-Analysegerät bei M-Scan, Inc. (West Chester, PA) unterworfen. Das Massenspektrum und das vorhergesagte Molekulargewicht des Proteins D4- α 1, wenn es 100% Hydroxyprolin anstelle von Prolin enthält, ist in [Fig. 70](#) angegeben. Das vorhergesagte Molekulargewicht des Proteins D4- α 1, das 100% Hydroxyprolin anstelle von Prolin enthält, beträgt 20 807,8 Da. Das experimentell bestimmte Molekulargewicht war 20 807,5 Da.

Beispiel 30

Konstruktion der Carboxy-terminalen 219 Aminosäuren des menschlichen Collagen-Typ I (α_1)-Fragment-Gens mit optimierter E. coli-Codonverwendung

[0238] Die Nukleotidsequenz des 657 Nukleotide langen Gens der Carboxy-terminalen 219 Aminosäuren des menschlichen Typ I (α_1)-Kollagens mit optimierter E. coli-Codonverwendung ist in [Fig. 71](#) gezeigt. Zur Synthese dieses Gens wurden einzigartige Restriktionsschnittstellen identifiziert oder ungefähr alle 150 Basenpaare erzeugt. Oligos aus ungefähr 80 Nukleotiden wurden auf einem Beckman Oligo 1000 DNA-Synthesegerät synthetisiert, gespalten und mit wässriger NH_4OH entschützt und mittels Elektrophorese in 7 M Harnstoff/12% Polyacrylamidgelen gereinigt. Jeder Satz an Oligos wurde so entworfen, daß er eine EcoRI-Restriktionsenzym-schnittstelle am 5'-Ende hat, eine einzigartige Restriktionsenzym-schnittstelle nahe dem 3'-Ende, gefolgt von der TAAT-Stoppsequenz und einer HindIII-Restriktionsenzym-schnittstelle ganz am 3'-Ende. Die ersten vier Oligos, welche die ersten 84 Aminosäuren der Carboxy-terminalen 219 Aminosäuren des menschlichen Typ I (α_1)-Kollagens mit optimierter E. coli-Codonverwendung umfassen, sind in [Fig. 81](#) angegeben (SEQ ID NO: 47–50).

[0239] Oligos N4-1 (SEQ ID NO: 47) und N4-2 (SEQ ID NO: 48) (1 μ g jeweils), wurden in 20 μ l T7-DNA-Polymerase-Puffer (40 mM Tris-HCl (pH 8,0), 5 mM MgCl_2 , 5 mM Dithiothreitol, 50 mM NaCl, 0,05 mg/ml Rinderserumalbumin) durch Erwärmen bei 90°C für 5 Minuten, gefolgt von langsamen Abkühlen auf Raumtemperatur aneinandergelagert. Nach einer kurzen Zentrifugierung bei 14 000 Upm, wurden 10 Einheiten T7-DNA-Polymerase und 2 μ l einer Lösung aus allen vier dNTPs (dATP, dGTP, dCTP, dTTP, jeweils 2,5 mM) zu den aneinandergelagerten Oligos gegeben. Die Verlängerungsreaktionen wurden bei 37°C für 30 min inkubiert und dann bei 70°C für 10 Minuten erwärmt. Nach dem Abkühlen auf Raumtemperatur, wurde HindIII-Puffer (5 μ l einer 10-fach Konzentration), 20 μ l H_2O und 10 Einheit HindIII-Restriktionsenzym hinzugegeben und die Röhrchen wurden bei 37°C für 10 Stunden inkubiert. HindIII-Puffer (2 μ l einer 10-fach Konzentration), 13,5 μ l 0,5 M Tris-HCl (pH 7,5), 1,8 μ l 1% Triton X100, 5,6 μ l H_2O und 20 U EcoRI wurden zu jedem Röhrchen hinzugegeben und die Inkubation wurde bei 37°C für 2 Stunden fortgesetzt. Die Verdau wurden einmal mit einem gleichen Volumen an Phenol, einmal mit Phenol/Chloroform/Isoamylalkohol und einmal mit Chloroform/Isoamylalkohol extrahiert. Nach einer Ethanol-Präzipitation wurde das Pellet in 10 μ l TE-Puffer (10 mM Tris-HCl (pH 8,0), 1 mM EDTA) resuspendiert. Vom resuspendierten Pellet wurde 4 μ l über Nacht bei 16°C mit Agarosegel gereinigtem EcoRI/HindIII-verdautes pBSKS⁺-Vektor (1 μ g) unter Verwendung von T4-DNA-Ligase (100 Einheiten) ligiert. Eine Hälfte des Transformationsgemisches wurde durch Wärmeschock in DH5 α -Zellen transformiert und 100 μ l der 1,0 ml-Transformationsmischung wurden auf Luria-Nährflüssigkeit (LB)-Agarplatten ausplattiert, welche 70 μ g/ml Ampicillin enthalten. Die Platten wurden über Nacht bei 37°C inkubiert. Die Ampicillin-resistenten Kolonien (6–12) wurden herausgepickt und über Nacht in LB-Medium gezüchtet, welche 70 μ g/ml Ampicillin enthalten. Plasmid-DNA wurde aus jeder Kultur durch Wizard Minipreps (Promega Corporation, Madison WI) isoliert und auf das Vorhandensein des ungefähr 120 Basenpaaren langen Insertionen durch Verdau mit EcoRI und HindIII durchmustert und durch Laufenlassen der verdauten Produkte auf Agarose-Elektrophoresegelelen. Klone mit Insertionen wurde mittels Standard-Dideoxyterminations-DNA-Sequenzierung überprüft. Der korrekte Klon wurde als pBSN4-1 bezeichnet.

[0240] Oligos N4-3 (SEQ ID NO: 49) und N4-4 (SEQ ID NO: 50) ([Fig. 81](#)) wurden synthetisiert, gereinigt, aneinandergelagert, verlängert und nach genau dem gleichen Verfahren, das oben für die Oligos N4-1 und N4-2 angegeben ist, in pBSKS⁺ kloniert. Das daraus hervorgehende Plasmid wurde als pBSN4-2A bezeichnet. Um die Abschnitte des Kollagen-Gens von pBSN4-1 und pBSN4-2A zusammenzuklonieren, wurde das Plasmid pBSN4-1 (1 µg) für 2 Stunden bei 37°C mit ApaI und HindIII verdaut. Der verdaute Vektor wurde mittels Agarosegelelektrophorese gereinigt. Plasmid pBSN4-2A (3 µg) wurde für 2 Stunden bei 37°C mit ApaI und HindIII verdaut und das Insert wurde mittels Agarosegelelektrophorese gereinigt. ApaI/HindIII-verdautes pBSN4-1 wurde mit dieser Insertion über Nacht bei 16°C mit T4-DNA-Ligase ligiert. Eine Hälfte der Ligationsmischung wurde in DH5α-Zellen transformiert und 1/10 des Transformationsgemisches wurde auf LB-Agarplatten ausgeplattiert, die 70 µg/ml Ampicillin enthalten. Nach einer Über-Nacht-Inkubation bei 37°C wurden Ampicillin-resistente Klone herausgepickt und auf das Vorhandensein der Insert-DNA, die oben beschrieben wurde, durchmustert. Klone wurde mittels Didesoxy-Terminationssequenzierung überprüft. Der korrekte Klon wurde als pBSN4-2 bezeichnet.

[0241] In ähnlicher Weise wurde der Rest des Gens für die Carboxy-terminalen 219 Aminosäuren des menschlichen Typ I (α₁)-Kollagens mit optimierter E. coli-Codonverwendung hergestellt, so daß die endgültige DNA-Sequenz die ist, die in [Fig. 71](#) (SEQ ID NO: 43) angegeben ist.

SEQUENZPROTOKOLL

(1) ALLGEMEINE ANGABEN:

(i) ANMELDER: GRUSKIN, ELLIOT A.
BUECHTER, DOUGLAS
BROKAW, JANE
ZHANG; GUANGHUI
PAOLELLA, DAVID

(ii) BEZEICHNUNG DER ERFINDUNG: Aminosäure-modifizierte
Polypeptide

(iii) ANZAHL DER SEQUENZEN: 50

(iv) BRIEFADRESSE:

(A) NAME: DILWORTH & BARRESE
(B) STRASSE: 333 EARLE OVINGTON BOULEVARD
(C) STADT: UNIONDALE
(D) BUNDESLAND: NY
(E) LAND: USA
(F) POSTLEITZAHL: 11553

(v) COMPUTERLESBARE FASSUNG:

(A) DATENTRÄGER: Floppy Disk
(B) COMPUTER: IBM-PC-kompatibel
(C) BETRIEBSSYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Ausgabe #1.0, Version #1.30

(vi) DATEN DER ANMELDUNG:

(A) ANMELDENUMMER:
(B) ANMELDEDATUM:
(C) KLASSIFIKATION:

(viii) ANGABEN ZUM ANWALT/VERTRETER:

(A) NAME: STEEN, JEFFREY S.

(ix) TELEKOMMUNIKATION:

(A) TELEFON: (516) 228-8484

(B) TELEFAX: (516) 228-8516

(2) ANGABEN ZU SEQ ID NO:1:

(i) SEQUENZCHARAKTERISTIKA:

(A) LÄNGE: 3170 Basenpaare

(B) ART: Nukleinsäure

(C) STRANGFORM: Einzelstrang

(D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: cDNA

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO:1:

CAGCTGTCTT ATGGCTATGA TGAGAAATCA ACCGGAGGAA TTTCCTGCC TGGCCCCATG	60
GGTCCCTCTG GTCCTCGTGG TCTCCCTGGC CCCCTGGTG CACCTGGTCC CCAAGGCTTC	120
CAAGGTCCCC CTGGTGAGCC TGGCGAGCCT GGAGCTTCAG GTCCCATGGG TCCCCGAGGT	180
CCCCCAGGTC CCCCTGGAAA GAATGGAGAT GATGGGGAAG CTGGAAAACC TGGTCGTCCT	240
GGTGAGCGTG GGCCTCCTGG GCCTCAGGGT GCTCGAGGAT TGCCCGGAAC AGCTGGCCTC	300
CCTGGAATGA AGGGACACAG AGGTTTCAGT GGTTTGGATG GTGCCAAGGG AGATGCTGGT	360
CCTGCTGGTC CTAAGGGTGA GCCTGGCAGC CCTGGTGAAA ATGGAGCTCC TGGTCAGATG	420

GGCCCCCGTG GCCTGCCTGG TGAGAGAGGT CGCCCTGGAG CCCCTGGCCC TGCTGGTGCT	480
CGTGGAATG ATGGTGCTAC TGGTGCTGCC GGGCCCCCTG GTCCCACCGG CCCCCTGGT	540
CCTCCTGGCT TCCCTGGTGC TGTGGTGCT AAGGGTGAAG CTGGTCCCCA AGGGCCCCGA	600
GGCTCTGAAG GTCCCCAGGG TGTGCGTGGT GAGCCTGGCC CCCCTGGCCC TGCTGGTGCT	660
GCTGGCCCTG CTGGAAACCC TGGTGCTGAT GGACAGCCTG GTGCTAAAGG TGCCAATGGT	720
GCTCCTGGTA TTGCTGGTGC TCCTGGCTTC CCTGGTGCCC GAGGCCCCTC TGGACCCCAG	780
GGCCCCGGCG GCCCTCCTGG TCCCAAGGGT AACAGCGGTG AACCTGGTGC TCCTGGCAGC	840
AAAGGAGACA CTGGTGCTAA GGGAGAGCCT GGCCCTGTTG GTGTTCAAGG ACCCCCTGGC	900
CCTGCTGGAG AGGAAGGAAA GCGAGGAGCT CGAGGTGAAC CCGGACCCAC TGGCCTGCCC	960
GGACCCCCTG GCGAGCGTGG TGGACCTGGT AGCCGTGGTT TCCCTGGCGC AGATGGTGTT	1020
GCTGGTCCCA AGGGTCCCGC TGGTGAACGT GGTTCCTCTG GCCCCGCTGG CCCCAGGA	1080
TCTCCTGGTG AAGCTGGTGC TCCCGGTGAA GCTGGTCTGC CTGGTGCCAA GGGTCTGACT	1140
GGAAGCCCTG GCAGCCCTGG TCCTGATGGC AAAACTGGCC CCCCTGGTCC CGCCGGTCAA	1200
GATGGTCGCC CCGGACCCCC AGGCCCACCT GGTGCCCCTG GTCAGGCTGG TGTGATGGGA	1260
TTCCCTGGAC CTAAAGGTGC TGCTGGAGAG CCCGGCAAGG CTGGAGAGCG AGGTGTTCCC	1320
GGACCCCCTG GCGCTGTCCG TCCTGCTGGC AAAGATGGAG AGGCTGGAGC TCAGGGACCC	1380
CCTGGCCCTG CTGGTCCCGC TGGCGAGAGA GGTGAACAAG GCCCTGCTGG CTCCCCCGGA	1440

TTCCAGGGTC TCCCTGGTCC TGCTGGTCCT CCAGGTGAAG CAGGCAAACC TGGTGAACAG	1500
GGTGTTCCTG GAGACCTTGG CGCCCCCTGGC CCCTCTGGAG CAAGAGGCGA GAGAGGTTTC	1560
CCTGGCGAGC GTGGTGTGCA AGGTCCCCCT GGTCTGCTG GACCCCGAGG GGCCAACGGT	1620
GCTCCCGGCA ACGATGGTGC TAAGGGTGAT GCTGGTGCCC CTGGAGCTCC CGGTAGCCAG	1680
GGCGCCCCCTG GCCTTCAGGG AATGCCTGGT GAACGTGGTG CAGCTGGTCT TCCAGGGCCT	1740
AAGGGTGACA GAGGTGATGC TGGTCCCAA GGTGCTGATG GCTCTCCTGG CAAAGATGGC	1800
GTCCGTGGTC TGACCGGCCC CATTGGTCCT CCTGGCCCTG CTGGTGCCCC TGGTGACAAG	1860
GGTGAAAGTG GTCCCAGCGG CCCTGCTGGT CCCACTGGAG CTCGTGGTGC CCCCGGAGAC	1920
CGTGGTGAGC CTGGTCCCCC CGGCCCTGCT GGCTTTGCTG GCCCCCCTGG TGCTGACGGC	1980
CAACCTGGTG CTAAAGGCGA ACCTGGTGAT GCTGGTGCCA AAGGCGATGC TGGTCCCCCT	2040
GGGCCTGCCG GACCCGCTGG ACCCCCTGGC CCCATTGGTA ATGTTGGTGC TCCTGGAGCC	2100
AAAGGTGCTC GGGCAGCGCT GGTCCCCCTG GTGCTACTGG TTTCCCTGGT GCTGCTGGCC	2160
GAGTCGGTCC TCCTGGCCCC TCTGGAAATG CTGGACCCCC TGGCCCTCCT GGTCTGCTG	2220
GCAAAGAAGG CGGCAAAGGT CCCCGTGGTG AACTGGCCC TGCTGGACGT CCTGGTGAAG	2280
TTGGTCCCCC TGGTCCCCCT GGCCCTGCTG GCGAGAAAGG ATCCCCCTGGT GCTGATGGTC	2340
CTGCTGGTGC TCCTGGTACT CCCGGGCCTC AAGGTATTGC TGGACAGCGT GGTGTGGTCG	2400
GCCTGCCTGG TCAGAGAGGA GAGAGAGGCT TCCCTGGTCT TCCTGGCCCC TCTGGTGAAC	2460

CTGGCAAACA AGGTCCCTCT GGAGCAAGTG GTGAACGTGG TCCCCCGGT CCCATGGGCC	2520
CCCCTGGATT GGCTGGACCC CCTGGTGAAT CTGGACGTGA GGGGGCTCCT GCTGCCGAAG	2580
GTTCCCCTGG ACGAGACGGT TCTCCTGGCG CCAAGGGTGA CCGTGGTGAG ACCGGCCCG	2640
CTGGACCCCC TGGTGCTCCT GGTGCTCCTG GTGCCCCCTGG CCCCCTTGGC CCTGCTGGCA	2700
AGAGTGGTGA TCGTGGTGAG ACTGGTCCTG CTGGTCCCGC CGGTCCCGTC GGCCCCGCTG	2760
GCGCCCGTGG CCCC GCCGA CCCCAAGGCC CCCGTGGTGA CAAGGGTGAG ACAGGCGAAC	2820
AGGGCGACAG AGGCATAAAG GGTCACCGTG GCTTCTCTGG CCTCCAGGGT CCCCCTGGCC	2880
CTCCTGGCTC TCCTGGTGAA CAAGGTCCCT CTGGAGCCTC TGGTCCTGCT GGTCCCCGAG	2940
GTCCCCCTGG CTCTGCTGGT GCTCCTGGCA AAGATGGACT CAACGGTCTC CCTGGCCCCA	3000
TTGGGCCCCC TGGTCCTCGC GGTCGCACTG GTGATGCTGG TCCTGTTGGT CCCCCGGCC	3060
CTCCTGGACC TCCTGGTCCC CCTGGTCCTC CCAGCGCTGG TTTCGACTTC AGCTTCCTCC	3120
CCCAGCCACC TCAAGAGAAG GCTCACGATG GTGGCCGCTA CTACCGGGCT	3170

(2) ANGABEN ZU SEQ ID NO:2:

(i) SEQUENZCHARAKTERISTIKA:

- (A) LÄNGE: 240 Basenpaare
- (B) ART: Nukleinsäure
- (C) STRANGFORM: Einzelstrang
- (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: cDNA

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO:2:

CAGCTGTCTT ATGGCTATGA TGAGAAATCA ACCGGAGGAA TTTCCGTGCC TGGCCCCATG	60
GGTCCCTCTG GTCCTCGTGG TCTCCCTGGC CCCCTGGTG CACCTGGTCC CCAAGGCTTC	120
CAAGGTCCCC CTGGTGAGCC TGGCGAGCCT GGAGCTTCAG GTCCCATGGG TCCCCGAGGT	180
CCCCCAGGTC CCCCTGGAAA GAATGGAGAT GATGGGGAAG CTGGAAAACC TGGTCGTCCT	240

(2) ANGABEN ZU SEQ ID NO:3:

(i) SEQUENZCHARAKTERISTIKA:

- (A) LÄNGE: 100 Basenpaare
- (B) ART: Nukleinsäure
- (C) STRANGFORM: Einzelstrang
- (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: cDNA

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO:3:

GGATCCATGG GGCTCGCTGG CCCACCGGGC GAACCGGGTC CGCCAGGCCC GAAAGGTCCG	60
CGTGGCGATA GCGGGCTCCC GGGCGATTCC TAATGGATCC	100

(2) ANGABEN ZU SEQ ID NO:4:

(i) SEQUENZCHARAKTERISTIKA:

- (A) LÄNGE: 21 Aminosäuren
- (B) ART: Aminosäure
- (C) STRANGFORM: Einzelstrang
- (D) TOPOLOGIE: unbekannt

(ii) ART DES MOLEKÜLS: Peptid

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO:4:

Gly Leu Ala Gly Pro Pro Gly Glu Pro Gly Pro Pro Gly Pro Lys Gly
 1 5 10 15

Pro Arg Gly Asp Ser
 20

(2) ANGABEN ZU SEQ ID NO:5:

(i) SEQUENZCHARAKTERISTIKA:

- (A) LÄNGE: 330 Basenpaare
- (B) ART: Nukleinsäure
- (C) STRANGFORM: Einzelstrang
- (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: cDNA

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO:5:

CAGCGGGCCA GGAAGAAGAA TAAGAACTGC CGGCGCCACT CGCTCTATGT GGACTTCAGC 60

GATGTGGGCT GGAATGACTG GATTGTGGCC CCACCAGGCT ACCAGGCCTT CTACTGCCAT 120

GGGGACTGCC CCTTTCCACT GGCTGACCAC CTCAACTCAA CCAACCATGC CATTGTGCAG 180

ACCCTGGTCA ATTCTGTCAA TTCCAGTATC CCCAAAGCCT GTTGTGTGCC CACTGAACTG 240

AGTGCCATCT CCATGCTGTA CCTGGATGAG TATGATAAGG TGGTACTGAA AAATTATCAG 300

GAGATGGTAG TAGAGGGATG TGGGTGCCGC 330

(2) ANGABEN ZU SEQ ID NO:6:

(i) SEQUENZCHARAKTERISTIKA:

- (A) LÄNGE: 1169 Aminosäuren
- (B) ART: Aminosäure
- (C) STRANGFORM: Einzelstrang
- (D) TOPOLOGIE: unbekannt

(ii) ART DES MOLEKÜLS: Peptid

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO:6:

Gln	Leu	Ser	Tyr	Gly	Tyr	Asp	Glu	Lys	Ser	Thr	Gly	Gly	Ile	Ser	Val
1				5					10					15	
Pro	Gly	Pro	Met	Gly	Pro	Ser	Gly	Pro	Arg	Gly	Leu	Pro	Gly	Pro	Pro
			20					25					30		
Gly	Ala	Pro	Gly	Pro	Gln	Gly	Phe	Gln	Gly	Pro	Pro	Gly	Glu	Pro	Gly
			35				40					45			
Glu	Pro	Gly	Ala	Ser	Gly	Pro	Met	Gly	Pro	Arg	Gly	Pro	Pro	Gly	Pro
			50				55				60				
Pro	Gly	Lys	Asn	Gly	Asp	Asp	Gly	Glu	Ala	Gly	Lys	Pro	Gly	Arg	Pro
65					70					75				80	

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

Ser Gly Pro Gln Gly Pro Gly Gly Pro Pro Gly Pro Lys Gly Asn Ser
 260 265 270

Gly Glu Pro Gly Ala Pro Gly Ser Lys Gly Asp Thr Gly Ala Lys Gly
 275 280 285

Glu Pro Gly Pro Val Gly Val Gln Gly Pro Pro Gly Pro Ala Gly Glu
 290 295 300

Glu Gly Lys Arg Gly Ala Arg Gly Glu Pro Gly Pro Thr Gly Leu Pro
 305 310 315 320

Gly Pro Pro Gly Glu Arg Gly Gly Pro Gly Ser Arg Gly Phe Pro Gly
 325 330 335

Ala Asp Gly Val Ala Gly Pro Lys Gly Pro Ala Gly Glu Arg Gly Ser
 340 345 350

Pro Gly Pro Ala Gly Pro Lys Gly Ser Pro Gly Glu Ala Gly Arg Pro
 355 360 365

Gly Glu Ala Gly Leu Pro Gly Ala Lys Gly Leu Thr Gly Ser Pro Gly
 370 375 380

Ser Pro Gly Pro Asp Gly Lys Thr Gly Pro Pro Gly Pro Ala Gly Gln
 385 390 395 400

Asp Gly Arg Pro Gly Pro Pro Gly Pro Pro Gly Ala Arg Gly Gln Ala
 405 410 415

Gly Val Met Gly Phe Pro Gly Pro Lys Gly Ala Ala Gly Glu Pro Gly
 420 425 430

Lys Ala Gly Glu Arg Gly Val Pro Gly Pro Pro Gly Ala Val Gly Pro			
435	440	445	
Ala Gly Lys Asp Gly Glu Ala Gly Ala Gln Gly Pro Pro Gly Pro Ala			
450	455	460	
Gly Pro Ala Gly Glu Arg Gly Glu Gln Gly Pro Ala Gly Ser Pro Gly			
465	470	475	480
Phe Gln Gly Leu Pro Gly Pro Ala Gly Pro Pro Gly Glu Ala Gly Lys			
485	490	495	
Pro Gly Glu Gln Gly Val Pro Gly Asp Leu Gly Ala Pro Gly Pro Ser			
500	505	510	
Gly Ala Arg Gly Glu Arg Gly Phe Pro Gly Glu Arg Gly Val Gln Gly			
515	520	525	
Pro Pro Gly Pro Ala Gly Pro Arg Gly Ala Asn Gly Ala Pro Gly Asn			
530	535	540	
Asp Gly Ala Lys Gly Asp Ala Gly Ala Pro Gly Ala Pro Gly Ser Gln			
545	550	555	560
Gly Ala Pro Gly Leu Gln Gly Met Pro Gly Glu Arg Gly Ala Ala Gly			
565	570	575	
Leu Pro Gly Pro Lys Gly Asp Arg Gly Asp Ala Gly Pro Lys Gly Ala			
580	585	590	
Asp Gly Ser Pro Gly Lys Asp Gly Val Arg Gly Leu Thr Gly Pro Ile			
595	600	605	

Gly Pro Pro Gly Pro Ala Gly Ala Pro Gly Asp Lys Gly Glu Ser Gly
610 615 620

Pro Ser Gly Pro Ala Gly Pro Thr Gly Ala Arg Gly Ala Pro Gly Asp
625 630 635 640

Arg Gly Glu Pro Gly Pro Pro Gly Pro Ala Gly Phe Ala Gly Pro Pro
645 650 655

Gly Ala Asp Gly Gln Pro Gly Ala Lys Gly Glu Pro Gly Asp Ala Gly
660 665 670

Ala Lys Gly Asp Ala Gly Pro Pro Gly Pro Ala Gly Pro Ala Gly Pro
675 680 685

Pro Gly Pro Ile Gly Asn Val Gly Ala Pro Gly Ala Lys Gly Ala Arg
690 695 700

Gly Ser Ala Gly Pro Pro Gly Ala Thr Gly Phe Pro Gly Ala Ala Gly
705 710 715 720

Arg Val Gly Pro Pro Gly Pro Ser Gly Asn Ala Gly Pro Pro Gly Pro
725 730 735

Pro Gly Pro Ala Gly Lys Glu Gly Gly Lys Gly Pro Arg Gly Glu Thr
740 745 750

Gly Pro Ala Gly Arg Pro Gly Glu Val Gly Pro Pro Gly Pro Pro Gly
755 760 765

Pro Ala Gly Glu Lys Gly Ser Pro Gly Ala Asp Gly Pro Ala Gly Ala
770 775 780

Pro Gly Thr Pro Gly Pro Gln Gly Ile Ala Gly Gln Arg Gly Val Val
 785 790 795 800

Gly Leu Pro Gly Gln Arg Gly Glu Arg Gly Phe Pro Gly Leu Pro Gly
 805 810 815

Pro Ser Gly Glu Pro Gly Lys Gln Gly Pro Ser Gly Ala Ser Gly Glu
 820 825 830

Arg Gly Pro Pro Gly Pro Met Gly Pro Pro Gly Leu Ala Gly Pro Pro
 835 840 845

Gly Glu Ser Gly Arg Glu Gly Ala Pro Ala Ala Glu Gly Ser Pro Gly
 850 855 860

Arg Asp Gly Ser Pro Gly Ala Lys Gly Asp Arg Gly Glu Thr Gly Pro
 865 870 875 880

Ala Gly Pro Pro Gly Ala Xaa Gly Ala Xaa Gly Ala Pro Gly Pro Val
 885 890 895

Gly Pro Ala Gly Lys Ser Gly Asp Arg Gly Glu Thr Gly Pro Ala Gly
 900 905 910

Pro Ala Gly Pro Val Gly Pro Ala Gly Ala Arg Gly Pro Ala Gly Pro
 915 920 925

Gln Gly Pro Arg Gly Asp Lys Gly Glu Thr Gly Glu Gln Gly Asp Arg
 930 935 940

Gly Ile Lys Gly His Arg Gly Phe Ser Gly Leu Gln Gly Pro Pro Gly
 945 950 955 960

Pro Pro Gly Ser Pro Gly Glu Gln Gly Pro Ser Gly Ala Ser Gly Pro
 965 970 975

Ala Gly Pro Arg Gly Pro Pro Gly Ser Ala Gly Ala Pro Gly Lys Asp
 980 985 990

Gly Leu Asn Gly Leu Pro Gly Pro Ile Gly Pro Pro Gly Pro Arg Gly
 995 1000 1005

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

Pro Gly Pro Pro Gly Pro Pro Ser Ala Gly Phe Asp Phe Ser Phe Leu
 1025 1030 1035 1040

Pro Gln Pro Pro Gln Glu Lys Ala His Asp Gly Gly Arg Tyr Tyr Arg
 1045 1050 1055

Ala Arg Ser Gln Arg Ala Arg Lys Lys Asn Lys Asn Cys Arg Arg His
 1060 1065 1070

Ser Leu Tyr Val Asp Phe Ser Asp Val Gly Trp Asn Asp Trp Ile Val
 1075 1080 1085

Ala Pro Pro Gly Tyr Gln Ala Phe Tyr Cys His Gly Asp Cys Pro Phe
 1090 1095 1100

Pro Leu Ala Asp His Leu Asn Ser Thr Asn His Ala Ile Val Gln Thr
 1105 1110 1115 1120

Leu Val Asn Ser Val Asn Ser Ser Ile Pro Lys Ala Cys Cys Val Pro
 1125 1130 1135

Thr Glu Leu Ser Ala Ile Ser Met Leu Tyr Leu Asp Glu Tyr Asp Lys
 1140 1145 1150

Val Val Leu Lys Asn Tyr Gln Glu Met Val Val Glu Gly Cys Gly Cys
 1155 1160 1165

Arg

(2) ANGABEN ZU SEQ ID NO:7:

(i) SEQUENZCHARAKTERISTIKA:

- (A) LÄNGE: 3531 Basenpaare
- (B) ART: Nukleinsäure
- (C) STRANGFORM: Einzelstrang
- (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: cDNA

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO:7:

GGGAAGGATT TCCATTCCCC AGCTGTCTTA TGGCTATGAT GAGAAATCAA CCGGAGGAAT	60
TTCCGTGCCT GGCCCCATGG GTCCCTCTGG TCCTCGTGGT CTCCCTGGCC CCCCTGGTGC	120
ACCTGGTCCC CAAGGCTTCC AAGGTCCCCC TGGTGAGCCT GGCAGCCTG GAGCTTCAGG	180
TCCCATGGGT CCCCAGGTC CCCAGGTCC CCCTGGAAAG AATGGAGATG ATGGGGAAGC	240
TGAAAAACCT GGTCGTCTG GTGAGCGTGG GCCTCCTGGG CTCAGGGTG CTCGAGGATT	300
GCCCCGAACA GCTGGCCTCC CTGGAATGAA GGGACACAGA GGTTCAGTG GTTTGGATGG	360
TGCCAAGGGA GATGCTGGTC CTGCTGGTCC TAAGGGTGAG CCTGGCAGCC CTGGTGAAAA	420

TGGAGCTCCT GGTGAGATGG GCCCCCGTGG CCTGCCTGGT GAGAGAGGTC GCCCTGGAGC	480
CCCTGGCCCT GCTGGTGCTC GTGGAAATGA TGGTGCTACT GGTGCTGCCG GGCCCCCTGG	540
TCCCACCGGC CCCGCTGGTC CTCCTGGCTT CCCTGGTGCT GTTGGTGCTA AGGGTGAAGC	600
TGGTCCCCAA GGGCCCCGAG GCTCTGAAGG TCCCAGGGT GTCCGTGGTG AGCCTGGCCC	660
CCCTGGCCCT GCTGGTGCTG CTGGCCCTGC TGGAAACCCT GGTGCTGATG GACAGCCTGG	720
TGCTAAAGGT GCCAATGGTG CTCCTGGTAT TGCTGGTGCT CCTGGCTTCC CTGGTGCCCCG	780
AGGCCCCCTCT GGACCCCAGG GCCCCGGCGG CCCTCCTGGT CCCAAGGGTA ACAGCGGTGA	840
ACCTGGTGCT CCTGGCAGCA AAGGAGACAC TGGTGCTAAG GGAGAGCCTG GCCCTGTTGG	900
TGTTCAAGGA CCCCCTGGCC CTGCTGGAGA GGAAGGAAAG CGAGGAGCTC GAGGTGAACC	960
CGGACCCACT GGCCTGCCCC GACCCCCTGG CGAGCGTGGT GGACCTGGTA GCCGTGGTTT	1020
CCCTGGCGCA GATGGTGTTG CTGGTCCCAA GGGTCCCCGT GGTGAACGTG GTTCTCCTGG	1080
CCCCGCTGGC CCCAAGGAT CTCCTGGTGA AGCTGGTCGT CCCGGTGAAG CTGGTCTGCC	1140
TGGTGCCAAG GGTCTGACTG GAAGCCCTGG CAGCCCTGGT CCTGATGGCA AACTGGCCC	1200
CCCTGGTCCC GCCGGTCAAG ATGGTCGCCC CGGACCCCCA GGCCACCTG GTGCCCCTGG	1260
TCAGGCTGGT GTGATGGGAT TCCCTGGACC TAAAGGTGCT GCTGGAGAGC CCGGCAAGGC	1320
TGGAGAGCGA GGTGTTCCCC GACCCCCTGG CGCTGTCGGT CCTGCTGGCA AAGATGGAGA	1380
GGCTGGAGCT CAGGGACCCC CTGGCCCTGC TGGTCCCCGT GGCAGAGAG GTGAACAAGG	1440

CCCTGCTGGC TCCCCCGGAT TCCAGGGTCT CCCTGGTCCT GCTGGTCCTC CAGGTGAAGC	1500
AGGCAAACCT GGTGAACAGG GTGTTCTTGG AGACCTTGGC GCCCCTGGCC CCTCTGGAGC	1560
AAGAGGCGAG AGAGGTTTCC CTGGCGAGCG TGGTGTGCAA GGTCCCCCTG GTCCTGCTGG	1620
ACCCCGAGGG GCCAACGGTG CTCCCCGCAA CGATGGTGCT AAGGGTGATG CTGGTGCCCC	1680
TGGAGCTCCC GGTAGCCAGG GCGCCCCCTGG CCTTCAGGGA ATGCCTGGTG AACGTGGTGC	1740
AGCTGGTCTT CCAGGGCCTA AGGGTGACAG AGGTGATGCT GGTCCCAAAG GTGCTGATGG	1800
CTCTCCTGGC AAAGATGGCG TCCGTGGTCT GACCGGCCCC ATTGGTCCTC CTGGCCCTGC	1860
TGGTGCCCCT GGTGACAAGG GTGAAAGTGG TCCCAGCGGC CCTGCTGGTC CCACTGGAGC	1920
TCGTGGTGCC CCCGGAGACC GTGGTGAGCC TGGTCCCCC GGCCCTGCTG GCTTTGCTGG	1980
CCCCCCTGGT GCTGACGGCC AACCTGGTGC TAAAGGCGAA CCTGGTGATG CTGGTGCCAA	2040
AGGCGATGGG TCCCCCTGGG CCTGCCGGAC CCGCTGGACC CCCTGGCCCC ATTGGTAATG	2100
TTGGTGCTCC TGGAGCCAAA GGTGCTCGCG GCAGCGCTGG TCCCCCTGGT GCTACTGGTT	2160
TCCCTGGTGC TGCTGGCCGA GTCGGTCCTC CTGGCCCCCTC TGGAAATGCT GGACCCCCTG	2220
GCCCTCCTGG TCCTGCTGGC AAAGAAGGCG GCAAAGGTCC CCGTGGTGAG ACTGGCCCTG	2280
CTGGACGTCC TGGTGAAGTT GGTCCCCCTG GTCCCCCTGG CCCTGCTGGC GAGAAAGGAT	2340
CCCCTGGTGC TGATGGTCCT GCTGGTGCTC CTGGTACTCC CGGGCCTCAA GGTATTGCTG	2400
GACAGCGTGG TGTGGTCGGC CTGCCTGGTC AGAGAGGAGA GAGAGGCTTC CCTGGTCTTC	2460

CTGGCCCCTC TGGTGAACCT GGCAAACAAG GTCCCTCTGG AGCAAGTGGT GAACGTGGTC	2520
CCCCCGGTCC CATGGGCCCC CCTGGATTGG CTGGACCCCC TGGTGAATCT GGACGTGAGG	2580
GGGCTCCTGC TGCCGAAGGT TCCCCTGGAC GAGACGGTTC TCCTGGCGCC AAGGGTGACC	2640
GTGGTGAGAC CGGCCCCGCT GGACCCCCTG GTGCTCTGGT GCTCTGGTGC CCCTGGCCCC	2700
GTTGGCCCTG CTGGCAAGAG TGGTGATCGT GGTGAGACTG GTCCTGCTGG TCCCGCCGGT	2760
CCCGTCGGCC CCGCTGGCGC CCGTGGCCCC GCCGGACCCC AAGGCCCCCG TGGTGACAAG	2820
GGTGAGACAG GCGAACAGGG CGACAGAGGC ATAAAGGGTC ACCGTGGCTT CTCTGGCCTC	2880
CAGGGTCCCC CTGGCCCTCC TGGCTCTCCT GGTGAACAAG GTCCCTCTGG AGCCTCTGGT	2940
CCTGCTGGTC CCCGAGGTCC CCCTGGCTCT GCTGGTGCTC CTGGCAAAGA TGGACTCAAC	3000
GGTCTCCCTG GCCCCATTGG GCCCCCTGGT CCTCGCGGTC GCACTGGTGA TGCTGGTCTT	3060
GTTGGTCCCC CCGGCCCTCC TGGACCTCCT GGTCCCCCTG GTCCTCCCAG CGCTGGTTTC	3120
GACTTCAGCT TCCTCCCCCA GCCACCTCAA GAGAAGGCTC ACGATGGTGG CCGCTACTAC	3180
CGGGCTAGAT CCCAGCGGGC CAGGAAGAAG AATAAGAACT GCCGGCGCCA CTCGCTCTAT	3240
GTGGACTTCA GCGATGTGGG CTGGAATGAC TGGATTGTGG CCCCACCAGG CTACCAGGCC	3300
TTCTACTGCC ATGGGGACTG CCCCTTTCCA CTGGCTGACC ACCTCAACTC AACCAACCAT	3360
GCCATTGTGC AGACCCTGGT CAATTCTGTC AATTCCAGTA TCCCCAAAGC CTGTTGTGTG	3420
CCCACTGAAC TGAGTGCCAT CTCCATGCTG TACCTGGATG AGTATGATAA GGTGGTACTG	3480

AAAAATTATC AGGAGATGGT AGTAGAGGGA TGTGGGTGCC GCTAAAAGCT T

3531

(2) ANGABEN ZU SEQ ID NO:8:

(i) SEQUENZCHARAKTERISTIKA:

(A) LÄNGE: 1171 Aminosäuren

(B) ART: Aminosäure

(C) STRANGFORM: Einzelstrang

(D) TOPOLOGIE: unbekannt

(ii) ART DES MOLEKÜLS: Peptid

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO:8:

Gln Leu Ser Tyr Gly Tyr Asp Glu Lys Ser Thr Gly Gly Ile Ser Val
1 5 10 15

Pro Gly Pro Met Gly Pro Ser Gly Pro Arg Gly Leu Pro Gly Pro Pro
20 25 30

Gly Ala Pro Gly Pro Gln Gly Phe Gln Gly Pro Pro Gly Glu Pro Gly
35 40 45

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

Pro Gly Lys Asn Gly Asp Asp Gly Glu Ala Gly Lys Pro Gly Arg Pro

65 70 75 80

Gly Glu Arg Gly Pro Pro Gly Pro Gln Gly Ala Arg Gly Leu Pro Gly
85 90 95

Thr Ala Gly Leu Pro Gly Met Lys Gly His Arg Gly Phe Ser Gly Leu
100 105 110

Asp Gly Ala Lys Gly Asp Ala Gly Pro Ala Gly Pro Lys Gly Glu Pro
 115 120 125

Gly Ser Pro Gly Glu Asn Gly Ala Pro Gly Gln Met Gly Pro Arg Gly
 130 135 140

Leu Pro Gly Glu Arg Gly Arg Pro Gly Ala Pro Gly Pro Ala Gly Ala
 145 150 155 160

Arg Gly Asn Asp Gly Ala Thr Gly Ala Ala Gly Pro Pro Gly Pro Thr
 165 170 175

Gly Pro Ala Gly Pro Pro Gly Phe Pro Gly Ala Val Gly Ala Lys Gly
 180 185 190

Glu Ala Gly Pro Gln Gly Pro Arg Gly Ser Glu Gly Pro Gln Gly Val
 195 200 205

Arg Gly Glu Pro Gly Pro Pro Gly Pro Ala Gly Ala Ala Gly Pro Ala
 210 215 220

Gly Asn Pro Gly Ala Asp Gly Gln Pro Gly Ala Lys Gly Ala Asn Gly
 225 230 235 240

Ala Pro Gly Ile Ala Gly Ala Pro Gly Phe Pro Gly Ala Arg Gly Pro
 245 250 255

Ser Gly Pro Gln Gly Pro Gly Gly Pro Pro Gly Pro Lys Gly Asn Ser
 260 265 270

Gly Glu Pro Gly Ala Pro Gly Ser Lys Gly Asp Thr Gly Ala Lys Gly
 275 280 285

Glu Pro Gly Pro Val Gly Val Gln Gly Pro Pro Gly Pro Ala Gly Glu
 290 295 300

Glu Gly Lys Arg Gly Ala Arg Gly Glu Pro Gly Pro Thr Gly Leu Pro
 305 310 315 320

Gly Pro Pro Gly Glu Arg Gly Gly Pro Gly Ser Arg Gly Phe Pro Gly
 325 330 335

Ala Asp Gly Val Ala Gly Pro Lys Gly Pro Ala Gly Glu Arg Gly Ser
 340 345 350

Pro Gly Pro Ala Gly Pro Lys Gly Ser Pro Gly Glu Ala Gly Arg Pro
 355 360 365

Gly Glu Ala Gly Leu Pro Gly Ala Lys Gly Leu Thr Gly Ser Pro Gly
 370 375 380

Ser Pro Gly Pro Asp Gly Lys Thr Gly Pro Pro Gly Pro Ala Gly Gln
 385 390 395 400

Asp Gly Arg Pro Gly Pro Pro Gly Pro Pro Gly Ala Arg Gly Gln Ala
 405 410 415

Gly Val Met Gly Phe Pro Gly Pro Lys Gly Ala Ala Gly Glu Pro Gly
 420 425 430

Lys Ala Gly Glu Arg Gly Val Pro Gly Pro Pro Gly Ala Val Gly Pro
 435 440 445

Ala Gly Lys Asp Gly Glu Ala Gly Ala Gln Gly Pro Pro Gly Pro Ala
 450 455 460

Gly Pro Ala Gly Glu Arg Gly Glu Gln Gly Pro Ala Gly Ser Pro Gly
 465 470 475 480

Phe Gln Gly Leu Pro Gly Pro Ala Gly Pro Pro Gly Glu Ala Gly Lys
 485 490 495

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

Gly Ala Arg Gly Glu Arg Gly Phe Pro Gly Glu Arg Gly Val Gln Gly
 515 520 525

Pro Pro Gly Pro Ala Gly Pro Arg Gly Ala Asn Gly Ala Pro Gly Asn
 530 535 540

Asp Gly Ala Lys Gly Asp Ala Gly Ala Pro Gly Ala Pro Gly Ser Gln
 545 550 555 560

Gly Ala Pro Gly Leu Gln Gly Met Pro Gly Glu Arg Gly Ala Ala Gly
 565 570 575

Leu Pro Gly Pro Lys Gly Asp Arg Gly Asp Ala Gly Pro Lys Gly Ala
 580 585 590

Asp Gly Ser Pro Gly Lys Asp Gly Val Arg Gly Leu Thr Gly Pro Ile
 595 600 605

Gly Pro Pro Gly Pro Ala Gly Ala Pro Gly Asp Lys Gly Glu Ser Gly
 610 615 620

Pro Ser Gly Pro Ala Gly Pro Thr Gly Ala Arg Gly Ala Pro Gly Asp
 625 630 635 640

Arg Gly Glu Pro Gly Pro Pro Gly Pro Ala Gly Phe Ala Gly Pro Pro	645	650	655
Gly Ala Asp Gly Gln Pro Gly Ala Lys Gly Glu Pro Gly Asp Ala Gly	660	665	670
Ala Lys Gly Asp Ala Gly Pro Pro Gly Pro Ala Gly Pro Ala Gly Pro	675	680	685
Pro Gly Pro Ile Gly Asn Val Gly Ala Pro Gly Ala Lys Gly Ala Arg	690	695	700
Gly Ser Ala Gly Pro Pro Gly Ala Thr Gly Phe Pro Gly Ala Ala Gly	705	710	715
Arg Val Gly Pro Pro Gly Pro Ser Gly Asn Ala Gly Pro Pro Gly Pro	725	730	735
Pro Gly Pro Ala Gly Lys Glu Gly Gly Lys Gly Pro Arg Gly Glu Thr	740	745	750
Gly Pro Ala Gly Arg Pro Gly Glu Val Gly Pro Pro Gly Pro Pro Gly	755	760	765
Pro Ala Gly Glu Lys Gly Ser Pro Gly Ala Asp Gly Pro Ala Gly Ala	770	775	780
Pro Gly Thr Pro Gly Pro Gln Gly Ile Ala Gly Gln Arg Gly Val Val	785	790	795
Gly Leu Pro Gly Gln Arg Gly Glu Arg Gly Phe Pro Gly Leu Pro Gly	805	810	815

Pro Ser Gly Glu	Pro Gly Lys Gln Gly	Pro Ser Gly Ala	Ser Gly Glu
820	825	830	
Arg Gly Pro Pro Gly	Pro Met Gly Pro Pro Gly	Leu Ala Gly	Pro Pro
835	840	845	
Gly Glu Ser Gly Arg Glu Gly	Ala Pro Gly Ala	Glu Gly Ser Pro Gly	
850	855	860	
Arg Asp Gly Ser Pro Gly	Ala Lys Gly Asp Arg Gly	Glu Thr Gly Pro	
865	870	875	880
Ala Gly Pro Pro Gly	Ala Pro Gly Ala Pro Gly	Pro Gly Pro Val	
885	890	895	
Gly Pro Ala Gly Lys Ser Gly	Asp Arg Gly Glu Thr Gly	Pro Ala Gly	
900	905	910	
Pro Ala Gly Pro Val Gly	Pro Ala Gly Ala Arg Gly	Pro Ala Gly Pro	
915	920	925	
Gln Gly Pro Arg Gly Asp Lys Gly	Glu Thr Gly Glu Gln Gly	Asp Arg	
930	935	940	
Gly Ile Lys Gly His Arg Gly	Phe Ser Gly Leu Gln Gly	Pro Pro Gly	
945	950	955	960
Pro Pro Gly Ser Pro Gly	Glu Gln Gly Pro Ser Gly	Ala Ser Gly Pro	
965	970	975	
Ala Gly Pro Arg Gly Pro Pro Gly	Ser Ala Gly Ala Pro Gly	Lys Asp	
980	985	990	

Gly Leu Asn Gly Leu Pro Gly Pro Ile Gly Pro Pro Gly Pro Arg Gly
 995 1000 1005

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

Pro Gly Pro Pro Gly Pro Pro Ser Ala Gly Phe Asp Phe Ser Phe Leu
 1025 1030 1035 1040

Pro Gln Pro Pro Gln Glu Lys Ala His Asp Gly Gly Arg Tyr Tyr Arg
 1045 1050 1055

Ala Arg Ser Ala Leu Asp Thr Asn Tyr Cys Phe Ser Ser Thr Glu Lys
 1060 1065 1070

Asn Cys Cys Val Arg Gln Leu Tyr Ile Asp Phe Arg Lys Asp Leu Gly
 1075 1080 1085

Trp Lys Trp Ile His Glu Pro Lys Gly Tyr His Ala Asn Phe Cys Leu
 1090 1095 1100

Gly Pro Cys Pro Tyr Ile Trp Ser Leu Asp Thr Gln Tyr Ser Lys Val
 1105 1110 1115 1120

Leu Ala Leu Tyr Asn Gln His Asn Pro Gly Ala Ser Ala Ala Pro Cys
 1125 1130 1135

Cys Val Pro Gln Ala Leu Glu Pro Leu Pro Ile Val Tyr Tyr Val Gly
 1140 1145 1150

Arg Lys Pro Lys Val Glu Gln Leu Ser Asn Met Ile Val Arg Ser Cys
 1155 1160 1165

Lys Cys Ser

1170

(2) ANGABEN ZU SEQ ID NO:9:

(i) SEQUENZCHARAKTERISTIKA:

- (A) LÄNGE: 3541 Basenpaare
- (B) ART: Nukleinsäure
- (C) STRANGFORM: Einzelstrang
- (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: cDNA

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO:9:

GGGAAGGATT TCCATTCCC AGCTGTCTTA TGGCTATGAT GAGAAATCAA CCGGAGGAAT	60
TTCCGTGCCT GGCCCCATGG GTCCCTCTGG TCCTCGTGGT CTCCCTGGCC CCCCTGGTGC	120
ACCTGGTCCC CAAGGCTTCC AAGGTCCCC TGGTGAGCCT GGCGAGCCTG GAGCTTCAGG	180
TCCCATGGGT CCCCAGGTC CCCCAGGTCC CCCTGGAAAG AATGGAGATG ATGGGGAAGC	240
TGGAAAACCT GGTCGTCCTG GTGAGCGTGG GCCTCCTGGG CTCAGGGTG CTCGAGGATT	300
GCCCGGAACA GCTGGCCTCC CTGGAATGAA GGGACACAGA GGTTTCAGTG GTTTGGATGG	360
TGCCAAGGGA GATGCTGGTC CTGCTGGTCC TAAGGGTGAG CCTGGCAGCC CTGGTGAAAA	420
TGGAGCTCCT GGTCAGATGG GCCCCGTGG CCTGCCTGGT GAGAGAGGTC GCCCTGGAGC	480
CCCTGGCCCT GCTGGTGCTC GTGGAAATGA TGGTGCTACT GGTGCTGCCG GGCCCCCTGG	540
TCCCACCGGC CCCGCTGGTC CTCCTGGCTT CCCTGGTGCT GTTGGTGCTA AGGGTGAAGC	600

TGGTCCCCAA GGGCCCCGAG GCTCTGAAGG TCCCCAGGGT GTGCGTGGTG AGCCTGGCCC	660
CCCTGGCCCT GCTGGTGCTG CTGGCCCTGC TGGAAACCCT GGTGCTGATG GACAGCCTGG	720
TGCTAAAGGT GCCAATGGTG CTCCTGGTAT TGCTGGTGCT CCTGGCTTCC CTGGTGCCCCG	780
AGGCCCTCTT GGACCCCAGG GCCCCGGCGG CCCTCCTGGT CCCAAGGGTA ACAGCGGTGA	840
ACCTGGTGCT CCTGGCAGCA AAGGAGACAC TGGTGCTAAG GGAGAGCCTG GCCCTGTTGG	900
TGTTCAAGGA CCCCCTGGCC CTGCTGGAGA GGAAGGAAAG CGAGGAGCTC GAGGTGAACC	960
CGGACCCACT GGCCTGCCCC GACCCCCTGG CGAGCGTGGT GGACCTGGTA GCCGTGGTTT	1020
CCCTGGCGCA GATGGTGTTG CTGGTCCCAA GGGTCCCGCT GGTGAACGTG GTTCTCCTGG	1080
CCCCGCTGGC CCCAAAGGAT CTCCTGGTGA AGCTGGTCGT CCCGGTGAAG CTGGTCTGCC	1140
TGGTGCCAAG GGTCTGACTG GAAGCCCTGG CAGCCCTGGT CCTGATGGCA AACTGGCCCC	1200
CCCTGGTCCC GCCGGTCAAG ATGGTCGCCC CGGACCCCCA GGCCACCTG GTGCCCCGTG	1260
TCAGGCTGGT GTGATGGGAT TCCCTGGACC TAAAGGTGCT GCTGGAGAGC CCGGCAAGGC	1320
TGGAGAGCGA GGTGTTCCCG GACCCCCTGG CGCTGTCGGT CCTGCTGGCA AAGATGGAGA	1380
GGCTGGAGCT CAGGGACCCC CTGGCCCTGC TGGTCCCGCT GGCAGAGAGG GTGAACAAGG	1440
CCCTGCTGGC TCCCCCGGAT TCCAGGGTCT CCCTGGTCCT GCTGGTCCTC CAGGTGAAGC	1500
AGGCAAACCT GGTGAACAGG GTGTTCCTGG AGACCTTGGC GCCCCTGGCC CCTCTGGAGC	1560
AAGAGGCGAG AGAGGTTTCC CTGGCGAGCG TGGTGTGCAA GGTCCCCCTG GTCCTGCTGG	1620

ACCCCGAGGG GCCAACGGTG CTCCCGGCAA CGATGGTGCT AAGGGTGATG CTGGTGCCCC	1680
TGGAGCTCCC GGTAGCCAGG GCGCCCCTGG CCTTCAGGGA ATGCCTGGTG AACGTGGTGC	1740
AGCTGGTCTT CCAGGGCCTA AGGGTGACAG AGGTGATGCT GGTCCCAAAG GTGCTGATGG	1800
CTCTCCTGGC AAAGATGGCG TCCGTGGTCT GACCGGCCCC ATTGGTCCTC CTGGCCCTGC	1860
TGGTGCCCCCT GGTGACAAGG GTGAAAGTGG TCCCAGCGGC CCTGCTGGTC CCACTGGAGC	1920
TCGTGGTGCC CCCGGAGACC GTGGTGAGCC TGGTCCCCCC GGCCCTGCTG GCTTTGCTGG	1980
CCCCCCTGGT GCTGACGGCC AACCTGGTGC TAAAGGCGAA CCTGGTGATG CTGGTGCCAA	2040
AGGCGATGCT GGTCCCCCTG GGCCTGCCGG ACCCGCTGGA CCCCTGGGCC CCATTGGTAA	2100
TGTTGGTGCT CCTGGAGCCA AAGGTGCTCG CGGCAGCGCT GGTCCCCCTG GTGCTACTGG	2160
TTCCCTGGT GCTGCTGGCC GAGTCGGTCC TCCTGGCCCC TCTGGAAATG CTGGACCCCC	2220
TGGCCCTCCT GGTCTGCTG GCAAAGAAGG CGGCAAAGGT CCCCGTGGTG AGACTGGCCC	2280
TGCTGGACGT CCTGGTGAAG TTGGTCCCCC TGGTCCCCCT GGCCCTGCTG GCGAGAAAGG	2340
ATCCCCTGGT GCTGATGGTC CTGCTGGTGC TCCTGGTACT CCCGGGCCTC AAGGTATTGC	2400
TGGACAGCGT GGTGTGGTCG GCCTGCCTGG TCAGAGAGGA GAGAGAGGCT TCCCTGGTCT	2460
TCCTGGCCCC TCTGGTGAAC CTGGCAAACA AGGTCCCTCT GGAGCAAGTG GTGAACGTGG	2520
TCCCCCGGT CCCATGGGCC CCCCTGGATT GGCTGGACCC CCTGGTGAAT CTGGACGTGA	2580
GGGGGCTCCT GCTGCCGAAG GTTCCCCTGG ACGAGACGGT TCTCCTGGCG CCAAGGGTGA	2640

CCGTGGTGAG ACCGGCCCCG CTGGACCCCC TGGTGCTCCT GGTGCTCCTG GTGCCCCTGG	2700
CCCCGTTGGC CCTGCTGGCA AGAGTGGTGA TCGTGGTGAG ACTGGTCCTG CTGGTCCCGC	2760
CGGTCCCGTC GGCCCCGCTG GCGCCCCGTG CCGCGCCGGA CCCCAGGCC CCCGTGGTGA	2820
CAAGGGTGAG ACAGGCQAAC AGGGCGACAG AGGCATAAAG GGTCAACCTG GCTTCTCTGG	2880
CCTCCAGGGT CCCCCTGGCC CTCCTGGCTC TCCTGGTGAA CAAGGTCCCT CTGGAGCCTC	2940
TGGTCCTGCT GGTCCCCGAG GTCCCCCTGG CTCTGCTGGT GCTCCTGGCA AAGATGGACT	3000
CAACGGTCTC CCTGGCCCCA TTGGGCCCCC TGGTCCTCGC GGTGCACTG GTGATGCTGG	3060
TCCTGTTGGT CCCCCCGGCC CTCCTGGACC TCCTGGTCCC CCTGGTCCTC CCAGCGCTGG	3120
TTTCGACTTC AGCTTCCTCC CCCAGCCACC TCAAGAGAAG GCTCACGATG GTGGCCGCTA	3180
CTACCGGGCT AGATCTGCCC TGGACACCAA CTATTGCTTC AGCTCCACGG AGAAGAAGTG	3240
CTGCGTGCGG CAGCTGTACA TTGACTTCCG CAAGGACCTC GGCTGGAAGT GGATCCACGA	3300
GCCCAAGGGC TACCATGCCA ACTTCTGCCT CGGGCCCTGC CCCTACATTT GGAGCCTGGA	3360
CACGCAGTAC AGCAAGGTCC TGGCCCTGTA CAACCAGCAT AACCCGGGCG CCTCGGCGGC	3420
GCCGTGCTGC GTGCCGCAGG CGCTGGAGCC GCTGCCCATC GTGTACTACG TGGGCCGCAA	3480
GCCCAAGGTG GAGCAGCTGT CCAACATGAT CGTGCGCTCC TGCAAGTGCA GCTGATCTAG	3540
A	3541

(2) ANGABEN ZU SEQ ID NO:10:

(i) SEQUENZCHARAKTERISTIKA:

- (A) LÄNGE: 1388 Aminosäuren
- (B) ART: Aminosäure
- (C) STRANGFORM: Einzelstrang
- (D) TOPOLOGIE: unbekannt

(ii) ART DES MOLEKÜLS: Peptid

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO:10:

Gln Leu Ser Tyr Gly Tyr Asp Glu Lys Ser Thr Gly Gly Ile Ser Val
 1 5 10 15

Pro Gly Pro Met Gly Pro Ser Gly Pro Arg Gly Leu Pro Gly Pro Pro
 20 25 30

Gly Ala Pro Gly Pro Gln Gly Phe Gln Gly Pro Pro Gly Glu Pro Gly
 35 40 45

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

Pro Gly Lys Asn Gly Asp Asp Gly Glu Ala Gly Lys Pro Gly Arg Pro
 65 70 75 80

Gly Glu Arg Gly Pro Pro Gly Pro Gln Gly Ala Arg Gly Leu Pro Gly
 85 90 95

Thr Ala Gly Leu Pro Gly Met Lys Gly His Arg Gly Phe Ser Gly Leu
 100 105 110

Asp Gly Ala Lys Gly Asp Ala Gly Pro Ala Gly Pro Lys Gly Glu Pro
 115 120 125

Gly Ser Pro Gly Glu Asn Gly Ala Pro Gly Gln Met Gly Pro Arg Gly
 130 135 140

Leu Pro Gly Glu Arg Gly Arg Pro Gly Ala Pro Gly Pro Ala Gly Ala
 145 150 155 160

Arg Gly Asn Asp Gly Ala Thr Gly Ala Ala Gly Pro Pro Gly Pro Thr
 165 170 175

Gly Pro Ala Gly Pro Pro Gly Phe Pro Gly Ala Val Gly Ala Lys Gly
 180 185 190

Glu Ala Gly Pro Gln Gly Pro Arg Gly Ser Glu Gly Pro Gln Gly Val
 195 200 205

Arg Gly Glu Pro Gly Pro Pro Gly Pro Ala Gly Ala Ala Gly Pro Ala
 210 215 220

Gly Asn Pro Gly Ala Asp Gly Gln Pro Gly Ala Lys Gly Ala Asn Gly
 225 230 235 240

Ala Pro Gly Ile Ala Gly Ala Pro Gly Phe Pro Gly Ala Arg Gly Pro
 245 250 255

Ser Gly Pro Gln Gly Pro Gly Gly Pro Pro Gly Pro Lys Gly Asn Ser
 260 265 270

Gly Glu Pro Gly Ala Pro Gly Ser Lys Gly Asp Thr Gly Ala Lys Gly
 275 280 285

Glu Pro Gly Pro Val Gly Val Gln Gly Pro Pro Gly Pro Ala Gly Glu
 290 295 300

Glu Gly Lys Arg Gly Ala Arg Gly Glu Pro Gly Pro Thr Gly Leu Pro
 305 310 315 320

Gly Pro Pro Gly Glu Arg Gly Gly Pro Gly Ser Arg Gly Phe Pro Gly
 325 330 335

Ala Asp Gly Val Ala Gly Pro Lys Gly Pro Ala Gly Glu Arg Gly Ser
 340 345 350

Pro Gly Pro Ala Gly Pro Lys Gly Ser Pro Gly Glu Ala Gly Arg Pro
 355 360 365

Gly Glu Ala Gly Leu Pro Gly Ala Lys Gly Leu Thr Gly Ser Pro Gly
 370 375 380

Ser Pro Gly Pro Asp Gly Lys Thr Gly Pro Pro Gly Pro Ala Gly Gln
 385 390 395 400

Asp Gly Arg Pro Gly Pro Pro Gly Pro Pro Gly Ala Arg Gly Gln Ala
 405 410 415

Gly Val Met Gly Phe Pro Gly Pro Lys Gly Ala Ala Gly Glu Pro Gly
 420 425 430

Lys Ala Gly Glu Arg Gly Val Pro Gly Pro Pro Gly Ala Val Gly Pro
 435 440 445

Ala Gly Lys Asp Gly Glu Ala Gly Ala Gln Gly Pro Pro Gly Pro Ala
 450 455 460

Gly Pro Ala Gly Glu Arg Gly Glu Gln Gly Pro Ala Gly Ser Pro Gly
 465 470 475 480

Phe Gln Gly Leu Pro Gly Pro Ala Gly Pro Pro Gly Glu Ala Gly Lys
 485 490 495

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

Gly Ala Arg Gly Glu Arg Gly Phe Pro Gly Glu Arg Gly Val Gln Gly
 515 520 525

Pro Pro Gly Pro Ala Gly Pro Arg Gly Ala Asn Gly Ala Pro Gly Asn
 530 535 540

Asp Gly Ala Lys Gly Asp Ala Gly Ala Pro Gly Ala Pro Gly Ser Gln
 545 550 555 560

Gly Ala Pro Gly Leu Gln Gly Met Pro Gly Glu Arg Gly Ala Ala Gly
 565 570 575

Leu Pro Gly Pro Lys Gly Asp Arg Gly Asp Ala Gly Pro Lys Gly Ala
 580 585 590

Asp Gly Ser Pro Gly Lys Asp Gly Val Arg Gly Leu Thr Gly Pro Ile
 595 600 605

Gly Pro Pro Gly Pro Ala Gly Ala Pro Gly Asp Lys Gly Glu Ser Gly
 610 615 620

Pro Ser Gly Pro Ala Gly Pro Thr Gly Ala Arg Gly Ala Pro Gly Asp
 625 630 635 640

Arg Gly Glu Pro Gly Pro Pro Gly Pro Ala Gly Phe Ala Gly Pro Pro	645	650	655
Gly Ala Asp Gly Gln Pro Gly Ala Lys Gly Glu Pro Gly Asp Ala Gly	660	665	670
Ala Lys Gly Asp Ala Gly Pro Pro Gly Pro Ala Gly Pro Ala Gly Pro	675	680	685
Pro Gly Pro Ile Gly Asn Val Gly Ala Pro Gly Ala Lys Gly Ala Arg	690	695	700
Gly Ser Ala Gly Pro Pro Gly Ala Thr Gly Phe Pro Gly Ala Ala Gly	705	710	715
Arg Val Gly Pro Pro Gly Pro Ser Gly Asn Ala Gly Pro Pro Gly Pro	725	730	735
Pro Gly Pro Ala Gly Lys Glu Gly Gly Lys Gly Pro Arg Gly Glu Thr	740	745	750
Gly Pro Ala Gly Arg Pro Gly Glu Val Gly Pro Pro Gly Pro Pro Gly	755	760	765
Pro Ala Gly Glu Lys Gly Ser Pro Gly Ala Asp Gly Pro Ala Gly Ala	770	775	780
Pro Gly Thr Pro Gly Pro Gln Gly Ile Ala Gly Gln Arg Gly Val Val	785	790	795
Gly Leu Pro Gly Gln Arg Gly Glu Arg Gly Phe Pro Gly Leu Pro Gly	805	810	815

Pro Ser Gly Glu Pro Gly Lys Gln Gly Pro Ser Gly Ala Ser Gly Glu
820 825 830

Arg Gly Pro Pro Gly Pro Met Gly Pro Pro Gly Leu Ala Gly Pro Pro
835 840 845

Gly Glu Ser Gly Arg Glu Gly Ala Pro Gly Ala Glu Gly Ser Pro Gly
850 855 860

Arg Asp Gly Ser Pro Gly Ala Lys Gly Asp Arg Gly Glu Thr Gly Pro
865 870 875 880

Ala Gly Pro Pro Gly Ala Pro Gly Ala Pro Gly Ala Pro Gly Pro Val
885 890 895

Gly Pro Ala Gly Lys Ser Gly Asp Arg Gly Glu Thr Gly Pro Ala Gly
900 905 910

Pro Ala Gly Pro Val Gly Pro Ala Gly Ala Arg Gly Pro Ala Gly Pro
915 920 925

Gln Gly Pro Arg Gly Asp Lys Gly Glu Thr Gly Glu Gln Gly Asp Arg
930 935 940

Gly Ile Lys Gly His Arg Gly Phe Ser Gly Leu Gln Gly Pro Pro Gly
945 950 955 960

Pro Pro Gly Ser Pro Gly Glu Gln Gly Pro Ser Gly Ala Ser Gly Pro
965 970 975

Ala Gly Pro Arg Gly Pro Pro Gly Ser Ala Gly Ala Pro Gly Lys Asp
980 985 990

Gly Leu Asn Gly Leu Pro Gly Pro Ile Gly Pro Pro Gly Pro Arg Gly

995

1000

1005

Arg Thr Gly Asp Ala Gly Pro Val Gly Pro Pro Gly Pro Pro Gly Pro

1010

1015

1020

Pro Gly Pro Pro Gly Pro Pro Ser Ala Gly Phe Asp Phe Ser Phe Leu

1025

1030

1035

1040

Pro Gln Pro Pro Gln Glu Lys Ala His Asp Gly Gly Arg Tyr Tyr Arg

1045

1050

1055

Ala Arg Ser Asp Glu Ala Ser Gly Ile Gly Pro Glu Val Pro Asp Asp

1060

1065

1070

Arg Asp Phe Glu Pro Ser Leu Gly Pro Val Cys Pro Phe Arg Cys Gln

1075

1080

1085

Cys His Leu Arg Val Val Gln Cys Ser Asp Leu Gly Leu Asp Lys Val

1090

1095

1100

Pro Lys Asp Leu Pro Pro Asp Thr Thr Leu Leu Asp Leu Gln Asn Asn

1105

1110

1115

1120

Lys Ile Thr Glu Ile Lys Asp Gly Asp Phe Lys Asn Leu Lys Asn Leu

1125

1130

1135

His Ala Leu Ile Leu Val Asn Asn Lys Ile Ser Lys Val Ser Pro Gly

1140

1145

1150

Ala Phe Thr Pro Leu Val Lys Leu Glu Arg Leu Tyr Leu Ser Lys Asn

1155

1160

1165

Gln Leu Lys Glu Leu Pro Glu Lys Met Pro Lys Thr Leu Gln Glu Leu
 1170 1175 1180

Arg Ala His Glu Asn Glu Ile Thr Lys Val Arg Lys Val Thr Phe Asn
 1185 1190 1195 1200

Gly Leu Asn Gln Met Ile Val Ile Glu Leu Gly Thr Asn Pro Leu Lys
 1205 1210 1215

Ser Ser Gly Ile Glu Asn Gly Ala Phe Gln Gly Met Lys Lys Leu Ser
 1220 1225 1230

Tyr Ile Arg Ile Ala Asp Thr Asn Ile Thr Ser Ile Pro Gln Gly Leu
 1235 1240 1245

Pro Pro Ser Leu Thr Glu Leu His Leu Asp Gly Asn Lys Ile Ser Arg
 1250 1255 1260

Val Asp Ala Ala Ser Leu Lys Gly Leu Asn Asn Leu Ala Lys Leu Gly
 1265 1270 1275 1280

Leu Ser Phe Asn Ser Ile Ser Ala Val Asp Asn Gly Ser Leu Ala Asn
 1285 1290 1295

Thr Pro His Leu Arg Glu Leu His Leu Asp Asn Asn Lys Leu Thr Arg
 1300 1305 1310

Val Pro Gly Gly Leu Ala Glu His Lys Tyr Ile Gln Val Val Tyr Leu
 1315 1320 1325

His Asn Asn Asn Ile Ser Val Val Gly Ser Ser Asp Phe Cys Pro Pro
 1330 1335 1340

Gly His Asn Thr Lys Lys Ala Ser Tyr Ser Gly Val Ser Leu Phe Ser
1345 1350 1355 1360

Asn Pro Val Gln Tyr Trp Glu Ile Gln Pro Ser Thr Phe Arg Cys Val
1365 1370 1375

Tyr Val Arg Ser Ala Ile Gln Leu Gly Asn Tyr Lys
1380 1385

(2) ANGABEN ZU SEQ ID NO:11:

(i) SEQUENZCHARAKTERISTIKA:

- (A) LÄNGE: 1107 Aminosäuren
- (B) ART: Aminosäure
- (C) STRANGFORM: Einzelstrang
- (D) TOPOLOGIE: unbekannt

(ii) ART DES MOLEKÜLS: Peptid

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO:11:

Gln Leu Ser Tyr Gly Tyr Asp Glu Lys Ser Thr Gly Gly Ile Ser Val
1 5 10 15

Pro Gly Pro Met Gly Pro Ser Gly Pro Arg Gly Leu Pro Gly Pro Pro
20 25 30

Gly Ala Pro Gly Pro Gln Gly Phe Gln Gly Pro Pro Gly Glu Pro Gly
35 40 45

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

Pro Gly Lys Asn Gly Asp Asp Gly Glu Ala Gly Lys Pro Gly Arg Pro
 65 70 75 80

Gly Glu Arg Gly Pro Pro Gly Pro Gln Gly Ala Arg Gly Leu Pro Gly
 85 90 95

Thr Ala Gly Leu Pro Gly Met Lys Gly His Arg Gly Phe Ser Gly Leu
 100 105 110

Asp Gly Ala Lys Gly Asp Ala Gly Pro Ala Gly Pro Lys Gly Glu Pro
 115 120 125

Gly Ser Pro Gly Glu Asn Gly Ala Pro Gly Gln Met Gly Pro Arg Gly
 130 135 140

Leu Pro Gly Glu Arg Gly Arg Pro Gly Ala Pro Gly Pro Ala Gly Ala
 145 150 155 160

Arg Gly Asn Asp Gly Ala Thr Gly Ala Ala Gly Pro Pro Gly Pro Thr
 165 170 175

Gly Pro Ala Gly Pro Pro Gly Phe Pro Gly Ala Val Gly Ala Lys Gly
 180 185 190

Glu Ala Gly Pro Gln Gly Pro Arg Gly Ser Glu Gly Pro Gln Gly Val
 195 200 205

Arg Gly Glu Pro Gly Pro Pro Gly Pro Ala Gly Ala Ala Gly Pro Ala
 210 215 220

Gly Asn Pro Gly Ala Asp Gly Gln Pro Gly Ala Lys Gly Ala Asn Gly
 225 230 235 240

Ala Pro Gly Ile Ala Gly Ala Pro Gly Phe Pro Gly Ala Arg Gly Pro
245 250 255

Ser Gly Pro Gln Gly Pro Gly Gly Pro Pro Gly Pro Lys Gly Asn Ser
260 265 270

Gly Glu Pro Gly Ala Pro Gly Ser Lys Gly Asp Thr Gly Ala Lys Gly
275 280 285

Glu Pro Gly Pro Val Gly Val Gln Gly Pro Pro Gly Pro Ala Gly Glu
290 295 300

Glu Gly Lys Arg Gly Ala Arg Gly Glu Pro Gly Pro Thr Gly Leu Pro
305 310 315 320

Gly Pro Pro Gly Glu Arg Gly Gly Pro Gly Ser Arg Gly Phe Pro Gly
325 330 335

Ala Asp Gly Val Ala Gly Pro Lys Gly Pro Ala Gly Glu Arg Gly Ser
340 345 350

Pro Gly Pro Ala Gly Pro Lys Gly Ser Pro Gly Glu Ala Gly Arg Pro
355 360 365

Gly Glu Ala Gly Leu Pro Gly Ala Lys Gly Leu Thr Gly Ser Pro Gly
370 375 380

Ser Pro Gly Pro Asp Gly Lys Thr Gly Pro Pro Gly Pro Ala Gly Gln
385 390 395 400

Asp Gly Arg Pro Gly Pro Pro Gly Pro Pro Gly Ala Arg Gly Gln Ala
405 410 415

Gly Val Met Gly Phe Pro Gly Pro Lys Gly Ala Ala Gly Glu Pro Gly			
420	425	430	
Lys Ala Gly Glu Arg Gly Val Pro Gly Pro Pro Gly Ala Val Gly Pro			
435	440	445	
Ala Gly Lys Asp Gly Glu Ala Gly Ala Gln Gly Pro Pro Gly Pro Ala			
450	455	460	
Gly Pro Ala Gly Glu Arg Gly Glu Gln Gly Pro Ala Gly Ser Pro Gly			
465	470	475	480
Phe Gln Gly Leu Pro Gly Pro Ala Gly Pro Pro Gly Glu Ala Gly Lys			
485	490	495	
Pro Gly Glu Gln Gly Val Pro Gly Asp Leu Gly Ala Pro Gly Pro Ser			
500	505	510	
Gly Ala Arg Gly Glu Arg Gly Phe Pro Gly Glu Arg Gly Val Gln Gly			
515	520	525	
Pro Pro Gly Pro Ala Gly Pro Arg Gly Ala Asn Gly Ala Pro Gly Asn			
530	535	540	
Asp Gly Ala Lys Gly Asp Ala Gly Ala Pro Gly Ala Pro Gly Ser Gln			
545	550	555	560
Gly Ala Pro Gly Leu Gln Gly Met Pro Gly Glu Arg Gly Ala Ala Gly			
565	570	575	
Leu Pro Gly Pro Lys Gly Asp Arg Gly Asp Ala Gly Pro Lys Gly Ala			
580	585	590	

Asp Gly Ser Pro Gly Lys Asp Gly Val Arg Gly Leu Thr Gly Pro Ile
 595 600 605

Gly Pro Pro Gly Pro Ala Gly Ala Pro Gly Asp Lys Gly Glu Ser Gly
 610 615 620

Pro Ser Gly Pro Ala Gly Pro Thr Gly Ala Arg Gly Ala Pro Gly Asp
 625 630 635 640

Arg Gly Glu Pro Gly Pro Pro Gly Pro Ala Gly Phe Ala Gly Pro Pro
 645 650 655

Gly Ala Asp Gly Gln Pro Gly Ala Lys Gly Glu Pro Gly Asp Ala Gly
 660 665 670

Ala Lys Gly Asp Ala Gly Pro Pro Gly Pro Ala Gly Pro Ala Gly Pro
 675 680 685

Pro Gly Pro Ile Gly Asn Val Gly Ala Pro Gly Ala Lys Gly Ala Arg
 690 695 700

Gly Ser Ala Gly Pro Pro Gly Ala Thr Gly Phe Pro Gly Ala Ala Gly
 705 710 715 720

Arg Val Gly Pro Pro Gly Pro Ser Gly Asn Ala Gly Pro Pro Gly Pro
 725 730 735

Pro Gly Pro Ala Gly Lys Glu Gly Gly Lys Gly Pro Arg Gly Glu Thr
 740 745 750

Gly Pro Ala Gly Arg Pro Gly Glu Val Gly Pro Pro Gly Pro Pro Gly
 755 760 765

Pro Ala Gly Glu Lys Gly Ser Pro Gly Ala Asp Gly Pro Ala Gly Ala
 770 775 780

Pro Gly Thr Pro Gly Pro Gln Gly Ile Ala Gly Gln Arg Gly Val Val
 785 790 795 800

Gly Leu Pro Gly Gln Arg Gly Glu Arg Gly Phe Pro Gly Leu Pro Gly
 805 810 815

Pro Ser Gly Glu Pro Gly Lys Gln Gly Pro Ser Gly Ala Ser Gly Glu
 820 825 830

Arg Gly Pro Pro Gly Pro Met Gly Pro Pro Gly Leu Ala Gly Pro Pro
 835 840 845

Gly Glu Ser Gly Arg Glu Gly Ala Pro Gly Ala Glu Gly Ser Pro Gly
 850 855 860

Arg Asp Gly Ser Pro Gly Ala Lys Gly Asp Arg Gly Glu Thr Gly Pro
 865 870 875 880

Ala Gly Pro Pro Gly Ala Pro Gly Ala Pro Gly Ala Pro Gly Pro Val
 885 890 895

Gly Pro Ala Gly Lys Ser Gly Asp Arg Gly Glu Thr Gly Pro Ala Gly
 900 905 910

Pro Ala Gly Pro Val Gly Pro Ala Gly Ala Arg Gly Pro Ala Gly Pro
 915 920 925

Gln Gly Pro Arg Gly Asp Lys Gly Glu Thr Gly Glu Gln Gly Asp Arg
 930 935 940

Gly Ile Lys Gly His Arg Gly Phe Ser Gly Leu Gln Gly Pro Pro Gly
 945 950 955 960

Pro Pro Gly Ser Pro Gly Glu Gln Gly Pro Ser Gly Ala Ser Gly Pro
 965 970 975

Ala Gly Pro Arg Gly Pro Pro Gly Ser Ala Gly Ala Pro Gly Lys Asp
 980 985 990

Gly Leu Asn Gly Leu Pro Gly Pro Ile Gly Pro Pro Gly Pro Arg Gly
 995 1000 1005

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

Pro Gly Pro Pro Gly Pro Pro Ser Ala Gly Phe Asp Phe Ser Phe Leu
 1025 1030 1035 1040

Pro Gln Pro Pro Gln Glu Lys Ala His Asp Gly Gly Arg Tyr Tyr Arg
 1045 1050 1055

Ala Arg Ser Pro Lys Asp Leu Pro Pro Asp Thr Thr Leu Leu Asp Leu
 1060 1065 1070

Gln Asn Asn Lys Ile Thr Glu Ile Lys Asp Gly Asp Phe Lys Asn Leu
 1075 1080 1085

Lys Asn Leu His Ala Leu Ile Leu Val Asn Asn Lys Ile Ser Lys Val
 1090 1095 1100

Ser Pro Gly
 1105

(2) ANGABEN ZU SEQ ID NO:12:

(i) SEQUENZCHARAKTERISTIKA:

- (A) LÄNGE: 4167 Basenpaare
- (B) ART: Nukleinsäure
- (C) STRANGFORM: Einzelstrang
- (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: cDNA

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO:12:

```

CAGCTGTCTT ATGGCTATGA TGAGAAATCA ACCGGAGGAA TTTCCGTGCC TGGCCCCATG      60
GGTCCCTCTG GTCCTCGTGG TCTCCCTGGC CCCCCTGGTG CACCTGGTCC CCAAGGCTTC      120
CAAGGTCCCC CTGGTGAGCC TGGCGAGCCT GGAGCTTCAG GTCCCATGGG TCCCCGAGGT      180
CCCCCAGGTC CCCCTGGAAA GAATGGAGAT GATGGGGAAG CTGGAAAACC TGGTCGTCTT      240
GGTGAGCGTG GGCCTCCTGC GCCTCAGGGT GCTCGAGGAT TGCCCGGAAC AGCTGGCCTC      300
CCTGGAATGA AGGGACACAG AGGTTTCAGT GGTTCGGATG GTGCCAAGGG AGATGCTGGT      360
CCTGCTGGTC CTAAGGGTGA GCCTGGCAGC CCTGGTGAAA ATGGAGCTCC TGGTCAGATG      420
GGCCCCCGTG GCCTGCCTGG TGAGAGAGGT CGCCCTGGAG CCCCTGGCCC TGCTGGTGCT      480
CGTGGAATG ATGGTGCTAC TGGTGCTGCC GGGCCCCCTG GTCCCACCGG CCCCCTGGT      540
CCTCCTGGCT TCCCTGGTGC TGTTGGTGCT AAGGGTGAAG CTGGTCCCCA AGGGCCCCGA      600

```

GGCTCTGAAG GTCCCCAGGG TGTGCGTGGT GAGCCTGGCC CCCCTGGCCC TGCTGGTGCT	660
GCTGGCCCTG CTGGAAACCC TGGTGCTGAT GGACAGCCTG GTGCTAAAGG TGCCAATGGT	720
GCTCCTGGTA TTGCTGGTGC TCCTGGCTTC CCTGGTGCCC GAGGCCCCCTC TGGACCCCAG	780
GGCCCCGGCG GCCCTCCTGG TCCCAAGGGT AACAGCGGTG AACCTGGTGC TCCTGGCAGC	840
AAAGGAGACA CTGGTGCTAA GGGAGAGCCT GGCCCTGTTG GTGTTCAAGG ACCCCCTGGC	900
CCTGCTGGAG AGCAAGGAAA GCGAGGAGCT CGAGGTGAAC CCGGACCCAC TGGCCTGCCC	960
GGACCCCCCTG GCGAGCGTGG TGGACCTGGT AGCCGTGGTT TCCCTGGCGC AGATGGTGTT	1020
GCTGGTCCCA AGGGTCCCGC TGGTGAACGT GGTTCCTCTG GCCCCGCTGG CCCCAAAGGA	1080
TCTCCTCGTG AAGCTGGTCG TCCCGGTGAA GCTGGTCTGC CTGGTGCCAA GGGTCTGACT	1140
GGAAGCCCTG GCAGCCCTGG TCCTGATGGC AAAACTGGCC CCCCTGGTCC CGCCGGTCAA	1200
GATGGTCGCC CCGGACCCCC AGGCCCACCT GGTGCCCCGTG GTCAGGCTGG TGTGATGGGA	1260
TTCCCTGGAC CTAAAGGTGC TGCTCGAGAG CCCGCAAGG CTGGAGAGCG AGGTGTTCCC	1320
GGACCCCCCTC GCGCTGTGCG TCCTGCTGGC AAAGATGGAG AGGCTGGAGC TCAGGGACCC	1380
CCTGGCCCTG CTGGTCCCGC TGGCGAGAGA GGTGAACAAG GCCCTGCTGG CTCCCCCGGA	1440
TTCCAGGGTC TCCCTGGTCC TGCTGGTCCT CCAGGTGAAG CAGGCAAACC TGGTGAACAG	1500
GGTGTTCTCTG GAGACCTTGG CGCCCCCTGGC CCCTCTGGAG CAAGAGGCGA GAGAGGTTTC	1560
CCTGGCGAGC GTGGTGTGCA AGGTCCCCCT GGTCTGCTG GACCCCGAGG GGCCAACGGT	1620

GCTCCCGCCA ACGATGCTGC TAAGGGTGAT GCTGGTGCCC CTGGAGCTCC CGGTAGCCAG	1680
GGCGCCCCTG GCCTTCAGGG AATGCCTGGT GAACGTGGTG CAGCTGGTCT TCCAGGGCCT	1740
AAGGGTGACA GAGGTGATGC TGGTCCCAA GGTGCTGATG GCTCTCCTGG CAAAGATGGC	1800
GTCCGTGGTC TGACCGACCC CATTGGTCCT CCTGGCCCTG CTGGTGCCCC TGGTGACAAG	1860
GGTGAAAGTG GTCCCAGCGG CCCTGCTGGT CCCACTGGAG CTCGTGGTGC CCCC GGAGAC	1920
CGTGGTGAGC CTGGTCCCCC CGGCCCTGCT GGCTTTGCTG GCCCCCTGG TGCTGACGGC	1980
CAACCTGGTG CTAAAGGCGA ACCTGGTGAT GCTGGTGCCA AAGGCGATGC TGGTCCCCCT	2040
GGGCCTGCCG GACCCGCTGG ACCCCCTGGC CCCATTGGTA ATGTTGGTGC TCCTGGAGCC	2100
AAACGTGCTC GCGGCAGCGC TGGTCCCCCT GGTGCTACTG GTTCCCTGG TGCTGCTGGC	2160
CGAGTCGGTC CTCCTGGCCC CTCTGGAAAT GCTGGACCCC CTGGCCCTCC TGGTCCTGCT	2220
GGCAAAGAAG GCGGCAAAGG TCCCCGTGGT GAGACTGGCC CTGCTGGACG TCCTGGTGAA	2280
GTTGGTCCCC CTGGTCCCCC TGGCCCTGCT GGCGAGAAAG GATCCCCCTGG TGCTGATGGT	2340
CCTGCTGGTG CTCCTGGTAC TCCCGGGCCT CAAGGTATTG CTGGACAGCG TGGTGTGGTC	2400
GGCCTGCCTG GTCAGAGAGG AGAGAGAGGC TTCCCTGGTC TTCTTGGCCC CTCTGGTGAA	2460
CCTGGCAAAC AAGGTCCCTC TGGAGCAAGT GGTGAACGTG GTCCCCCCGG TCCCATGGGC	2520
CCCCCTGGAT TGGCTGGACC CCCTGGTGAA TCTGGACGTG AGGGGGCTCC TGCTGCCGAA	2580
GGTCCCCCTG GACGAGACGG TTCTCCTGGC GCCAAGGGTG ACCGTGGTGA GACCGGCCCC	2640

GCTGGACCCC CTGGTGCTCC TGGTGCTCCT GGTGCCCCCTG GCCCCGTTGG CCCTGCTGGC	2700
AAGAGTGGTG ATCGTGGTGA GACTGGTCCT GCTGGTCCCG CCGGTCCCGT CGGCCCCGCT	2760
GGCGCCCCGTG GCCCCGCCGG ACCCCAAGGC CCCCCTGGTG ACAAGGGTGA GACAGGCGAA	2820
CAGGGCGACA GAGGCATAAA GGGTCACCGT GGCTTCTCTG GCCTCCAGGG TCCCCCTGGC	2880
CCTCCTGGCT CTCCTGGTGA ACAAGGTCCC TCTGGAGCCT CTGGTCCTGC TGGTCCCCGA	2940
GGTCCCCCTG GCTCTGCTGG TGCTCCTGGC AAAGATGGAC TCAACGGTCT CCCTGGCCCC	3000
ATTGGGCCCC CTGGTCCTCG CGGTGCACT GGTGATGCTG GTCCTGTTGG TCCCCCGGC	3060
CCTCCTGGAC CTCCTGGTCC CCCTGGTCCT CCCAGGCTG GTTTCGACTT CAGCTTCCTC	3120
CCCCAGCCAC CTCAAGAGAA GGCTCAGAT GGTGGCCGCT ACTACGGGC TAGATCCGAT	3180
GAGGCTTCTG GGATAGCCCC AGAAGTTCCT GATGACCGC ACTTCGAGCC CTCCCTAGGC	3240
CCAGTGTGCC CCTTCCGCTG TCAATGCCAT CTTGAGTGG TCCAGTGTTT TGATTGGGT	3300
CTGGACAAAG TGCCAAAGGA TCTTCCCCCT GACACAATC TGCTAGACCT GCAAAACAAC	3360
AAAATAACCG AAATCAAAGA TGGAGACTTT AAGAACCTGA AGAACCTTCA CGCATTGATT	3420
CTTGTCAACA ATAAAATTAG CAAAGTTAGT CCTGGAGCAT TTACACCTTT GGTGAAGTTG	3480
GAACGACTTT ATCTGTCCAA GAATCAGCTG AAGGAATTGC CAGAAAAAAT GCCCAAAC	3540
CTTCAGGAGC TGCCTGCCCA TGAGAATGAG ATCACCAGG TCGGAAAAGT TACTTTCAAT	3600
GGACTGAACC AGATGATTGT CATAGAACTG GGCACCAATC CGCTGAAGAG CTCAGGAATT	3660

GAAAATGGGG CTTTCCAGGG AATGAAGAAG CTCTCCTACA TCCGCATTGC TGATACCAAT	3720
ATCACCAGCA TTCCTCAAGG TCTTCCTCCT TCCCTTACGG AATTACATCT TGATGGCAAC	3780
AAAATCAGCA GAGTTGATGC AGCTAGCCTG AAAGGACTGA ATAATTTGGC TAAGTTGGGA	3840
TTGAGTTTCA ACAGCATCTC TGCTGTTGAC AATGGCTCTC TGGCCAACAC GCCTCATCTG	3900
AGGGAGCTTC ACTTGGACAA CAACAAGCTT ACCAGAGTAC CTGGTGGGCT GGCAGAGCAT	3960
AAGTACATCC AGGTTGTCTA CCTTCATAAC AACAAATATCT CTGTAGTTGG ATCAAGTGAC	4020
TTCTGCCCCAC CTGGACACAA CACCAAAAAG GCTTCTTATT CGGGTGTGAG TCTTTTCAGC	4080
AACCCGGTCC AGTACTGGGA GATACAGCCA TCCACCTTCA GATGTGTCTA CGTGCGCTCT	4140
GCCATTCAAC TCGGAAACTA TAAGTAA	4167

(2) ANGABEN ZU SEQ ID NO:13:

(i) SEQUENZCHARAKTERISTIKA:

- (A) LÄNGE: 3349 Basenpaare
- (B) ART: Nukleinsäure
- (C) STRANGFORM: Einzelstrang
- (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: cDNA

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO:13:

GGGAAGGATT TCCATTTCCTC AGCTGTCTTA TGGCTATGAT GAGAAATCAA CCGGAGGAAT	60

TTCCGTGCCT GGCCCCATGG GTCCCTCTGG TCCTCGTGGT CTCCCTGGCC CCCCTGGTGC	120
ACCTGGTCCC CAAGGCTTCC AAGGTCCCCC TGGTGAGCCT GGCGAGCCTG GAGCTTCAGG	180
TCCCATGGGT CCCCAGAGTC CCCCAGGTCC CCCTGGAAAG AATGGAGATG ATGGGGAAGC	240
TGGAAAACCT GGTCGTCTTG GTGAGCGTGG GCCTCCTGGG CTCAGGGTG CTCGAGGATT	300
GCCCGAACA GCTGGCCTCC CTGGAATGAA GGGACACAGA GGTTTCAGTG GTTTGGATGG	360
TGCCAAGGGA GATGCTGGTC CTGCTGGTCC TAAGGGTGAG CCTGGCAGCC CTGGTGAAAA	420
TGGAGCTCCT GGTCAGATGG GCCCCCGTGG CCTGCCTGGT GAGAGAGGTC GCCCTGGAGC	480
CCCTGGCCCT GCTGGTGCTC GTGGAAATGA TGGTGCTACT GGTGCTGCCG GGCCCCCTGG	540
TCCCACCGGC CCCGCTGGTC CTCCTGGCTT CCCTGGTGCT GTTGGTGCTA AGGGTGAAGC	600
TGGTCCCCAA GGGCCCCGAG GCTCTGAAGG TCCCCAGGGT GTGCGTGGTG AGCCTGGCCC	660
CCCTGGCCCT GCTGGTGCTG CTGGCCCTGC TGGAAACCCT GGTGCTGATG GACAGCCTGG	720
TGCTAAAGGT GCCAATGGTG CTCCTGGTAT TGCTGGTGCT CCTGGCTTCC CTGGTGCCCCG	780
AGGCCCCCTCT GGACCCCAGG GCCCCGGCGG CCCTCCTGGT CCAAGGGTA ACAGCGGTGA	840
ACCTGGTGCT CCTGGCAGCA AAGGAGACAC TGGTGCTAAG GGAGAGCCTG GCCCTGTTGG	900
TGTTCAAGGA CCCCCTGGCC CTGCTGGAGA GGAAGGAAAG CGAGGAGCTC GAGGTGAACC	960
CGGACCCACT GGCCTGCCCC GACCCCCTGG CGAGCGTGGT GGACCTGGTA GCCGTGGTTT	1020
CCCTGGCGCA GATGGTGTTG CTGGTCCCAA GGGTCCCGCT GGTGAACGTG GTTCTCCTGG	1080

CCCCGCTGGC CCCAAAGGAT CTCCTGGTGA AGCTGGTCGT CCCGGTGAAG CTGGTCTGCC	1140
TGGTGCCAAG GGTCTGACTG GAAGCCCTGG CAGCCCTGGT CCTGATGGCA AAACCTGGCCC	1200
CCCTGGTCCC GCCGGTCAAG ATGGTCGCCC CGGACCCCCA GGCCACCTG GTGCCCCTGG	1260
TCAGGCTGGT GTGATGGGAT TCCCTGGACC TAAAGGTGCT GCTGGAGAGC CCGGCAAGGC	1320
TGGAGAGCGA GGTGTTCCCG GACCCCTGG CGCTGTCGGT CCTGCTGGCA AAGATGGAGA	1380
GGCTGGAGCT CAGGGACCCC CTGGCCCTGC TGGTCCCGCT GGCGAGAGAG GTGAACAAGG	1440
CCCTGCTGGC TCCCCCGGAT TCCAGGGTCT CCCTGGTCCT GCTGGTCCTC CAGGTGAAGC	1500
AGGCAAACCT GGTGAACAGG GTGTTCTTGG AGACCTTGGC GCCCCTGGCC CCTCTGGAGC	1560
AAGAGGCGAG AGAGGTTTCC CTGGCGAGCG TGGTGTGCAA GGTCCCCCTG GTCCTGCTGG	1620
ACCCCGAGGG GCCAACGGTG CTCCCGGCAA CGATGGTGCT AAGGGTGATG CTGGTGCCCC	1680
TGGAGCTCCC GGTAGCCAGG GCGCCCCCTGG CCTTCAGGGA ATGCCTGGTG AACGTGGTGC	1740
AGCTGGTCTT CCAGGGCCTA AGGGTGACAG AGGTGATGCT GGTCCCAAAG GTGCTGATGG	1800
CTCTCCTGGC AAAGATGGCG TCCGTGGTCT GACCGGCCCC ATTGGTCCTC CTGGCCCTGC	1860
TGGTGCCCCCT GGTGACAAGG GTGAAAGTGG TCCCAGCGGC CCTGCTGGTC CCACTGGAGC	1920
TCGTGGTGCC CCCGGAGACC GTGGTGAGCC TGGTCCCCCC GGCCCTGCTG GCTTTGCTGG	1980
CCCCCCTGGT GCTGACGGCC AACCTGGTGC TAAAGGCGAA CCTGGTGATG CTGGTGCCAA	2040
AGGCGATGCT GGTCCCCCTG GGCCTGCCGG ACCCGCTGGA CCCCCTGGCC CCATTGGTAA	2100

TGTTGGTGCT CCTGGAGCCA AAGGTGCTCG CGGCAGCGCT GGTCCCCCTG GTGCTACTGG	2160
TTTCCCTGGT GCTGCTGGCC GAGTCGGTCC TCCTGGCCCC TCTGGAAATG CTGGACCCCC	2220
TGGCCCTCCT GGTCTGCTG GCAAAGAAGG CGGCAAAGGT CCCCCTGGTG AACTTGGCCC	2280
TGCTGGACGT CCTGGTGAAG TTGGTCCCCC TGGTCCCCCT GGCCCTGCTG GCGAGAAAGG	2340
ATCCCCTGGT GCTGATGGTC CTGCTGGTGC TCCTGGTACT CCCGGGCCTC AAGGTATTGC	2400
TGGACAGCGT GGTGTGGTCG GCCTGCCTGG TCAGAGAGGA GAGAGAGGCT TCCCTGGTCT	2460
TCCTGGCCCC TCTGGTGAAC CTGGCAAACA AGGTCCCTCT GGAGCAAGTG GTGAACGTGG	2520
TCCCCCGGT CCCATGGGCC CCCCTGGATT GGCTGGACCC CCTGGTGAAT CTGGACGTGA	2580
GGGGGCTCCT GCTGCCGAAG GTTCCCCTGG ACGAGACGGT TCTCCTGGCG CCAAGGGTGA	2640
CCGTGGTGAG ACCGGCCCCG CTGGACCCCC TGGTGCTCCT GGTGCTCCTG GTGCCCCTGG	2700
CCCCGTGGC CCTGCTGGCA AGAGTGGTGA TCGTGGTGAG ACTGGTCCTG CTGGTCCCGC	2760
CGGTCCCGTC GGCCCCGCTG GCGCCCCGTG CCCC GCCGGA CCCC AAGGCC CCCGTGGTGA	2820
CAAGGGTGAG ACAGGCGAAC AGGGCGACAG AGGCATAAAG GGTCAACCGTG GCTTCTCTGG	2880
CCTCCAGGGT CCCCCTGGCC CTCCTGGCTC TCCTGGTGAA CAAGGTCCCT CTGGAGCCTC	2940
TGGTCTGCT GGTCCCCGAG GTCCCCCTGG CTCTGCTGGT GCTCCTGGCA AAGATGGACT	3000
CAACGGTCTC CCTGGCCCCA TTGGGCCCCC TGGTCCTCGC GGTGCGACTG GTGATGCTGG	3060
TCCTGTTGGT CCCCCCGGCC CTCCTGGACC TCCTGGTCCC CCTGGTCCTC CCAGCGCTGG	3120

TTTCGACTTC AGCTTCCTCC CCCAGCCACC TCAAGAGAAG GCTCACGATG GTGGCCGCTA 3180

CTACCGGGCT AGATCTCCAA AGGATCTTCC CCCTGACACA ACTCTGCTAG ACCTGCAAAA 3240

CAACAAAATA ACCGAAATCA AAGATGGAGA CTTTAAGAAC CTGAAGAACC TTCACGCATT 3300

GATTCTTGTC AACAATAAAA TTAGCAAAGT TAGTCCTGGA TAACTGCAG 3349

(2) ANGABEN ZU SEQ ID NO:14:

(i) SEQUENZCHARAKTERISTIKA:

- (A) LÄNGE: 57 Basenpaare
- (B) ART: Nukleinsäure
- (C) STRANGFORM: Einzelstrang
- (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: cDNA

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO:14:

ATCGAGGGAA GGATTTGAGA ATTCGGATCC TCTAGAGTCG ACCTGCAGGC AAGCTTG 57

(2) ANGABEN ZU SEQ ID NO:15:

(i) SEQUENZCHARAKTERISTIKA:

- (A) LÄNGE: 3171 Basenpaare
- (B) ART: Nukleinsäure
- (C) STRANGFORM: Einzelstrang
- (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: cDNA

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO:15:

CAGCTGTCTT ATGGCTATGA TGAGAAATCA ACCGGAGGAA TTTCCGTGCC TGGCCCCATG	60
GGTCCCTCTG GTCCTCGTGG TCTCCCTGGC CCCCTGGTG CACCTGGTCC CCAAGGCTTC	120
CAAGGTCCCC CTGGTGAGCC TGGCGAGCCT GGAGCTTCAG GTCCCATGGG TCCCCGAGGT	180
CCCCCAGGTC CCCCTGGAAA GAATGGAGAT GATGGGGAAG CTGGAAAACC TGGTCGTCCT	240
GGTGAGCGTG GGCCTCCTGG GCCTCAGGGT GCTCGAGGAT TGCCCGGAAC AGCTGGCCTC	300
CCTGGAATGA AGGGACACAG AGGTTTCAGT GGTTCGGATG GTGCCAAGGG AGATGCTGGT	360
CCTGCTGGTC CTAAGGGTGA GCCTGGCAGC CCTGGTGAAA ATGGAGCTCC TGGTCAGATG	420
GGCCCCCGTG GCCTGCCTGG TGAGAGAGGT CGCCCTGGAG CCCCTGGCCC TGCTGGTGCT	480
CGTGAAATG ATGGTGCTAC TGGTGCTGCC GGGCCCCCTG GTCCCACCGG CCCCCTGGT	540
CCTCCTGGCT TCCCTGGTGC TGTGGTGCT AAGGGTGAAG CTGGTCCCCA AGGGCCCCGA	600
GGCTCTGAAG GTCCCCAGGG TGTGCGTGGT GAGCCTGGCC CCCCTGGCCC TGCTGGTGCT	660
GCTGGCCCTG CTGGAACCC TGGTGCTGAT GGACAGCCTG GTGCTAAAGG TGCCAATGGT	720
GCTCCTGGTA TTGCTGGTGC TCCTGGCTTC CCTGGTGCCC GAGGCCCCCTC TGGACCCCAG	780
GGCCCCGGCG GCCCTCCTGG TCCCAAGGGT AACAGCGGTG AACCTGGTGC TCCTGGCAGC	840
AAAGGAGACA CTGGTGCTAA GGGAGAGCCT GGCCCTGTTG GTGTTCAAGG ACCCCCTGGC	900

CCTGCTGGAG AGGAAGGAAA GCGAGGAGCT CGAGGTGAAC CCGGACCCAC TGGCCTGCCC	960
GGACCCCTG GCGAGCGTGG TGGACCTGGT AGCCGTGGTT TCCCTGGCGC AGATGGTGTT	1020
GCTGGTCCCA AGGGTCCCGC TGGTGAACGT GGTTCCTCTG GCCCCGCTGG CCCCAAAGGA	1080
TCTCCTGGTG AAGCTGGTCG TCCCGGTGAA GCTGGTCTGC CTGGTGCCAA GGGTCTGACT	1140
GGAAGCCCTG GCAGCCCTGG TCCTGATGGC AAAACTGGCC CCCCTGGTCC CGCCGGTCAA	1200
GATGGTCGCC CCGGACCCCC AGGCCCACCT GGTGCCCCGTG GTCAGGCTGG TGTGATGGGA	1260
TTCCCTGGAC CTAAAGGTGC TGCTGGAGAG CCCGGCAAGG CTGGAGAGCG AGGTGTTCCC	1320
GGACCCCTG GCGCTGTCGG TCCTGCTGGC AAAGATGGAG AGGCTGGAGC TCAGGGACCC	1380
CCTGGCCCTG CTGGTCCCGC TGGCGAGAGA GGTGAACAAG GCCCTGCTGG CTCCCCCGGA	1440
TTCCAGGGTC TCCCTGGTCC TGCTGGTCCT CCAGGTGAAG CAGGCAAACC TGGTGAACAG	1500
GGTGTTCCTG GAGACCTTGG CGCCCCTGGC CCCTCTGGAG CAAGAGGCGA GAGAGGTTTC	1560
CCTGGCGAGC GTGGTGTGCA AGGTCCCCCT GGTCTGCTG GACCCCGAGG GGCCAACGGT	1620
GCTCCCGGCA ACGATGGTGC TAAGGGTGAT GCTGGTGCCC CTGGAGCTCC CGGTAGCCAG	1680
GGCGCCCCTG GCCTTCAGGG AATGCCTGGT GAACGTGGTG CAGCTGGTCT TCCAGGGCCT	1740
AAGGGTGACA GAGGTGATGC TGGTCCCAA GGTGCTGATG GCTCTCCTGG CAAAGATGGC	1800
GTCCGTGGTC TGACCGGCCC CATTGGTCCT CCTGGCCCTG CTGGTGCCCC TGGTGACAAG	1860
GGTGAAAGTG GTCCCAGCGG CCCTGCTGGT CCCACTGGAG CTCGTGGTGC CCCCCGAGAC	1920

CGTGGTGAGC CTGGTCCCCC CGGCCCTGCT GGCTTTGCTG GCCCCCCTGG TGCTGACGGC	1980
CAACCTGGTG CTAAAGGCGA ACCTGGTGAT GCTGGTGCCA AAGGCGATGC TGGTCCCCCT	2040
GGGCCTGCCG GACCCGCTGG ACCCCCTGGC CCCATTGGTA ATGTTGGTGC TCCTGGAGCC	2100
AAAGGTGCTC GCGGCAGCGC TGGTCCCCCT GGTGCTACTG GTTTCCTGG TGCTGCTGGC	2160
CGAGTCGGTC CTCCTGGCCC CTCTGGAAT GCTGGACCCC CTGGCCCTCC TGGTCCTGCT	2220
GGCAAAGAAG GCGGCAAAGG TCCCCGTGGT GAGACTGGCC CTGCTGGACG TCCTGGTGAA	2280
GTTGGTCCCC CTGGTCCCCC TGGCCCTGCT GGCGAGAAAG GATCCCCCTGG TGCTGATGGT	2340
CCTGCTGGTG CTCCTGGTAC TCCCGGGCCT CAAGGTATTG CTGGACAGCG TGGTGTGGTC	2400
GGCCTGCCTG GTCAGAGAGG AGAGAGAGGC TTCCCTGGTC TTCCTGGCCC CTCTGGTGAA	2460
CCTGGCAAAC AAGGTCCCTC TGGAGCAAGT GGTGAACGTG GTCCCCCGG TCCCATGGGC	2520
CCCCCTGGAT TGGCTGGACC CCCTGGTGAA TCTGGACGTG AGGGGGCTCC TGCTGCCGAA	2580
GGTTCCTCTG GACGAGACGG TTCTCCTGGC GCCAAGGGTG ACCGTGGTGA GACCGGCCCC	2640
GCTGGACCCC CTGGTGCTCC TGGTGCTCCT GGTGCCCCTG GCCCCGTTGG CCCTGCTGGC	2700
AAGAGTGGTG ATCGTGGTGA GACTGGTCCT GCTGGTCCCC CCGGTCCCGT CGGCCCCGCT	2760
GGCGCCCGTG GCCCCGCCGG ACCCCAAGGC CCCCCTGGTG ACAAGGGTGA GACAGGCGAA	2820
CAGGGCGACA GAGGCATAAA GGGTCACCGT GGCTTCTCTG GCCTCCAGGG TCCCCCTGGC	2880
CCTCCTGGCT CTCCTGGTGA ACAAGGTCCC TCTGGAGCCT CTGGTCCTGC TGGTCCCCGA	2940

```

GGTCCCCCTG GCTCTGCTGG TGCTCCTGGC AAAGATGGAC TCAACGGTCT CCCTGGCCCC 3000

ATTGGGCCCC CTGGTCCTCG CGGTCGCACT GGTGATGCTG GTCCTGTTGG TCCCCCGGC 3060

CCTCCTGGAC CTCCTGGTCC CCCTGGTCCT CCCAGCGCTG GTTTCGACTT CAGCTTCCTC 3120

CCCCAGCCAC CTCAAGAGAA GGCTCACGAT GGTGGCCGCT ACTACCGGGC T 3171

```

(2) ANGABEN ZU SEQ ID NO:16:

(i) SEQUENZCHARAKTERISTIKA:

- (A) LÄNGE: 1057 Aminosäuren
- (B) ART: Aminosäure
- (C) STRANGFORM: Einzelstrang
- (D) TOPOLOGIE: unbekannt

(ii) ART DES MOLEKÜLS: Peptid

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO:16:

```

Gln Leu Ser Tyr Gly Tyr Asp Glu Lys Ser Thr Gly Gly Ile Ser Val
1           5           10           15

Pro Gly Pro Met Gly Pro Ser Gly Pro Arg Gly Leu Pro Gly Pro Pro
          20           25           30

Gly Ala Pro Gly Pro Gln Gly Phe Gln Gly Pro Pro Gly Glu Pro Gly
          35           40           45

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

```


Pro Gly Lys Asn Gly Asp Asp Gly Glu Ala Gly Lys Pro Gly Arg Pro
 65 70 75 80

Gly Glu Arg Gly Pro Pro Gly Pro Gln Gly Ala Arg Gly Leu Pro Gly
 85 90 95

Thr Ala Gly Leu Pro Gly Met Lys Gly His Arg Gly Phe Ser Gly Leu
 100 105 110

Asp Gly Ala Lys Gly Asp Ala Gly Pro Ala Gly Pro Lys Gly Glu Pro
 115 120 125

Gly Ser Pro Gly Glu Asn Gly Ala Pro Gly Gln Met Gly Pro Arg Gly
 130 135 140

Leu Pro Gly Glu Arg Gly Arg Pro Gly Ala Pro Gly Pro Ala Gly Ala
 145 150 155 160

Arg Gly Asn Asp Gly Ala Thr Gly Ala Ala Gly Pro Pro Gly Pro Thr
 165 170 175

Gly Pro Ala Gly Pro Pro Gly Phe Pro Gly Ala Val Gly Ala Lys Gly
 180 185 190

Glu Ala Gly Pro Gln Gly Pro Arg Gly Ser Glu Gly Pro Gln Gly Val
 195 200 205

Arg Gly Glu Pro Gly Pro Pro Gly Pro Ala Gly Ala Ala Gly Pro Ala
 210 215 220

Gly Asn Pro Gly Ala Asp Gly Gln Pro Gly Ala Lys Gly Ala Asn Gly
 225 230 235 240

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

Gly Val Met Gly Phe Pro Gly Pro Lys Gly Ala Ala Gly Glu Pro Gly			
420	425	430	
Lys Ala Gly Glu Arg Gly Val Pro Gly Pro Pro Gly Ala Val Gly Pro			
435	440	445	
Ala Gly Lys Asp Gly Glu Ala Gly Ala Gln Gly Pro Pro Gly Pro Ala			
450	455	460	
Gly Pro Ala Gly Glu Arg Gly Glu Gln Gly Pro Ala Gly Ser Pro Gly			
465	470	475	480
Phe Gln Gly Leu Pro Gly Pro Ala Gly Pro Pro Gly Glu Ala Gly Lys			
485	490	495	
Pro Gly Glu Gln Gly Val Pro Gly Asp Leu Gly Ala Pro Gly Pro Ser			
500	505	510	
Gly Ala Arg Gly Glu Arg Gly Phe Pro Gly Glu Arg Gly Val Gln Gly			
515	520	525	
Pro Pro Gly Pro Ala Gly Pro Arg Gly Ala Asn Gly Ala Pro Gly Asn			
530	535	540	
Asp Gly Ala Lys Gly Asp Ala Gly Ala Pro Gly Ala Pro Gly Ser Gln			
545	550	555	560
Gly Ala Pro Gly Leu Gln Gly Met Pro Gly Glu Arg Gly Ala Ala Gly			
565	570	575	
Leu Pro Gly Pro Lys Gly Asp Arg Gly Asp Ala Gly Pro Lys Gly Ala			
580	585	590	

Asp Gly Ser Pro Gly Lys Asp Gly Val Arg Gly Leu Thr Gly Pro Ile
 595 600 605

Gly Pro Pro Gly Pro Ala Gly Ala Pro Gly Asp Lys Gly Glu Ser Gly
 610 615 620

Pro Ser Gly Pro Ala Gly Pro Thr Gly Ala Arg Gly Ala Pro Gly Asp
 625 630 635 640

Arg Gly Glu Pro Gly Pro Pro Gly Pro Ala Gly Phe Ala Gly Pro Pro
 645 650 655

Gly Ala Asp Gly Gln Pro Gly Ala Lys Gly Glu Pro Gly Asp Ala Gly
 660 665 670

Ala Lys Gly Asp Ala Gly Pro Pro Gly Pro Ala Gly Pro Ala Gly Pro
 675 680 685

Pro Gly Pro Ile Gly Asn Val Gly Ala Pro Gly Ala Lys Gly Ala Arg
 690 695 700

Gly Ser Ala Gly Pro Pro Gly Ala Thr Gly Phe Pro Gly Ala Ala Gly
 705 710 715 720

Arg Val Gly Pro Pro Gly Pro Ser Gly Asn Ala Gly Pro Pro Gly Pro
 725 730 735

Pro Gly Pro Ala Gly Lys Glu Gly Gly Lys Gly Pro Arg Gly Glu Thr
 740 745 750

Gly Pro Ala Gly Arg Pro Gly Glu Val Gly Pro Pro Gly Pro Pro Gly
 755 760 765

Pro Ala Gly Glu Lys Gly Ser Pro Gly Ala Asp Gly Pro Ala Gly Ala
 770 775 780

Pro Gly Thr Pro Gly Pro Gln Gly Ile Ala Gly Gln Arg Gly Val Val
 785 790 795 800

Gly Leu Pro Gly Gln Arg Gly Glu Arg Gly Phe Pro Gly Leu Pro Gly
 805 810 815

Pro Ser Gly Glu Pro Gly Lys Gln Gly Pro Ser Gly Ala Ser Gly Glu
 820 825 830

Arg Gly Pro Pro Gly Pro Met Gly Pro Pro Gly Leu Ala Gly Pro Pro
 835 840 845

Gly Glu Ser Gly Arg Glu Gly Ala Pro Ala Ala Glu Gly Ser Pro Gly
 850 855 860

Arg Asp Gly Ser Pro Gly Ala Lys Gly Asp Arg Gly Glu Thr Gly Pro
 865 870 875 880

Ala Gly Pro Pro Gly Ala Pro Gly Ala Pro Gly Ala Pro Gly Pro Val
 885 890 895

Gly Pro Ala Gly Lys Ser Gly Asp Arg Gly Glu Thr Gly Pro Ala Gly
 900 905 910

Pro Ala Gly Pro Val Gly Pro Ala Gly Ala Arg Gly Pro Ala Gly Pro
 915 920 925

Gln Gly Pro Arg Gly Asp Lys Gly Glu Thr Gly Glu Gln Gly Asp Arg
 930 935 940

Gly Ile Lys Gly His Arg Gly Phe Ser Gly Leu Gln Gly Pro Pro Gly
 945 950 955 960

Pro Pro Gly Ser Pro Gly Glu Gln Gly Pro Ser Gly Ala Ser Gly Pro
 965 970 975

Ala Gly Pro Arg Gly Pro Pro Gly Ser Ala Gly Ala Pro Gly Lys Asp
 980 985 990

Gly Leu Asn Gly Leu Pro Gly Pro Ile Gly Pro Pro Gly Pro Arg Gly
 995 1000 1005

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

Pro Gly Pro Pro Gly Pro Pro Ser Ala Gly Phe Asp Phe Ser Phe Leu
 1025 1030 1035 1040

Pro Gln Pro Pro Gln Glu Lys Ala His Asp Gly Gly Arg Tyr Tyr Arg
 1045 1050 1055

Ala

(2) ANGABEN ZU SEQ ID NO:17:

(i) SEQUENZCHARAKTERISTIKA:

- (A) LÄNGE: 46 Aminosäuren
- (B) ART: Aminosäure
- (C) STRANGFORM: Einzelstrang
- (D) TOPOLOGIE: unbekannt

(ii) ART DES MOLEKÜLS: Peptid

(ix) MERKMAL:

- (A) NAME/SCHLÜSSEL: Bereich
- (B) LAGE: 1..2
- (D) SONSTIGE ANGABEN: /Anmerkung =
"Aminosäuresequenz für Glutathion-S-
Transferase"

(ix) MERKMAL:

- (A) NAME/SCHLÜSSEL: Bereich
- (B) LAGE: 4..5
- (D) SONSTIGE ANGABEN: /Anmerkung = "338

Wiederholungen des folgenden Triplets Gly-X-y, wobei
ungefähr 35 % des X- und Y-Positionen von Prolin und 4-
Hydroxyprolin besetzt sind."

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO:18:

Xaa	Met	Gly	Xaa	Tyr	Ser	Ala	Gly	Phe	Asp	Phe	Ser	Phe	Leu	Pro	Gln
1				5					10					15	
Pro	Pro	Gln	Glu	Lys	Ala	His	Asp	Gly	Gly	Arg	Tyr	Tyr	Arg	Ala	
			20					25					30		

(2) ANGABEN ZU SEQ ID NO:19:

(i) SEQUENZCHARAKTERISTIKA:

- (A) LÄNGE: 3171 Basenpaare
- (B) ART: Nukleinsäure
- (C) STRANGFORM: Doppelstrang
- (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: DNA (genomisch)

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO:19:

CAGCTGAGCT ATGGCTATGA TGAAAAAGC ACCGGCGGCA TCAGCGTGCC GGGCCCGATG	60
GGTCCGAGCG GCCCTCGTGG CCTGCCGGGC CCGCCAGGTG CGCCCGGTCC GCAGGGCTTT	120
CAGGGTCCGC CGGGCGAACC GGGCGAACCT GGTGCGAGCG GCCCGATGGG CCCGCGCGGC	180
CCGCCGGGTC CGCCAGGCAA AAACGGCGAT GATGGCGAAG CGGGCAAACC GGGACGTCCG	240
GGTGAACGTG GCCCCCGGG CCCGCGGGC GCGCGCGGAC TGCCGGGTAC TCGGGGACTG	300
CCGGGCATGA AAGGCCACCG CGGTTTCTCT GGTCTGGATG GTGCGAAAGG TGATGCGGGT	360
CCGGCGGGTC CGAAAGGTGA GCCGGGCAGC CCGGGCGAAA ACGGCGCGCC GGGTCAGATG	420
GGCCCGCGTG GCCTGCCTGG TGAACGCGGT CGCCCGGGCG CCCCAGGGCC AGCTGGCGCA	480
CGTGGCAACG ATGGTGCGAC CGGTGCGGCC GGTCCACCGG GCCCGACGGG CCCGGCGGGT	540
CCCCCGGGCT TTCGGGTGC GGTGGGTGCG AAAGGCGAAG CAGGTCCGCA GGGGCCGCGC	600
GGGAGCGAGG GTCCTCAGGG CGTTCGTGGT GAACCGGGCC CGCCAGGGCC GGCGGGTGCG	660
GCGGGCCCGG CTGGTAACCC TGGCGCGGAC GGTGAGCCAG GTGCGAAAGG TGCCAACGGC	720
GCGCCGGGTA TTGCAGGTGC ACCGGGCTTC CCGGGTGCCC GCGGCCCGTC CGGCCCGCAG	780
GGCCCGGGCG GCCCGCCCGG CCCGAAAGGG AACAGCGGTG AACCGGGTGC GCCAGGCAGC	840
AAAGGCGACA CCGGTGCGAA AGGTGAACCG GCGCCAGTGG GTGTTCAAGG CCCGCCGGGC	900
CCGGCGGGCG AGGAAGGCAA ACGCGGTGCT CGCGGTGAAC CGGGCCCGAC CGGCCTGCCT	960

GGCCCGCCGG GAGAACGTGG TGGCCCGGGT AGCCGCGGTT TTCCGGGCGC GGATGGTGTG	1020
GCGGGCCCGA AAGGTCCGGC GGGTGAACGT GGTAGCCCGG GCCCGGCGGG CCCAAAAGGC	1080
AGCCCGGGCG AGGCAGGACG TCCGGGTGAA GCGGGTCTCC CGGGCGCCAA AGGTCTGACC	1140
GGCTCTCCGG GCAGCCCGGG TCCGGATGGC AAAACGGGCC CGCCTGGTCC GGCCGGCCAG	1200
GATGGTCGCC CGGGCCCGCC GGGCCCGCCG GGTGCCCCGTG GTCAGGCGGG TGTCATGGGC	1260
TTTCAGGCC CCAAAGGTGC GCGGGGTGAA CCGGGCAAAG CGGGCGAACG CGGTGTCCCG	1320
GGTCCGCCGG GCGCTGTCCG GCCGGCGGGC AAAGATGGCG AAGCGGGCGC GCAAGGCCCG	1380
CCGGGACCAG CGGGTCCGGC GGGCGAGCGC GGTGAACAGG GCCCGGCAGG CAGCCCGGGT	1440
TTCCAGGGTC TGCCGGGCCC TGCGGGTCCA CCGGGTGAAG CGGGCAAACC GGGGGAACAA	1500
GGTGTGCCGG GCGACCTGGG CGCCCCAGGC CCGAGCGGCG CGCGCGGCGA ACGCGGTTTC	1560
CCGGGCGAAC GTGGTGTGCA GGGCCCCGCC GGGCCGGCTG GTCCGCGCGG CGCCAACGGC	1620
GCGCCGGGCA ACGATGGTGC GAAAGGTGAT GCGGGTGCCC CAGGTGCGCC GGGCAGCCAG	1680
GGCGCCCCCG GGCTGCAAGG CATGCCGGGT GAACGTGGTG CCGCGGGTCT ACCGGGTCCG	1740
AAAGGCGACC GCGGTGATGC GGGTCCAAAA GGTGCGGATG GCTCCCCTGG CAAAGATGGC	1800
GTTCGTGGTC TGACCGGCCC GATCGGCCCG CCGGGCCCCG CAGGTGCCCC GGGTGACAAA	1860
GGTGAAAGCG GTCCGAGCGG CCCAGCGGGC CCCACTGGTG CGCGTGGTGC CCCGGGCGAC	1920
CGTGGTGAAC CGGGTCCGCC GGGCCCGGCG GGCTTTGCGG GCCCGCCAGG CGCTGACGGC	1980

CAGCCGGGTG CGAAAGGCGA ACCGGGGGAT GCGGGTGCTA AAGGCGACGC GGGTCCGCCG	2040
GGCCCTGCCG GCCCGGCGGG CCCGCCAGGC CCGATTGGCA ACGTGGGTGC GCCGGGTGCC	2100
AAAGGTGCGC GCGGCAGCGC TGGTCCGCCG GCGCGACCG GTTTCCCCCG TCGGCGGGG	2160
CGCGTGGGTC CGCCAGGCC GAGCGGTAAC GCGGGTCCGC CAGGTCCGCC TGGCCCGGCT	2220
GGCAAAGAGG GCGGCAAAGG TCCGCGTGGT GAAACCGGCC CTGCGGGACG TCCAGGTGAA	2280
GTGGGTCCGC CGGGCCCGCC GGGCCCGGCG GGCAGAAAAG GTAGCCCGGG TCGGATGGT	2340
CCCGCCGGTG CGCCAGGCAC GCCGGGTCCG CAAGGTATCG CTGGCCAGCG TGGTGTCTC	2400
GGGCTGCCGG GTCAGCGCGG CGAACGCGGC TTTCCGGGTC TGCCGGGCCC GAGCGGTGAG	2460
CCGGGCAAAC AGGGTCCATC TGGCGCGAGC GGTGAACGTG GCCCGCCGGG TCCCATGGGC	2520
CCGCCGGGTC TGGCGGGCCC TCCGGGTGAA AGCGGTCGTG AAGGCGCGCC GGGTGCCGAA	2580
GGCAGCCCAG GCCGCGACGG TAGCCCGGGG GCCAAAGGGG ATCGTGGTGA AACCGGCCCG	2640
GCGGGCCCCC CGGGTGACCC GGGCGCGCCC GGTGCCCCAG GCCCGGTGGG CCCGCGGGC	2700
AAAAGCGGTG ATCGTGGTGA GACCGGTCCG GCGGGCCCCG CCGGTCCGGT GGGCCCAGCG	2760
GGCGCCCGTG GCCCGGCCGG TCCGCAGGGC CCGCGGGGTG ACAAAGGTGA AACGGGCGAA	2820
CAGGGCGACC GTGGCATTAA AGGCCACCGT GGCTTCAGCG GCCTGCAGGG TCCACCGGGC	2880
CCGCCGGGCA GTCCGGGTGA ACAGGGTCCG TCCGGAGCCA GCGGGCCGGC GGGCCCACGC	2940
GGTCCGCCGG GCAGCGCGGG CGCGCCGGGC AAAGACGGTC TGAACGGTCT GCCGGGCCCC	3000

ATCGGCCCCG CCGGCCCACG CGGCCGCACC GGTGATGCGG GTCCGGTGGG TCCCCGGGC 3060

CCGCCGGGCC CGCCAGGCCG GCCGGGACCG CCGAGCGCGG GTTTCGACTT CAGCTTCCTG 3120

CCGCAGCCGC CGCAGGAGAA AGCGCACGAC GCGGTCGCT ACTACCGTGC G 3171

(2) ANGABEN ZU SEQ ID NO:20:

(i) SEQUENZCHARAKTERISTIKA:

- (A) LÄNGE: 1057 Aminosäuren
- (B) ART: Aminosäure
- (C) STRANGFORM: Einzelstrang
- (D) TOPOLOGIE: unbekannt

(ii) ART DES MOLEKÜLS: Peptid

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO:20:

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

Pro Gly Lys Asn Gly Asp Asp Gly Glu Ala Gly Lys Pro Gly Arg Pro
65 70 75 80

Gly Glu Arg Gly Pro Pro Gly Pro Gln Gly Ala Arg Gly Leu Pro Gly
85 90 95

Thr Ala Gly Leu Pro Gly Met Lys Gly His Arg Gly Phe Ser Gly Leu
100 105 110

Asp Gly Ala Lys Gly Asp Ala Gly Pro Ala Gly Pro Lys Gly Glu Pro
115 120 125

Gly Ser Pro Gly Glu Asn Gly Ala Pro Gly Gln Met Gly Pro Arg Gly
130 135 140

Leu Pro Gly Glu Arg Gly Arg Pro Gly Ala Pro Gly Pro Ala Gly Ala
145 150 155 160

Arg Gly Asn Asp Gly Ala Thr Gly Ala Ala Gly Pro Pro Gly Pro Thr
165 170 175

Gly Pro Ala Gly Pro Pro Gly Phe Pro Gly Ala Val Gly Ala Lys Gly
180 185 190

Glu Ala Gly Pro Gln Gly Pro Arg Gly Ser Glu Gly Pro Gln Gly Val
195 200 205

Arg Gly Glu Pro Gly Pro Pro Gly Pro Ala Gly Ala Ala Gly Pro Ala
210 215 220

Gly Asn Pro Gly Ala Asp Gly Gln Pro Gly Ala Lys Gly Ala Asn Gly
225 230 235 240

Ala Pro Gly Ile Ala Gly Ala Pro Gly Phe Pro Gly Ala Arg Gly Pro
 245 250 255

Ser Gly Pro Gln Gly Pro Gly Gly Pro Pro Gly Pro Lys Gly Asn Ser
 260 265 270

Gly Glu Pro Gly Ala Pro Gly Ser Lys Gly Asp Thr Gly Ala Lys Gly
 275 280 285

Glu Pro Gly Pro Val Gly Val Gln Gly Pro Pro Gly Pro Ala Gly Glu
 290 295 300

Glu Gly Lys Arg Gly Ala Arg Gly Glu Pro Gly Pro Thr Gly Leu Pro
 305 310 315 320

Gly Pro Pro Gly Glu Arg Gly Gly Pro Gly Ser Arg Gly Phe Pro Gly
 325 330 335

Ala Asp Gly Val Ala Gly Pro Lys Gly Pro Ala Gly Glu Arg Gly Ser
 340 345 350

Pro Gly Pro Ala Gly Pro Lys Gly Ser Pro Gly Glu Ala Gly Arg Pro
 355 360 365

Gly Glu Ala Gly Leu Pro Gly Ala Lys Gly Leu Thr Gly Ser Pro Gly
 370 375 380

Ser Pro Gly Pro Asp Gly Lys Thr Gly Pro Pro Gly Pro Ala Gly Gln
 385 390 395 400

Asp Gly Arg Pro Gly Pro Pro Gly Pro Pro Gly Ala Arg Gly Gln Ala
 405 410 415

Gly Val Met Gly Phe Pro Gly Pro Lys Gly Ala Ala Gly Glu Pro Gly			
420	425	430	
Lys Ala Gly Glu Arg Gly Val Pro Gly Pro Pro Gly Ala Val Gly Pro			
435	440	445	
Ala Gly Lys Asp Gly Glu Ala Gly Ala Gln Gly Pro Pro Gly Pro Ala			
450	455	460	
Gly Pro Ala Gly Glu Arg Gly Glu Gln Gly Pro Ala Gly Ser Pro Gly			
465	470	475	480
Phe Gln Gly Leu Pro Gly Pro Ala Gly Pro Pro Gly Glu Ala Gly Lys			
485	490	495	
Pro Gly Glu Gln Gly Val Pro Gly Asp Leu Gly Ala Pro Gly Pro Ser			
500	505	510	
Gly Ala Arg Gly Glu Arg Gly Phe Pro Gly Glu Arg Gly Val Gln Gly			
515	520	525	
Pro Pro Gly Pro Ala Gly Pro Arg Gly Ala Asn Gly Ala Pro Gly Asn			
530	535	540	
Asp Gly Ala Lys Gly Asp Ala Gly Ala Pro Gly Ala Pro Gly Ser Gln			
545	550	555	560
Gly Ala Pro Gly Leu Gln Gly Met Pro Gly Glu Arg Gly Ala Ala Gly			
565	570	575	
Leu Pro Gly Pro Lys Gly Asp Arg Gly Asp Ala Gly Pro Lys Gly Ala			
580	585	590	

Asp Gly Ser Pro Gly Lys Asp Gly Val Arg Gly Leu Thr Gly Pro Ile			
595	600	605	
Gly Pro Pro Gly Pro Ala Gly Ala Pro Gly Asp Lys Gly Glu Ser Gly			
610	615	620	
Pro Ser Gly Pro Ala Gly Pro Thr Gly Ala Arg Gly Ala Pro Gly Asp			
625	630	635	640
Arg Gly Glu Pro Gly Pro Pro Gly Pro Ala Gly Phe Ala Gly Pro Pro			
645	650	655	
Gly Ala Asp Gly Gln Pro Gly Ala Lys Gly Glu Pro Gly Asp Ala Gly			
660	665	670	
Ala Lys Gly Asp Ala Gly Pro Pro Gly Pro Ala Gly Pro Ala Gly Pro			
675	680	685	
Pro Gly Pro Ile Gly Asn Val Gly Ala Pro Gly Ala Lys Gly Ala Arg			
690	695	700	
Gly Ser Ala Gly Pro Pro Gly Ala Thr Gly Phe Pro Gly Ala Ala Gly			
705	710	715	720
Arg Val Gly Pro Pro Gly Pro Ser Gly Asn Ala Gly Pro Pro Gly Pro			
725	730	735	
Pro Gly Pro Ala Gly Lys Glu Gly Gly Lys Gly Pro Arg Gly Glu Thr			
740	745	750	
Gly Pro Ala Gly Arg Pro Gly Glu Val Gly Pro Pro Gly Pro Pro Gly			
755	760	765	

Pro Ala Gly Glu Lys Gly Ser Pro Gly Ala Asp Gly Pro Ala Gly Ala			
770	775	780	
Pro Gly Thr Pro Gly Pro Gln Gly Ile Ala Gly Gln Arg Gly Val Val			
785	790	795	800
Gly Leu Pro Gly Gln Arg Gly Glu Arg Gly Phe Pro Gly Leu Pro Gly			
805	810	815	
Pro Ser Gly Glu Pro Gly Lys Gln Gly Pro Ser Gly Ala Ser Gly Glu			
820	825	830	
Arg Gly Pro Pro Gly Pro Met Gly Pro Pro Gly Leu Ala Gly Pro Pro			
835	840	845	
Gly Glu Ser Gly Arg Glu Gly Ala Pro Gly Ala Glu Gly Ser Pro Gly			
850	855	860	
Arg Asp Gly Ser Pro Gly Ala Lys Gly Asp Arg Gly Glu Thr Gly Pro			
865	870	875	880
Ala Gly Pro Pro Gly Ala Pro Gly Ala Pro Gly Ala Pro Gly Pro Val			
885	890	895	
Gly Pro Ala Gly Lys Ser Gly Asp Arg Gly Glu Thr Gly Pro Ala Gly			
900	905	910	
Pro Ala Gly Pro Val Gly Pro Ala Gly Ala Arg Gly Pro Ala Gly Pro			
915	920	925	
Gln Gly Pro Arg Gly Asp Lys Gly Glu Thr Gly Glu Gln Gly Asp Arg			
930	935	940	

Gly Ile Lys Gly His Arg Gly Phe Ser Gly Leu Gln Gly Pro Pro Gly
945 950 955 960

Pro Pro Gly Ser Pro Gly Glu Gln Gly Pro Ser Gly Ala Ser Gly Pro
965 970 975

Ala Gly Pro Arg Gly Pro Pro Gly Ser Ala Gly Ala Pro Gly Lys Asp
980 985 990

Gly Leu Asn Gly Leu Pro Gly Pro Ile Gly Pro Pro Gly Pro Arg Gly
995 1000 1005

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

Pro Gly Pro Pro Gly Pro Pro Ser Ala Gly Phe Asp Phe Ser Phe Leu
1025 1030 1035 1040

Pro Gln Pro Pro Gln Glu Lys Ala His Asp Gly Gly Arg Tyr Tyr Arg
1045 1050 1055

Ala

(2) ANGABEN ZU SEQ ID NO:21:

(i) SEQUENZCHARAKTERISTIKA:

- (A) LÄNGE: 79 Basenpaare
- (B) ART: Nukleinsäure
- (C) STRANGFORM: Einzelstrang
- (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: cDNA

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO:21:

GGAATTCATG CAGCTGAGCT ATGGCTATGA TGAAAAAGC ACCGGCGGCA TCAGCGTGCC 60

GGGCCCCGATG GGTCCGAGC 79

(2) ANGABEN ZU SEQ ID NO:22:

(i) SEQUENZCHARAKTERISTIKA:

- (A) LÄNGE: 75 Basenpaare
- (B) ART: Nukleinsäure
- (C) STRANGFORM: Einzelstrang
- (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: cDNA

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO:22:

GGCCCCGGGCT ACCCAGGCTC GCCGGGCGCA CCGGACGGCC CGGGCGGTCC AGCGGGGGCCA 60

GCATTATTCTG AACCC 75

(2) ANGABEN ZU SEQ ID NO:23:

(i) SEQUENZCHARAKTERISTIKA:

- (A) LÄNGE: 81 Basenpaare
- (B) ART: Nukleinsäure
- (C) STRANGFORM: Einzelstrang
- (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: cDNA

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO:23:

GGAATTCCGG GTCCGCAGGG CTTTCAGGGT CCGCCGGGCG AACCTGGTGC GAGCGGCCCCG 60

ATGGGCCCCG GCGGCCCGCC C 81

(2) ANGABEN ZU SEQ ID NO:24:

(i) SEQUENZCHARAKTERISTIKA:

- (A) LÄNGE: 87 Basenpaare
- (B) ART: Nukleinsäure
- (C) STRANGFORM: Einzelstrang
- (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: cDNA

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO:24:

TACCCGGGCG CGCCGGGCGG CCCAGGCGGT CCGTTTTTGC CGCTACTACC GTTCGCCCCGT 60

TTGGCCCTGC AGGCATTATT CGAACCC 87

(2) ANGABEN ZU SEQ ID NO:25:

(i) SEQUENZCHARAKTERISTIKA:

- (A) LÄNGE: 111 Basenpaare
- (B) ART: Nukleinsäure
- (C) STRANGFORM: Einzelstrang
- (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: cDNA

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO:25:

CAGCTGAGCT ATGGCTATGA TGAAAAAGC ACCGGCGGCA TCAGCGTGCC GGGCCCCGATG 60

GGTCCGAGCG GCCCTCGTGG CCTGCCGGGC CCGCCAGGTG CGCCCGGTCC G 111

(2) ANGABEN ZU SEQ ID NO:26:

(i) SEQUENZCHARAKTERISTIKA:

(A) LÄNGE: 37 Aminosäuren

(B) ART: Aminosäure

(C) STRANGFORM: Einzelstrang

(D) TOPOLOGIE: unbekannt

(ii) ART DES MOLEKÜLS: Peptid

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO:26:

Gln Leu Ser Tyr Gly Tyr Asp Glu Lys Ser Thr Gly Gly Ile Ser Val
1 5 10 15

Pro Gly Pro Met Gly Pro Ser Gly Pro Arg Gly Leu Pro Gly Pro Pro
20 25 30

Gly Ala Pro Gly Pro
35

(2) ANGABEN ZU SEQ ID NO:27:

(i) SEQUENZCHARAKTERISTIKA:

(A) LÄNGE: 240 Basenpaare

(B) ART: Nukleinsäure

(C) STRANGFORM: Einzelstrang

(D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: cDNA

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO:27:

```

CAGCTGAGCT ATGGCTATGA TGAAAAAGC ACCGGCGGCA TCAGCGTGCC GGGCCCGATG      60
GGTCCGAGCG GCCCTCGTGG CCTGCCGGGC CCGCCAGGTG CGCCCGGTCC GCAGGGCTTT      120
CAGGGTCCGC CGGGCGAACC GGGCGAACCT GGTGCGAGCG GCCCGATGGG CCCGCGCGGC      180
CCGCCGGGTC CGCCAGGCAA AAACGGCGAT GATGGCGAAG CGGGCAAACC GGGACGTCCG      240
    
```

(2) ANGABEN ZU SEQ ID NO:28:

(i) SEQUENZCHARAKTERISTIKA:

(A) LÄNGE: 80 Aminosäuren

(B) ART: Aminosäure

(C) STRANGFORM: Einzelstrang

(D) TOPOLOGIE: unbekannt

(ii) ART DES MOLEKÜLS: Peptid

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO:28:

```

Gln Leu Ser Tyr Gly Tyr Asp Glu Lys Ser Thr Gly Gly Ile Ser Val
1           5           10           15
Pro Gly Pro Met Gly Pro Ser Gly Pro Arg Gly Leu Pro Gly Pro Pro
          20           25           30
    
```

Gly Ala Pro Gly Pro Gln Gly Phe Gln Gly Pro Pro Gly Glu Pro Gly
 35 40 45
 Glu Pro Gly Ala Ser Gly Pro Met Gly Pro Arg Gly Pro Pro Gly Pro
 50 55 60
 Pro Gly Lys Asn Gly Asp Asp Gly Glu Ala Gly Lys Pro Gly Arg Pro
 65 70 75 80

(2) ANGABEN ZU SEQ ID NO:29:

(i) SEQUENZCHARAKTERISTIKA:

- (A) LÄNGE: 3120 Basenpaare
- (B) ART: Nukleinsäure
- (C) STRANGFORM: Einzelstrang
- (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: cDNA

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO:29:

CAGTATGATG GAAAAGGAGT TGGACTTGGC CCTGGACCAA TGGGCTTAAT GGGACCTAGA 60
 GGCCACCTG GTGCAGCTGG AGCCCCAGGC CCTCAAGGTT TCCAAGGACC TGCTGGTGAG 120
 CCTGGTGAAC CTGGTCAAAC TGGTCCTGCA GGTGCTCGTG GTCCAGCTGG CCTCCTGGC 180
 AAGGCTGGTG AAGATGGTCA CCCTGGAAAA CCCGGAACGAC CTGGTGAGAG AGGAGTTGTT 240
 GGACCACAGG GTGCTCGTGG TTTCCCTGGA ACTCCTGGAC TTCCTGGCTT CAAAGGCATT 300
 AGGGGACACA ATGGTCTGGA TGGATTGAAG GGACAGCCCG GTGCTCCTGG TGTGAAGGGT 360

GAACCTGGTG CCCCTGGTGA AAATGGAACT CCAGGTCAAA CAGGAGCCCG TGGGCTTCCT	420
GGTGAGAGAG GACGTGTTGG TGCCCCCTGGC CCAGCTGGTG CCCGTGGCAG TGATGGAAGT	480
GTGGGTCCCCG TGGGTCCCTGC TGGTCCCATT GGGTCTGCTG GCCCTCCAGG CTTCCCAGGT	540
GCCCCCTGGCC CCAAGGGTGA AATTGGAGCT GTTGGAACG CTGGTCCTGC TGGTCCCGCC	600
GGTCCCCGTG GTGAAGTGGG TCTTCCAGGC CTCTCCGGCC CCGTTGGACC TCCTGGTAAT	660
CCTGGAGCAA ACGGCCTTAC TGGTGCCAAG GGTGCTGCTG GCCTTCCCGG CGTTGCTGGG	720
GCTCCCCGCC TCCCTGGACC CCGCGGTATT CCTGGCCCTG TTGGTGCTGC CGGTGCTACT	780
GGTGCCAGAG GACTTGTTGG TGAGCCTGGT CCAGCTGGCT CCAAAGGAGA GAGCGGTAAC	840
AAGGGTGAGC CCGGCTCTGC TGGGCCCCAA GGTCTCCTG GTCCAGTGG TGAAGAAGGA	900
AAGAGAGGCC CTAATGGGGA AGCTGGATCT GCCGGCCCTC CAGGACCTCC TGGGCTGAGA	960
GGTAGTCCTG GTTCTCGTGG TCTTCCTGGA GCTGATGGCA GAGCTGGCGT CATGGGCCCT	1020
CCTGGTAGTC GTGGTGCAAG TGGCCCTGCT GGAGTCCGAG GACCTAATGG AGATGCTGGT	1080
CGCCCTGGGG AGCCTGGTCT CATGGGACCC AGAGGTCTTC CTGGTTCCCC TGGAATATC	1140
GGCCCCGCTG GAAAAGAAGG TCCTGTCGGC CTCCCTGGCA TCGACGGCAG GCCTGGCCCA	1200
ATTGGCCCAG CTGGAGCAAG AGGAGAGCCT GGCAACATTG GATTCCTGG ACCCAAAGGC	1260
CCCACTGGTG ATCCTGGCAA AAACGGTGAT AAAGGTCATG CTGGTCTTGC TGGTGCTCGG	1320
GGTGCTCCAG GTCCTGATGG AAACAATGGT GCTCAGGGAC CTCCTGGACC ACAGGGTGTT	1380

CAAGGTGGAA AAGGTGAACA GGGTCCCGCT GGTCTCCAG GCTTCCAGGG TCTGCCTGGC	1440
CCCTCAGGTC CCGCTGGTGA AGTTGGCAAA CCAGGAGAAA GGGGTCTCCA TGGTGAGTTT	1500
GGTCTCCCTG GTCCTGCTGG TCCAAGAGGG GAACGCGGTC CCCCAGGTGA GAGTGGTGCT	1560
GCCGGTCCTA CTGGTCCTAT TGAAGCCGA GGTCTTCTG GACCCCAGG GCCTGATGGA	1620
AACAAGGGTG AACCTGGTGT GGTGGTGCT GTGGGCACTG CTGGTCCATC TGGTCCTAGT	1680
GGACTCCCAG GAGAGAGGGG TGCTGCTGGC ATACCTGGAG GCAAGGGAGA AAAGGGTGAA	1740
CCTGGTCTCA GAGGTGAAAT TGGTAACCT GGCAGAGATG GTGCTCGTGG TGCTCATGGT	1800
GCTGTAGGTG CCCCTGGTCC TGCTGGAGCC ACAGGTGACC GGGGCGAAGC TGGGGCTGCT	1860
GGTCTCTGCTG GTCCTGCTGG TCCTCGGGGA AGCCCTGGTG AACGTGGCGA GGTCCGTCCT	1920
GCTGGCCCCA ACGGATTGTC TGGTCCGGCT GGTGCTGCTG GTCAACCGGG TGCTAAAGGA	1980
GAAAGAGGAG CCAAAGGGCC TAAGGGTGAA AACGGTGTTG TTGGTCCCAC AGGCCCCGTT	2040
GGAGCTGCTG GCCCAGCTGG TCCAAATGGT CCCCCCGGTC CTGCTGGAAG TCGTGGTGAT	2100
GGAGGCCCCC CTGGTATGAC TGGTTTCCCT GGTGCTGCTG GACGGACTGG TCCCCCAGGA	2160
CCCTCTGGTA TTTCTGGCCC TCCTGGTCCC CCTGGTCCTG CTGGGAAAGA AGGGCTTCGT	2220
GGTCCTCGTG GTGACCAAGG TCCAGTTGGC CGAACTGGAG AAGTAGGTGC AGTTGGTCCC	2280
CCTGGCTTCG CTGGTGAGAA GGGTCCCTCT GGAGAGGCTG GTACTGCTGG ACCTCCTGGC	2340
ACTCCAGGTC CTCAGGGTCT TCTTGGTGCT CCTGGTATTC TGGGTCTCCC TGGCTCGAGA	2400

GGTGAACGTG GTCTACCTGG TGTGCTGGT GCTGTGGGTG AACCTGGTCC TCTTGGCATT	2460
GCCGGCCCTC CTGGGGCCCG TGGTCCTCCT GGTGCTGTGG GTAGTCCTGG AGTCAACGGT	2520
GCTCCTGGTG AAGCTGGTCG TGATGGCAAC CCTGGGAACG ATGGTCCCCC AGGTCGCGAT	2580
GGTCAACCCG GACACAAGGG AGAGCGCGGT TACCCTGGCA ATATTGGTCC CGTTGGTGCT	2640
GCAGGTGCAC CTGGTCCTCA TGGCCCCGTG GGTCTGCTG GCAAACATGG AAACCGTGGT	2700
GAAACTGGTC CTTCTGGTCC TGTGGTCCT GCTGGTGCTG TTGGCCCAAG AGGTCCTAGT	2760
GGCCCACAAG GCATTCGTGG CGATAAGGGA GAGCCCGGTG AAAAGGGGCC CAGAGGTCTT	2820
CCTGGCTTAA AGGGACACAA TGGATTGCAA GGTCTGCCTG GTATCGCTGG TCACCATGGT	2880
GATCAAGGTG CTCCTGGCTC CGTGGGTCCT GCTGGTCCTA GGGGCCCTGC TGGTCCTTCT	2940
GGCCCTGCTG GAAAAGATGG TCGCACTGGA CATCCTGGTA CGGTTGGACC TGCTGGCATT	3000
CGAGGCCCTC AGGGTCACCA AGGCCCTGCT GGCCCCCTG GTCCCCCTGG CCCTCCTGGA	3060
CCTCCAGGTG TAAGCGGTGG TGGTTATGAC TTTGGTTACG ATGGAGACTT CTACAGGGCT	3120

(2) ANGABEN ZU SEQ ID NO:30:

(i) SEQUENZCHARAKTERISTIKA:

- (A) LÄNGE: 1040 Aminosäuren
- (B) ART: Aminosäure
- (C) STRANGFORM: Einzelstrang
- (D) TOPOLOGIE: unbekannt

(ii) ART DES MOLEKÜLS: Peptid

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO:30:

Gln Tyr Asp Gly Lys Gly Val Gly Leu Gly Pro Gly Pro Met Gly Leu
 1 5 10 15

Met Gly Pro Arg Gly Pro Pro Gly Ala Ala Gly Ala Pro Gly Pro Gln
 20 25 30

Gly Phe Gln Gly Pro Ala Gly Glu Pro Gly Glu Pro Gly Gln Thr Gly
 35 40 45

Pro Ala Gly Ala Arg Gly Pro Ala Gly Pro Pro Gly Lys Ala Gly Glu
 50 55 60

Asp Gly His Pro Gly Lys Pro Gly Arg Pro Gly Glu Arg Gly Val Val
 65 70 75 80

Gly Pro Gln Gly Ala Arg Gly Phe Pro Gly Thr Pro Gly Leu Pro Gly
 85 90 95

Phe Lys Gly Ile Arg Gly His Asn Gly Leu Asp Gly Leu Lys Gly Gln
 100 105 110

Pro Gly Ala Pro Gly Val Lys Gly Glu Pro Gly Ala Pro Gly Glu Asn
 115 120 125

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

Arg Val Gly Ala Pro Gly Pro Ala Gly Ala Arg Gly Ser Asp Gly Ser
 145 150 155 160

Val Gly Pro Val Gly Pro Ala Gly Pro Ile Gly Ser Ala Gly Pro Pro	165	170	175
Gly Phe Pro Gly Ala Pro Gly Pro Lys Gly Glu Ile Gly Ala Val Gly	180	185	190
Asn Ala Gly Pro Ala Gly Pro Ala Gly Pro Arg Gly Glu Val Gly Leu	195	200	205
Pro Gly Leu Ser Gly Pro Val Gly Pro Pro Gly Asn Pro Gly Ala Asn	210	215	220
Gly Leu Thr Gly Ala Lys Gly Ala Ala Gly Leu Pro Gly Val Ala Gly	225	230	235
Ala Pro Gly Leu Pro Gly Pro Arg Gly Ile Pro Gly Pro Val Gly Ala	245	250	255
Ala Gly Ala Thr Gly Ala Arg Gly Leu Val Gly Glu Pro Gly Pro Ala	260	265	270
Gly Ser Lys Gly Glu Ser Gly Asn Lys Gly Glu Pro Gly Ser Ala Gly	275	280	285
Pro Gln Gly Pro Pro Gly Pro Ser Gly Glu Glu Gly Lys Arg Gly Pro	290	295	300
Asn Gly Glu Ala Gly Ser Ala Gly Pro Pro Gly Pro Pro Gly Leu Arg	305	310	315
Gly Ser Pro Gly Ser Arg Gly Leu Pro Gly Ala Asp Gly Arg Ala Gly	325	330	335

Val Met Gly Pro Pro Gly Ser Arg Gly Ala Ser Gly Pro Ala Gly Val

340

345

350

Arg Gly Pro Asn Gly Asp Ala Gly Arg Pro Gly Glu Pro Gly Leu Met

355

360

365

Gly Pro Arg Gly Leu Pro Gly Ser Pro Gly Asn Ile Gly Pro Ala Gly

370

375

380

Lys Glu Gly Pro Val Gly Leu Pro Gly Ile Asp Gly Arg Pro Gly Pro

385

390

395

400

Ile Gly Pro Ala Gly Ala Arg Gly Glu Pro Gly Asn Ile Gly Phe Pro

405

410

415

Gly Pro Lys Gly Pro Thr Gly Asp Pro Gly Lys Asn Gly Asp Lys Gly

420

425

430

His Ala Gly Leu Ala Gly Ala Arg Gly Ala Pro Gly Pro Asp Gly Asn

435

440

445

Asn Gly Ala Gln Gly Pro Pro Gly Pro Gln Gly Val Gln Gly Gly Lys

450

455

460

Gly Glu Gln Gly Pro Ala Gly Pro Pro Gly Phe Gln Gly Leu Pro Gly

465

470

475

480

Pro Ser Gly Pro Ala Gly Glu Val Gly Lys Pro Gly Glu Arg Gly Leu

485

490

495

His Gly Glu Phe Gly Leu Pro Gly Pro Ala Gly Pro Arg Gly Glu Arg

500

505

510

Gly Pro Pro Gly Glu Ser Gly Ala Ala Gly Pro Thr Gly Pro Ile Gly			
515	520	525	
Ser Arg Gly Pro Ser Gly Pro Pro Gly Pro Asp Gly Asn Lys Gly Glu			
530	535	540	
Pro Gly Val Val Gly Ala Val Gly Thr Ala Gly Pro Ser Gly Pro Ser			
545	550	555	560
Gly Leu Pro Gly Glu Arg Gly Ala Ala Gly Ile Pro Gly Gly Lys Gly			
565	570	575	
Glu Lys Gly Glu Pro Gly Leu Arg Gly Glu Ile Gly Asn Pro Gly Arg			
580	585	590	
Asp Gly Ala Arg Gly Ala His Gly Ala Val Gly Ala Pro Gly Pro Ala			
595	600	605	
Gly Ala Thr Gly Asp Arg Gly Glu Ala Gly Ala Ala Gly Pro Ala Gly			
610	615	620	
Pro Ala Gly Pro Arg Gly Ser Pro Gly Glu Arg Gly Glu Val Gly Pro			
625	630	635	640
Ala Gly Pro Asn Gly Phe Ala Gly Pro Ala Gly Ala Ala Gly Gln Pro			
645	650	655	
Gly Ala Lys Gly Glu Arg Gly Ala Lys Gly Pro Lys Gly Glu Asn Gly			
660	665	670	
Val Val Gly Pro Thr Gly Pro Val Gly Ala Ala Gly Pro Ala Gly Pro			
675	680	685	

Asn Gly Pro Pro Gly Pro Ala Gly Ser Arg Gly Asp Gly Gly Pro Pro
690 695 700

Gly Met Thr Gly Phe Pro Gly Ala Ala Gly Arg Thr Gly Pro Pro Gly
705 710 715 720

Pro Ser Gly Ile Ser Gly Pro Pro Gly Pro Pro Gly Pro Ala Gly Lys
725 730 735

Glu Gly Leu Arg Gly Pro Arg Gly Asp Gln Gly Pro Val Gly Arg Thr
740 745 750

Gly Glu Val Gly Ala Val Gly Pro Pro Gly Phe Ala Gly Glu Lys Gly
755 760 765

Pro Ser Gly Glu Ala Gly Thr Ala Gly Pro Pro Gly Thr Pro Gly Pro
770 775 780

Gln Gly Leu Leu Gly Ala Pro Gly Ile Leu Gly Leu Pro Gly Ser Arg
785 790 795 800

Gly Glu Arg Gly Leu Pro Gly Val Ala Gly Ala Val Gly Glu Pro Gly
805 810 815

Pro Leu Gly Ile Ala Gly Pro Pro Gly Ala Arg Gly Pro Pro Gly Ala
820 825 830

Val Gly Ser Pro Gly Val Asn Gly Ala Pro Gly Glu Ala Gly Arg Asp
835 840 845

Gly Asn Pro Gly Asn Asp Gly Pro Pro Gly Arg Asp Gly Gln Pro Gly
850 855 860

His Lys Gly Glu Arg Gly Tyr Pro Gly Asn Ile Gly Pro Val Gly Ala
 865 870 875 880

Ala Gly Ala Pro Gly Pro His Gly Pro Val Gly Pro Ala Gly Lys His
 885 890 895

Gly Asn Arg Gly Glu Thr Gly Pro Ser Gly Pro Val Gly Pro Ala Gly
 900 905 910

Ala Val Gly Pro Arg Gly Pro Ser Gly Pro Gln Gly Ile Arg Gly Asp
 915 920 925

Lys Gly Glu Pro Gly Glu Lys Gly Pro Arg Gly Leu Pro Gly Leu Lys
 930 935 940

Gly His Asn Gly Leu Gln Gly Leu Pro Gly Ile Ala Gly His His Gly
 945 950 955 960

Asp Gln Gly Ala Pro Gly Ser Val Gly Pro Ala Gly Pro Arg Gly Pro
 965 970 975

Ala Gly Pro Ser Gly Pro Ala Gly Lys Asp Gly Arg Thr Gly His Pro
 980 985 990

Gly Thr Val Gly Pro Ala Gly Ile Arg Gly Pro Gln Gly His Gln Gly
 995 1000 1005

Pro Ala Gly Pro Pro Gly Pro Pro Gly Pro Pro Gly Pro Pro Gly Val
 1010 1015 1020

Ser Gly Gly Gly Tyr Asp Phe Gly Tyr Asp Gly Asp Phe Tyr Arg Ala
 1025 1030 1035 1040

(2) ANGABEN ZU SEQ ID NO:31:

(i) SEQUENZCHARAKTERISTIKA:

- (A) LÄNGE: 3120 Basenpaare
- (B) ART: Nukleinsäure
- (C) STRANGFORM: Einzelstrang
- (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: cDNA

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO:31:

```

CAGTACGACG GTAAAGGCGT AGGCCTGGGT CCGGGTCCGA TGGGCCTGAT GGGTCCACGT      60
GGCCCCACCGG GTGCAGCAGG TGCGCCGGGT CCGCAGGGCT TCCAAGGTCC GGCGGGTGAA      120
CCGGGCGAAC CGGGTCAGAC GGGTCCGGCG GGTGCTCGCG GTCCGGCTGG CCCACCGGGC      180
AAAGCTGGCG AAGACGGTCA CCCGGGTAAG CCAGGCCGCC CGGGCGAACG TGGCGTCGTG      240
GGTCCGCAAG GTGCGCGTGG TTTCCCGGGC ACGCCGGGTC TGCCGGGTTT CAAAGGCATT      300
CGTGGTCACA ACGGTCTGGA CGGTCTGAAA GGCCAACCGG GTGCTCCGGG CGTCAAAGGC      360
GAACCGGGTG CCCCAGGCGA AAACGGTACG CCGGGCCAGA CTGGTGCGCG TGGTCTGCCG      420
GGTGAACGCG GCCGTGTTGG CGCTCCGGGT CCGGCTGGCG CGCGTGSCAG CGATGGCTCC      480
GTCGGTCCGG TTGGCCCTGC GGGTCCGATT GGTTCGCTG GCCCTCCGGG TTTCCCGGGT      540
GCGCCGGGTC CGAAGGGTGA GATCGGCGCG GTTGGCAACG CAGGCCCCGC TGGTCCAGCC      600
GGCCCTCGTG GCGAAGTCGG TCTGCCGGGT CTGAGCGGTC CGGTAGGCCC ACCGGGTAAC      660

```

CCGGGCGCAA ACGGCCTGAC GGGTGCAAAA GGTGCGGCTG GCCTGCCGGG CGTTGCCGGT	720
GCCCCGGGCC TGCCGGGTCC GCGCGGTATT CCGGGTCCGG TAGGCGCAGC CCGTGCAACT	780
GGTGCCCGTG GCCTGGTTGG CGAACCAGGT CCGGCGGGTT CTAAAGGCGA AAGCGGTAAC	840
AAAGGTGAGC CGGGTTCCGC GGGCCCCGAG GGTCCGCCCG GTCCGAGCGG CGAAGAAGGT	900
AAACGTGGTC CGAACGGCGA GGCTGGTTCC GCAGGCCCTC CCGGTCCGCC GGGTCTGCGT	960
GGCAGCCCGG GTAGCCGTGG CCTGCCGGGC GCGGACGGCC GTGCGGGCGT GATGGGTCCG	1020
CCGGGTTCCT GTGGTGCTTC TGGTCCGGCT GGTGTCCGTG GTCCGAATGG CGACGCGGGC	1080
CGTCCGGGTG AACCGGGCCT GATGGGTCCG CGTGGCCTGC CGGGTAGCCC GGGTAACATT	1140
GGTCCGGCGG GTAAGGAGGG TCCGGTAGGT CTGCCGGGTA TTGATGGTCG TCCGGGTCCG	1200
ATCGGCCCTG CGGGCGCTCG TGGCGAGCCG GGTAACATCG GTTTTCCGGG TCCGAAGGGT	1260
CCGACGGGCG ACCCGGGCAA GAACGGTGAT AAAGGCCATG CAGGTCTGGC AGGTGCCCGT	1320
GGTGCAACCG GTCCGGATGG TAACAATGGT GCGCAGGGTC CGCCGGGTCC GCAGGGCGTA	1380
CAGGGTGGCA AAGGTGAACA GGGTCCGGCA GGGCCACCGG GCTTCCAGGG TCTGCCGGGT	1440
CCGAGCGGCC CGGCTGGTGA AGTGGGCAAA CCGGGCGAAC GTGGCCTCCA TGGCGAGTTT	1500
GGCCTGCCGG GTCCGGCCGG TCCGCGTGGT GAGCGCGGCC CTCCGGGCGA ATCCGGCGCG	1560
GCAGGTCCGA CCGGCCCCGAT TGGTTCCCGT GGTCCGAGCG GCCCACCAGG TCCGGACGGC	1620
AACAAAGGCG AGCCGGGTGT TGTGGTGCT GTTGGTACCG CCGGCCCGTC TGGTCCGAGC	1680

GGTCTGCCCG GCGAACGCGG TGCCGCTGGT ATTCCGGGCG GCAAAGGTGA AAAAGGTGAA	1740
CCGGGTCTGC GCGGTGAGAT TGGCAACCCG GGCCGTGACG GTGCTCGCGG TGCACACGGC	1800
GCGGTTGGCG CACCGGGTCC GGCAGGCGCG ACTGGTGATC GTGGCGAAGC TGGTGCAGCG	1860
GGTCCGGCGG GTCCGGCCCG CCCTCGCGGT TCCCCGGGCG AACGCGGCGA AGTCGGCCCCG	1920
GCTGGCCCCA ATGGCTTTGC TGGCCCAGCG GGCGCTGCGG GCCAACCGGG TGCGAAAGGT	1980
GAGCGCGGTG CCAAAGGCCC GAAAGGTGAA AATGGTGTAG TTGGTCCGAC GGGTCCGGTT	2040
GGTGCGGCTG GTCCGGCTGG CCCGAATGGT CCGCCGGGTC CGGCAGGCAG CCGTGCGCAT	2100
GGTGGCCCCAC CGGGCATGAC CGGTTTCCCT GGCGCGGCCG GTCGCACCGG CCCGCCGGGT	2160
CCGTCTGGCA TTTCTGGCCC ACCGGGTCCG CCGGGTCCGG CGGGCAAAGA AGGTCTGCGT	2220
GGCCACGCG GCGACCAGGG TCCGGTGGGC CGTACCGGCG AAGTCGGTGC TGTGGCCCT	2280
CCGGGCTTTG CGGGTGAGAA AGGTCCGAGC GGTGAAGCTG GCACCGCAGG CCCGCCGGGT	2340
ACGCCGGGTC CGCAAGGTCT GCTGGGTGCT CCGGGTATCC TGGGCCTGCC GGGCTCCCGT	2400
GGCGAACGCG GTCTGCCGGG CGTTGCAGGC GCTGTAGGCG AACCGGGTCC GCTGGGTATC	2460
GCGGGTCCGC CGGGTGCGCG TGGTCCGCCG GGTGCCGTGG GCTCTCCGGG TGTTAACGGC	2520
GCCCCTGGTG AAGCGGGCCG CGACGGCAAT CCGGGCAACG ATGGTCCGCC GGGTCGTGAT	2580
GGTCAGCCGG GTCACAAAGG TGAGCGTGGC TACCCGGGTA ACATCGGTCC GGTGGTGCG	2640
GCCGCGCCTC CGGGTCCGCA CGGTCCGGTA GGCCCAGCCG GCAAACACGG TAACCGTGGT	2700

```

GAAACGGGTC CGTCCGGTCC GGTAGGTCCG GCGGGTGCTG TTGGTCCACG CGGCCCGTCC 2760
GGCCCGCAGG GTATTGCGCG TGACAAAGGC GAACCGGGCG AAAAAGGTCC GCGTGGTCTG 2820
CCGGGCCTTA AGGGCCACAA CCGTCTGCAA GGTCTGCCGG GTATCGCGGG TCACCACGGT 2880
GATCAGGGTG CTCCGGGTTC CGTTGGTCCG GCCGGTCCGC GTGGCCCCGC TGGTCCGTCT 2940
GGTCCGGCCG GTAAAGACGG CCGTACGGGC CACCCGGGTA CCGTGGGTCC GGCCGGCATT 3000
CGCGGTCCGC AAGGTCACCA GGGTCCGGCG GGTCCGCCGG GTCCGCCGGG TCCGCCGGGT 3060
CCGCCGGGTG TTAGCGGTGG CGGTTATGAT TTTGGTTATG ACGGTGATTT CTATCGTGCG 3120

```

(2) ANGABEN ZU SEQ ID NO:32:

(i) SEQUENZCHARAKTERISTIKA:

- (A) LÄNGE: 1040 Aminosäuren
- (B) ART: Aminosäure
- (C) STRANGFORM: Einzelstrang
- (D) TOPOLOGIE: unbekannt

(ii) ART DES MOLEKÜLS: Peptid

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO:32:

```

Gln Tyr Asp Gly Lys Gly Val Gly Leu Gly Pro Gly Pro Met Gly Leu
1           5           10          15

```

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

Asn Ala Gly Pro Ala Gly Pro Ala Gly Pro Arg Gly Glu Val Gly Leu			
195	200	205	
Pro Gly Leu Ser Gly Pro Val Gly Pro Pro Gly Asn Pro Gly Ala Asn			
210	215	220	
Gly Leu Thr Gly Ala Lys Gly Ala Ala Gly Leu Pro Gly Val Ala Gly			
225	230	235	240
Ala Pro Gly Leu Pro Gly Pro Arg Gly Ile Pro Gly Pro Val Gly Ala			
245	250	255	
Ala Gly Ala Thr Gly Ala Arg Gly Leu Val Gly Glu Pro Gly Pro Ala			
260	265	270	
Gly Ser Lys Gly Glu Ser Gly Asn Lys Gly Glu Pro Gly Ser Ala Gly			
275	280	285	
Pro Gln Gly Pro Pro Gly Pro Ser Gly Glu Glu Gly Lys Arg Gly Pro			
290	295	300	
Asn Gly Glu Ala Gly Ser Ala Gly Pro Pro Gly Pro Pro Gly Leu Arg			
305	310	315	320
Gly Ser Pro Gly Ser Arg Gly Leu Pro Gly Ala Asp Gly Arg Ala Gly			
325	330	335	
Val Met Gly Pro Pro Gly Ser Arg Gly Ala Ser Gly Pro Ala Gly Val			
340	345	350	
Arg Gly Pro Asn Gly Asp Ala Gly Arg Pro Gly Glu Pro Gly Leu Met			
355	360	365	

Gly Pro Arg Gly Leu Pro Gly Ser Pro Gly Asn Ile Gly Pro Ala Gly
 370 375 380

Lys Glu Gly Pro Val Gly Leu Pro Gly Ile Asp Gly Arg Pro Gly Pro
 385 390 395 400

Ile Gly Pro Ala Gly Ala Arg Gly Glu Pro Gly Asn Ile Gly Phe Pro
 405 410 415

Gly Pro Lys Gly Pro Thr Gly Asp Pro Gly Lys Asn Gly Asp Lys Gly
 420 425 430

His Ala Gly Leu Ala Gly Ala Arg Gly Ala Pro Gly Pro Asp Gly Asn
 435 440 445

Asn Gly Ala Gln Gly Pro Pro Gly Pro Gln Gly Val Gln Gly Gly Lys
 450 455 460

Gly Glu Gln Gly Pro Ala Gly Pro Pro Gly Phe Gln Gly Leu Pro Gly
 465 470 475 480

Pro Ser Gly Pro Ala Gly Glu Val Gly Lys Pro Gly Glu Arg Gly Leu
 485 490 495

His Gly Glu Phe Gly Leu Pro Gly Pro Ala Gly Pro Arg Gly Glu Arg
 500 505 510

Gly Pro Pro Gly Glu Ser Gly Ala Ala Gly Pro Thr Gly Pro Ile Gly
 515 520 525

Ser Arg Gly Pro Ser Gly Pro Pro Gly Pro Asp Gly Asn Lys Gly Glu
 530 535 540

Pro Gly Val Val Gly Ala Val Gly Thr Ala Gly Pro Ser Gly Pro Ser
 545 550 555 560

Gly Leu Pro Gly Glu Arg Gly Ala Ala Gly Ile Pro Gly Gly Lys Gly
 565 570 575

Glu Lys Gly Glu Pro Gly Leu Arg Gly Glu Ile Gly Asn Pro Gly Arg
 580 585 590

Asp Gly Ala Arg Gly Ala His Gly Ala Val Gly Ala Pro Gly Pro Ala
 595 600 605

Gly Ala Thr Gly Asp Arg Gly Glu Ala Gly Ala Ala Gly Pro Ala Gly
 610 615 620

Pro Ala Gly Pro Arg Gly Ser Pro Gly Glu Arg Gly Glu Val Gly Pro
 625 630 635 640

Ala Gly Pro Asn Gly Phe Ala Gly Pro Ala Gly Ala Ala Gly Gln Pro
 645 650 655

Gly Ala Lys Gly Glu Arg Gly Ala Lys Gly Pro Lys Gly Glu Asn Gly
 660 665 670

Val Val Gly Pro Thr Gly Pro Val Gly Ala Ala Gly Pro Ala Gly Pro
 675 680 685

Asn Gly Pro Pro Gly Pro Ala Gly Ser Arg Gly Asp Gly Gly Pro Pro
 690 695 700

Gly Met Thr Gly Phe Pro Gly Ala Ala Gly Arg Thr Gly Pro Pro Gly
 705 710 715 720

Pro Ser Gly Ile Ser Gly Pro Pro Gly Pro Pro Gly Pro Ala Gly Lys

725

730

735

Glu Gly Leu Arg Gly Pro Arg Gly Asp Gln Gly Pro Val Gly Arg Thr

740

745

750

Gly Glu Val Gly Ala Val Gly Pro Pro Gly Phe Ala Gly Glu Lys Gly

755

760

765

Pro Ser Gly Glu Ala Gly Thr Ala Gly Pro Pro Gly Thr Pro Gly Pro

770

775

780

Gln Gly Leu Leu Gly Ala Pro Gly Ile Leu Gly Leu Pro Gly Ser Arg

785

790

795

800

Gly Glu Arg Gly Leu Pro Gly Val Ala Gly Ala Val Gly Glu Pro Gly

805

810

815

Pro Leu Gly Ile Ala Gly Pro Pro Gly Ala Arg Gly Pro Pro Gly Ala

820

825

830

Val Gly Ser Pro Gly Val Asn Gly Ala Pro Gly Glu Ala Gly Arg Asp

835

840

845

Gly Asn Pro Gly Asn Asp Gly Pro Pro Gly Arg Asp Gly Gln Pro Gly

850

855

860

His Lys Gly Glu Arg Gly Tyr Pro Gly Asn Ile Gly Pro Val Gly Ala

865

870

875

880

Ala Gly Ala Pro Gly Pro His Gly Pro Val Gly Pro Ala Gly Lys His

885

890

895

Gly Asn Arg Gly Glu Thr Gly Pro Ser Gly Pro Val Gly Pro Ala Gly			
900	905	910	
Ala Val Gly Pro Arg Gly Pro Ser Gly Pro Gln Gly Ile Arg Gly Asp			
915	920	925	
Lys Gly Glu Pro Gly Glu Lys Gly Pro Arg Gly Leu Pro Gly Leu Lys			
930	935	940	
Gly His Asn Gly Leu Gln Gly Leu Pro Gly Ile Ala Gly His His Gly			
945	950	955	960
Asp Gln Gly Ala Pro Gly Ser Val Gly Pro Ala Gly Pro Arg Gly Pro			
965	970	975	
Ala Gly Pro Ser Gly Pro Ala Gly Lys Asp Gly Arg Thr Gly His Pro			
980	985	990	
Gly Thr Val Gly Pro Ala Gly Ile Arg Gly Pro Gln Gly His Gln Gly			
995	1000	1005	
Pro Ala Gly Pro Pro Gly Pro Pro Gly Pro Pro Gly Pro Pro Gly Val			
1010	1015	1020	
Ser Gly Gly Gly Tyr Asp Phe Gly Tyr Asp Gly Asp Phe Tyr Arg Ala			
1025	1030	1035	1040

(2) ANGABEN ZU SEQ ID NO:33:

- (i) SEQUENZCHARAKTERISTIKA:
 - (A) LÄNGE: 76 Basenpaare
 - (B) ART: Nukleinsäure
 - (C) STRANGFORM: Einzelstrang

(D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: cDNA

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO:33:

```
GGAATTCATG CAGTATGATG GCAAAGGCGT CGGCCTCGGC CCGGGCCCAA TGGGCCTCAT 60
GGGCCCCGCGC GGCCCA 76
```

(2) ANGABEN ZU SEQ ID NO:34:

(i) SEQUENZCHARAKTERISTIKA:

- (A) LÄNGE: 79 Basenpaare
- (B) ART: Nukleinsäure
- (C) STRANGFORM: Einzelstrang
- (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: cDNA

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO:34:

```
CCGGGCGCGC CGGGTGGCCC ACGTCGACCG CGGGGTCCGG GCGTTCCAAA GGTCCCGGGA 60
CGGCCAATTA TTCGAACCC 79
```

(2) ANGABEN ZU SEQ ID NO:35:

(i) SEQUENZCHARAKTERISTIKA:

- (A) LÄNGE: 82 Basenpaare
- (B) ART: Nukleinsäure

(C) STRANGFORM: Einzelstrang

(D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: cDNA

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO:35:

GGAATTCGCC GGTGAGCCGG GTGAACCGGG CCAAACGGGT CCGGCAGGTC CACGTGGTCC 60

AGCGGGCCCCG CCTGGCAAGG CG 82

(2) ANGABEN ZU SEQ ID NO:36:

(i) SEQUENZCHARAKTERISTIKA:

(A) LÄNGE: 84 Basenpaare

(B) ART: Nukleinsäure

(C) STRANGFORM: Einzelstrang

(D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: cDNA

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO:36:

CCGGGCGGAC CGTTCCGCC ACTTCTACCG GTGGGACCGT TTGGCCCGGC GGGCCACTCG 60

CACCGCATCA CATTATTCGA ACCC 84

(2) ANGABEN ZU SEQ ID NO:37:

(i) SEQUENZCHARAKTERISTIKA:

(A) LÄNGE: 240 Basenpaare

- (B) ART: Nukleinsäure
- (C) STRANGFORM: Einzelstrang
- (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: cDNA

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO:37:

```

CAGTATGATG GCAAAGGCGT CGGCCTCGGC CCGGGCCCAA TGGGCCTCAT GGGCCGCGC      60
GGCCCACCGG GTGCAGCTGG CGCCCCAGGC CCGCAAGGTT TCCAGGGCCC TGCCGGTGAG      120
CCGGGTGAAC CCGGCCAAAC GGGTCCGGCA GGTGCACGTG GTCCAGCGGG CCCGCCTGGC      180
AAGGCGGGTG AAGATGGCCA CCCTGGCAA CCGGGCCGCC CGGGTGAGCG TGGCGTAGTG      240
  
```

(2) ANGABEN ZU SEQ ID NO:38:

- (i) SEQUENZCHARAKTERISTIKA:
 - (A) LÄNGE: 80 Aminosäuren
 - (B) ART: Aminosäure
 - (C) STRANGFORM: Einzelstrang
 - (D) TOPOLOGIE: unbekannt

(ii) ART DES MOLEKÜLS: Peptid

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO:38:

```

Gln Tyr Asp Gly Lys Gly Val Gly Leu Gly Pro Gly Pro Met Gly Leu
1           5           10           15
  
```

Met Gly Pro Arg Gly Pro Pro Gly Ala Ala Gly Ala Pro Gly Pro Gln

20

25

30

Gly Phe Gln Gly Pro Ala Gly Glu Pro Gly Glu Pro Gly Gln Thr Gly

35

40

45

Pro Ala Gly Ala Arg Gly Pro Ala Gly Pro Pro Gly Lys Ala Gly Glu

50

55

60

Asp Gly His Pro Gly Lys Pro Gly Arg Pro Gly Glu Arg Gly Val Val

65

70

75

80

(2) ANGABEN ZU SEQ ID NO:39:

(i) SEQUENZCHARAKTERISTIKA:

(A) LÄNGE: 276 Basenpaare

(B) ART: Nukleinsäure

(C) STRANGFORM: Einzelstrang

(D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: cDNA

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO:39:

ATGGGGCTCG CTGGCCCACC GGGCGAACC GGTCCGCCAG GCCCGAAAGG TCCGCGTGGC	60
GATAGCGGGC TCGCTGGCCC ACCGGGCGAA CCGGGTCCGC CAGGCCCGAA AGGTCCGCGT	120
GGCGATAGCG GGCTCGCTGG CCCACCGGGC GAACCGGGTC CGCCAGGCCC GAAAGGTCCG	180
CGTGGCGATA GCGGGCTCGC TGGCCCACCG GGCGAACC GG TCCGCCAGG CCCGAAAGGT	240
CCGCGTGGCG ATAGCGGGCT CCCGGGCGAT TCCTAA	276

(2) ANGABEN ZU SEQ ID NO:40:

(i) SEQUENZCHARAKTERISTIKA:

(A) LÄNGE: 91 Aminosäuren

(B) ART: Aminosäure

(C) STRANGFORM: Einzelstrang

(D) TOPOLOGIE: unbekannt

(ii) ART DES MOLEKÜLS: Peptid

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO:40:

Met Gly Leu Ala Gly Pro Pro Gly Glu Pro Gly Pro Pro Gly Pro Lys

Met Gly Leu Ala Gly Pro Pro Gly Glu Pro Gly Pro Pro Gly Pro Lys

1 5 10 15 y

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

20 25 30 o

Pro Pro Gly Pro Lys Gly Pro Arg Gly Asp Ser Gly Leu Ala Gly Pro

35 40 45 r

Pro Gly Glu Pro Gly Pro Pro Gly Pro Lys Gly Pro Arg Gly Asp Ser

50 55 60 y

Gly Leu Ala Gly Pro Pro Gly Glu Pro Gly Pro Pro Gly Pro Lys Gly

65 70 75 80

Pro Arg Gly Asp Ser Gly Leu Pro Gly Asp Ser

85 90

(2) ANGABEN ZU SEQ ID NO:41:

(i) SEQUENZCHARAKTERISTIKA:

- (A) LÄNGE: 13 Aminosäuren
(B) ART: Aminosäure
(C) STRANGFORM: Einzelstrang
(D) TOPOLOGIE: unbekannt

(ii) ART DES MOLEKÜLS: Peptid

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO:41:

Gly Pro Pro Gly Leu Ala Gly Pro Pro Gly Glu Ser Gly
1 5 10

(2) ANGABEN ZU SEQ ID NO:42:

(i) SEQUENZCHARAKTERISTIKA:

- (A) LÄNGE: 13 Aminosäuren
(B) ART: Aminosäure
(C) STRANGFORM: Einzelstrang
(D) TOPOLOGIE: unbekannt

(ii) ART DES MOLEKÜLS: Peptid

(ix) MERKMAL:

- (A) NAME/SCHLÜSSEL: Modifizierte-Stelle
(B) LAGE: 2..3
(D) SONSTIGE ANGABEN: /Produkt = "4-Hydroxyprolin"

(ix) MERKMAL:

- (A) NAME/SCHLÜSSEL: Modifiziert-Stelle
- (B) LAGE: 8..9
- (D) SONSTIGE ANGABEN: /Produkt = "Xaa =
4-Hydroxyprolin"

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO:42:

Gly Xaa Xaa Gly Leu Ala Gly Xaa Xaa Gly Glu Ser Gly
 1 5 10

(2) ANGABEN ZU SEQ ID NO:43:

(i) SEQUENZCHARAKTERISTIKA:

- (A) LÄNGE: 660 Basenpaare
- (B) ART: Nukleinsäure
- (C) STRANGFORM: Einzelstrang
- (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: cDNA

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO:43:

```

ATGGGCCCCG CGGGTCTGGC GGGCCCTCCG GGTGAAAGCG GTCGTGAAGG CGCGCCGGGT      60
GCCGAAGGCA GCCCAGGCCG CGACGGTAGC CCGGGGGCCA AAGGGGATCG TGGTGAAACC      120
GGCCCGGCGG GCCCCCGGG TGCACCGGGC GCGCCGGGTG CCCCAGGCCG GGTGGGCCCC      180
GCGGGCAAAA GCGGTGATCG TGGTGAGACC GGTCCGGCGG GCCCGGCCGG TCCGGTGGGC      240
CCAGCGGGCG CCCGTGGCCC GGCCGGTCCG CAGGGCCCGC GGGGTGACAA AGGTGAAACG      300
GGCGAACAGG GCGACCGTGG CATTAAAGGC CACCGTGGCT TCAGCGGCCT GCAGGGTCCA      360

```

```

CCGGGCCCCG CCGGCAGTCC GGGTGAACAG GGTCCGTCCG GAGCCAGCGG GCCGGCGGGC      420
CCACGCGGTC CGCCGGGCAG CGCGGGCGCG CCGGGCAAAG ACGGTCTGAA CGGTCTGCCG      480
GGCCCCGATCG GCCCGCCGGG CCCACGCGGC CGCACCGGTG ATGCGGGTCC GGTGGGTCCC      540
CCGGGCCCCG CCGGCCCGCC AGGCCCCCGG GGACCGCCGA GCGCGGGTTT CGACTTCAGC      600
TTCCTGCCGC AGCCGCCGCA GGAGAAAGCG CACGACGGCG GTCGCTACTA CCGTGCGTAA      660

```

(2) ANGABEN ZU SEQ ID NO:44:

(i) SEQUENZCHARAKTERISTIKA:

- (A) LÄNGE: 219 Aminosäuren
- (B) ART: Aminosäure
- (C) STRANGFORM: Einzelstrang
- (D) TOPOLOGIE: unbekannt

(ii) ART DES MOLEKÜLS: Peptid

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO:44:

```

Met Gly Pro Pro Gly Leu Ala Gly Pro Pro Gly Glu Ser Gly Arg Glu
1           5           10           15

Gly Ala Pro Gly Ala Glu Gly Ser Pro Gly Arg Asp Gly Ser Pro Gly
          20           25           30

Ala Lys Gly Asp Arg Gly Glu Thr Gly Pro Ala Gly Pro Pro Gly Ala
          35           40           45

Pro Gly Ala Pro Gly Ala Pro Gly Pro Val Gly Pro Ala Gly Lys Ser
          50           55           60-

```

Gly Asp Arg Gly Glu Thr Gly Pro Ala Gly Pro Ala Gly Pro Val Gly
 65 70 75 80

Pro Ala Gly Ala Arg Gly Pro Ala Gly Pro Gln Gly Pro Arg Gly Asp
 85 90 95

Lys Gly Glu Thr Gly Glu Gln Gly Asp Arg Gly Ile Lys Gly His Arg
 100 105 110

Gly Phe Ser Gly Leu Gln Gly Pro Pro Gly Pro Pro Gly Ser Pro Gly
 115 120 125

Glu Gln Gly Pro Ser Gly Ala Ser Gly Pro Ala Gly Pro Arg Gly Pro
 130 135 140

Pro Gly Ser Ala Gly Ala Pro Gly Lys Asp Gly Leu Asn Gly Leu Pro
 145 150 155 160

Gly Pro Ile Gly Pro Pro Gly Pro Arg Gly Arg Thr Gly Asp Ala Gly
 165 170 175

Pro Val Gly Pro Pro Gly Pro Pro Gly Pro Pro Gly Pro Pro Gly Pro
 180 185 190

Pro Ser Ala Gly Phe Asp Phe Ser Phe Leu Pro Gln Pro Pro Gln Glu
 195 200 205

Lys Ala His Asp Gly Gly Arg Tyr Tyr Arg Ala
 210 215

(2) ANGABEN ZU SEQ ID NO:45:

(i) SEQUENZCHARAKTERISTIKA:

(A) LÄNGE: 627 Basenpaare

- (B) ART: Nukleinsäure
- (C) STRANGFORM: Einzelstrang
- (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: cDNA

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO:45:

ATGGGCTCTC CGGGTGTAA CGGCGCCCT GGTGAAGCG GCCGCGACGG CAATCCGGGC	60
AACGATGGTC CGCCGGGTCG TGATGGTCAG CCGGGTCACA AAGGTGAGCG TGGCTACCCG	120
GGTAACATCG GTCCGGTTGG TCGGGCCGGC GCTCCGGGTC CGCACGGTCC GGTAGGCCCA	180
GCCGGCAAAC ACGGTAACCG TGGTGAAACG GGTCCGTCCG GTCCGGTAGG TCCGGCGGGT	240
GCTGTTGGTC CACGCGGCCG GTCCGGCCCG CAGGGTATTC GCGGTGACAA AGGCGAACCG	300
GGCGAAAAAG GTCCGCGTGG TCTGCCGGGC CTTAAGGGCC ACAACGGTCT GCAAGGTCTG	360
CCGGGTATCG CCGGTCACCA CGGTGATCAG GGTGCTCCGG GTTCCGTTGG TCCGGCCGGT	420
CCGCGTGGCC CGGCTGGTCC GTCTGGTCCG GCCGGTAAAG ACGGCCGTAC GGGCCACCCG	480
GGTACGGTGG GTCCGGCCCG CATTCGCGGT CCGCAAGGTC ACCAGGGTCC GGCGGGTCCG	540
CCGGGTCCGC CCGGTCCGCC GGGTCCGCCG GGTGTTAGCG GTGGCGGTTA TGATTTTGGT	600
TATGACGGTG ATTTCTATCG TCGTAA	627

(2) ANGABEN ZU SEQ ID NO:46:

(i) SEQUENZCHARAKTERISTIKA:

- (A) LÄNGE: 219 Aminosäuren
- (B) ART: Aminosäure
- (C) STRANGFORM: Einzelstrang
- (D) TOPOLOGIE: unbekannt

(ii) ART DES MOLEKÜLS: Peptid

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO:46:

Met Gly Pro Pro Gly Leu Ala Gly Pro Pro Gly Glu Ser Gly Arg Glu
1 5 10 15

Gly Ala Pro Gly Ala Glu Gly Ser Pro Gly Arg Asp Gly Ser Pro Gly
20 25 30

Ala Lys Gly Asp Arg Gly Glu Thr Gly Pro Ala Gly Pro Pro Gly Ala
35 40 45

Pro Gly Ala Pro Gly Ala Pro Gly Pro Val Gly Pro Ala Gly Lys Ser
50 55 60

Gly Asp Arg Gly Glu Thr Gly Pro Ala Gly Pro Ala Gly Pro Val Gly
65 70 75 80

Pro Ala Gly Ala Arg Gly Pro Ala Gly Pro Gln Gly Pro Arg Gly Asp
85 90 95

Lys Gly Glu Thr Gly Glu Gln Gly Asp Arg Gly Ile Lys Gly His Arg
100 105 110

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

(2) ANGABEN ZU SEQ ID NO:47:

- (i) SEQUENZCHARAKTERISTIKA:
 - (A) LÄNGE: 95 Basenpaare
 - (B) ART: Nukleinsäure
 - (C) STRANGFORM: Einzelstrang
 - (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: cDNA

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO:47:

GGAATTCTCC CATGGGCCCC CCGGGTCTGG CGGGCCCTCC GGGTGAAAGC GGTCGTGAAG 60

GCGCGCCGGG TGCCGAAGGC AGCCCAGGCC GCGAC 95

(2) ANGABEN ZU SEQ ID NO:48:

(i) SEQUENZCHARAKTERISTIKA:

- (A) LÄNGE: 97 Basenpaare
- (B) ART: Nukleinsäure
- (C) STRANGFORM: Einzelstrang
- (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: cDNA

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO:48:

CTTCGTCGG GTCCGGCGCT GCCATCGGGC CCCCGGTTTC CCCTAGCACC ACTTTGGCCG 60

GGCCGCCCCG GGGGCCCACG TGGCATTATT CGAACCC 97

(2) ANGABEN ZU SEQ ID NO:49:

(i) SEQUENZCHARAKTERISTIKA:

- (A) LÄNGE: 91 Basenpaare
- (B) ART: Nukleinsäure
- (C) STRANGFORM: Einzelstrang
- (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: cDNA

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO:49

GGAATTCGGT GCACCGGGCG CGCCGGGTGC CCCAGGCCCG GTGGGCCCGG CGGGCAAAAG 60

CGGTGATCGT GGCAGACCG GTCCGGCGGG C 91

(2) ANGABEN ZU SEQ ID NO:50:

(i) SEQUENZCHARAKTERISTIKA:

(A) LÄNGE: 91 Basenpaare

(B) ART: Nukleinsäure

(C) STRANGFORM: Einzelstrang

(D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: cDNA

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO:50:

CTCTGGCCAG GCCGCCCGGG CCGGCCAGGC CACCCGGGTC GCCCGCGGGC ACCGGGCCGG 60

CCAGGCGTCC CGGGCGCCAT TATTCGAACC C 91

SEQUENZPROTOKOLL

<110> United States Surgical Corporation

<120> Aminosäure-modifizierte Polypeptide

<130> CIP

<140> EP99119184

<141> 07.10.1999

<150> 09/169,768

<151> 09.10.1998

<160> 50

<170> PatentIn Ver. 2.1

<210> 1

<211> 3170

<212> DNA

<213> Homo sapiens

<400> 1

cagctgtctt atggctatga tgagaaatca accggaggaa tttccgtgcc tggcccatg 60
ggtcctcttg gtcctcgtgg tctccctggc cccctgggtg cacctgggtcc ccaaggcttc 120
caagggtccc ctggtgagcc tggcgagcct ggagcttcag gtcccatggg tcccagaggt 180
ccccaggtc ccctggaaa gaatggagat gatggggaag ctggaaaacc tggtcgtcct 240

ggtgagcgtg ggcctcctgg gcctcagggg gctcgaggat tgcccgggaa acgtggcctc 300
 cctggaatga agggacacag aggtttcagt ggtttggatg gtgccaaggg agatgctggg 360
 cctgctgggc ctaaggggtga gcctggcagc cctgggtgaaa atggagctcc tggtcagatg 420
 ggcccccggtg gcctgcctgg tgagagaggt cgccctggag cccctggccc tgctgggtgct 480
 cgtggaaatg atgggtgctac tgggtgctgcc gggccccctg gtcccaaccg ccccgctggg 540
 cctcctgggt tccctgggtgc tgttgggtgct aaggggtgaag ctgggtccca agggccccga 600
 ggctctgaag gtccccaggg tgtgcgtggg gagcctggcc cccctggccc tgctgggtgct 660
 gctggccctg ctggaaaccc tgggtgctgat ggacagcctg gtgctaaagg tgccaatggg 720
 gctcctggta ttgctgggtgc tcctggcttc cctgggtgcc gagggccctc tggaccccag 780
 ggccccggcg gccctcctgg tcccaagggg aacagcgggtg aacctgggtgc tcctggcagc 840
 aaaggagaca ctgggtgctaa gggagagcct ggccctgttg gtgttcaagg acccctggc 900
 cctgctggag aggaaggaaa gcgaggagct cgagggtgaac ccggaccacac tggcctgccc 960
 ggacccctg gcgagcgtgg tggacctggg agcctgggtt tccctggcgc agatgggtgtt 1020
 gctgggtcca aggggtccgc tgggtgaacgt ggttctcctg gcccgcgtgg ccccaaagga 1080
 tctcctgggtg aagctgggtc tcccggtgaa gctgggtctgc ctgggtgcaa gggctgact 1140
 ggaagccctg gcagccctgg tcctgatggc aaaactggcc cccctgggtc cgccgggtcaa 1200
 gatggctgcc ccggaccccc agggccacct ggtgcccgtg gtcaggctgg tgtgatggga 1260
 ttccctggac ctaaagggtgc tgctggagag cccggcaagg ctggagagcg aggtgttccc 1320
 ggacccctg gcgctgtcgg tcctgctggc aaagatggag aggctggagc tcagggaccc 1380
 cctggccctg ctgggtccgc tggcgagaga ggtgaacaag gccctgctgg ctccccgga 1440
 ttccagggtc tccctgggtc tgctgggtct ccagggtgaag caggcaaacc tgggtgaacag 1500
 ggtgttctcgt gagaccttgg cgccctggc ccctctggag caagaggcga gagaggtttc 1 560
 cctggcgagc gtgggtgtgca aggtccccct ggtcctgctg gaccccgagg ggccaacggg 1620
 gctcccggca acgatgggtgc taagggtgat gctgggtgcc ctggagctcc cggtagccag 1680
 ggcgccccctg gccttcaggg aatgcctggg gaacgtgggtg cagctgggtc tccagggcct 1740
 aagggtgaca gaggtgatgc tgggtccaaa ggtgctgatg gctctcctgg caaagatggc 1800
 gtccgtgggtc tgaccggccc cattgggtct cctggccctg ctgggtgccc tgggtgacaag 1860

ggtgaaagtg gtcccagcgg ccctgctggt cccactggag ctctggtgac ccccgagac 1920
 cgtggtgagc ctggtccccc cggccctgct ggctttgctg gccccctgg tgctga cggc 1980
 caacctggtg ctaaaggcga acctggtgat gctggtgcca aaggcgatgc tggccccct 2040
 gggcctgccg gaccgctgg accccctggc cccattggta atgttgggtgc tcctggagcc 2100
 aaaggtgctc gggcagcgt ggtccccctg gtgctactgg tttccctggt gctgctggcc 2160
 gagtcggtcc tcctggcccc tctggaaatg ctggaccccc tggccctcct ggtcctgctg 2220
 gcaaagaagg cggcaaaggc ccccggtggt agactggccc tgctggacgt cctggtgaag 2280
 ttgggtcccc tgggtccccct ggccctgctg gcgagaaagg atccctggt gctgatggtc 2340
 ctgctggtgc tcctggtact cccgggcctc aagggtattgc tggacagcgt ggtgtggtcg 2400
 gcctgcctgg tcagagagga gagagaggct tcctggtct tcctggcccc tctggtgaac 2460
 ctggcaaaca aggtccctct ggagcaagtg gtgaacgtgg tcccccggt cccatgggccc 2520
 cccctggatt ggctggaccc cctggtgaat ctggacgtga gggggctcct gctgccgaag 2580
 gttcccctgg acgagacggt tctcctggcg ccaaggggtga ccgtggtgag accggccccg 2640
 ctggaccccc tgggtgctcct ggtgctcctg gtgccccctgg ccccggtggc cctgctggca 2700
 agagtgggtga tcgtggtgag actggtcctg ctggtccgc cgggtccgctc ggccccgctg 2760
 gcgcccgtgg ccccgccgga ccccaaggcc cccgtggtga caagg gtgag acaggcgaac 2820
 agggcgacag aggcataaag ggtcaccgtg gcttctctgg cctccagggt cccctggcc 2880
 ctctggctc tcctggtgaa caaggctcct ctggagcctc tggctcctgct ggtccccgag 2940
 gtccccctgg ctctgctggt gctcctggca aagatggact caacggtctc cctggcccca 3000
 ttgggcccc cgggtcctgc ggtcgactg gtgatgctgg tcctgttggc cccccggcc 3060
 ctctggacc tcctggtccc cctggtcctc ccagcgctgg ttctgacttc agcttcctcc 3120
 cccagccacc tcaagagaag gctcacgatg gtggccgcta ctaccgggct 3170

<210> 2
 <211> 240
 <212> DNA
 <213> Homo sapiens

<400> 2

cagctgtctt atggctatga tgagaaatca accggaggaa ttcccggtgcc tggcccatg 60
 ggccccctctg gtcctcgtgg tctccctggc cccctgggtg cacctgggtcc ccaaggcttc 120
 caaggtcccc ctggtgagcc tggcgagcct ggagcttcag gtcccatggg tccccgaggt 180
 cccccaggtc cccctggaaa gaatggagat gatggggaag ctggaaaacc tggtcgtcct 240

<210> 3

<211> 100

<212> DNA

<213> Künstliche Sequenz

<220>

<223> Beschreibung der künstlichen Sequenz: DNA-Sequenz
 codierend ein Gen-Kollagen-artiges Peptid

<400> 3

ggatccatgg ggctcgctgg cccaccgggc gaaccgggtc cgccaggccc gaaagggtccg 60
 cgtggcgata gcggggtccc gggcgattcc taatggatcc 100

<210> 4

<211> 21

<212> PRT

<213> Künstliche Sequenz

<220>

<223> Beschreibung der künstlichen Sequenz:

Aminosäuresequenz eines Kollagen-artigen Peptids

<400> 4

Gly Leu Ala Gly Pro Pro Gly Glu Pro Gly Pro Pro Gly Pro Lys Gly

1

5

10

15

Pro Arg Gly Asp Ser

20

<210> 5

<211> 330

<212> DNA

<213> Homo sapiens

<400> 5

cagcgggcca ggaagaagaa taagaactgc cggcgccact cgctctatgt ggacttcagc 60
 gatgtgggct ggaatgactg gattgtggcc ccaccaggct accaggcctt ctactgccat 120
 ggggactgcc cctttccact ggctgaccac ctcaactcaa ccaaccatgc cattgtgcag 180
 accctgggtca attctgtcaa ttccagtatc cccaaagcct gttgtgtgcc cactgaactg 240
 agtgccatct ccatgctgta cctggatgag tatgataagg tgggtactgaa aaattatcag 300
 gagatggtag tagaggggatg tgggtgccgc 330

<210> 6

<211> 1169

<212> PRT

<213> Künstliche Sequenz

<220>

<223> Beschreibung der künstlichen Sequenz:

Aminosäuresequenz eines Kollagen-I-(alpha1)/

BMP-2B chimären Proteins

<400> 6

Gln Leu Ser Tyr Gly Tyr Asp Glu Lys Ser Thr Gly Gly Ile Ser Val

1 5 10 15

Pro Gly Pro Met Gly Pro Ser Gly Pro Arg Gly Leu Pro Gly Pro Pro

20 25 30

Gly Ala Pro Gly Pro Gln Gly Phe Gln Gly Pro Pro Gly Glu Pro Gly

35 40 45

Glu Pro Gly Ala Ser Gly Pro Met Gly Pro Arg Gly Pro Pro Gly Pro

50 55 60

Pro Gly Lys Asn Gly Asp Asp Gly Glu Ala Gly Lys Pro Gly Arg Pro

65 70 75 80

Gly Glu Arg Gly Pro Pro Gly Pro Gln Gly Ala Arg Gly Leu Pro Gly

85 90 95

Thr Ala Gly Leu Pro Gly Met Lys Gly His Arg Gly Phe Ser Gly Leu			
100	105	110	
Asp Gly Ala Lys Gly Asp Ala Gly Pro Ala Gly Pro Lys Gly Glu Pro			
115	120	125	
Gly Ser Pro Gly Glu Asn Gly Ala Pro Gly Gln Met Gly Pro Arg Gly			
130	135	140	
Leu Pro Gly Glu Arg Gly Arg Pro Gly Ala Pro Gly Pro Ala Gly Ala			
145	150	155	160
Arg Gly Asn Asp Gly Ala Thr Gly Ala Ala Gly Pro Pro Gly Pro Thr			
165	170	175	
Gly Pro Ala Gly Pro Pro Gly Phe Pro Gly Ala Val Gly Ala Lys Gly			
180	185	190	
Glu Ala Gly Pro Gln Gly Pro Arg Gly Ser Glu Gly Pro Gln Gly Val			
195	200	205	
Arg Gly Glu Pro Gly Pro Pro Gly Pro Ala Gly Ala Ala Gly Pro Ala			
210	215	220	
Gly Asn Pro Gly Ala Asp Gly Gln Pro Gly Ala Lys Gly Ala Asn Gly			
225	230	235	240
Ala Pro Gly Ile Ala Gly Ala Pro Gly Phe Pro Gly Ala Arg Gly Pro			

245

250

255

Ser Gly Pro Gln Gly Pro Gly Gly Pro Pro Gly Pro Lys Gly Asn Ser

260

265

270

Gly Glu Pro Gly Ala Pro Gly Ser Lys Gly Asp Thr Gly Ala Lys Gly

275

280

285

Glu Pro Gly Pro Val Gly Val Gln Gly Pro Pro Gly Pro Ala Gly Glu

290

295

300

Glu Gly Lys Arg Gly Ala Arg Gly Glu Pro Gly Pro Thr Gly Leu Pro

305

310

315

320

Gly Pro Pro Gly Glu Arg Gly Gly Pro Gly Ser Arg Gly Phe Pro Gly

325

330

335

Ala Asp Gly Val Ala Gly Pro Lys Gly Pro Ala Gly Glu Arg Gly Ser

340

345

350

Pro Gly Pro Ala Gly Pro Lys Gly Ser Pro Gly Glu Ala Gly Arg Pro

355

360

365

Gly Glu Ala Gly Leu Pro Gly Ala Lys Gly Leu Thr Gly Ser Pro Gly

370

375

380

Ser Pro Gly Pro Asp Gly Lys Thr Gly Pro Pro Gly Pro Ala Gly Gln

385	390	395	400
Asp Gly Arg Pro Gly Pro Pro Gly Pro Pro Gly Ala Arg Gly Gln Ala			
	405	410	415
Gly Val Met Gly Phe Pro Gly Pro Lys Gly Ala Ala Gly Glu Pro Gly			
	420	425	430
Lys Ala Gly Glu Arg Gly Val Pro Gly Pro Pro Gly Ala Val Gly Pro			
	435	440	445
Ala Gly Lys Asp Gly Glu Ala Gly Ala Gln Gly Pro Pro Gly Pro Ala			
	450	455	460
Gly Pro Ala Gly Glu Arg Gly Glu Gln Gly Pro Ala Gly Ser Pro Gly			
465	470	475	480
Phe Gln Gly Leu Pro Gly Pro Ala Gly Pro Pro Gly Glu Ala Gly Lys			
	485	490	495
Pro Gly Glu Gln Gly Val Pro Gly Asp Leu Gly Ala Pro Gly Pro Ser			
	500	505	510
Gly Ala Arg Gly Glu Arg Gly Phe Pro Gly Glu Arg Gly Val Gln Gly			
	515	520	525
Pro Pro Gly Pro Ala Gly Pro Arg Gly Ala Asn Gly Ala Pro Gly Asn			
	530	535	540

Asp Gly Ala Lys Gly Asp Ala Gly Ala Pro Gly Ala Pro Gly Ser Gln

545 550 555 560

Gly Ala Pro Gly Leu Gln Gly Met Pro Gly Glu Arg Gly Ala Ala Gly

565 570 575

Leu Pro Gly Pro Lys Gly Asp Arg Gly Asp Ala Gly Pro Lys Gly Ala

580 585 590

Asp Gly Ser Pro Gly Lys Asp Gly Val Arg Gly Leu Thr Gly Pro Ile

595 600 605

Gly Pro Pro Gly Pro Ala Gly Ala Pro Gly Asp Lys Gly Glu Ser Gly

610 615 620

Pro Ser Gly Pro Ala Gly Pro Thr Gly Ala Arg Gly Ala Pro Gly Asp

625 630 635 640

Arg Gly Glu Pro Gly Pro Pro Gly Pro Ala Gly Phe Ala Gly Pro Pro

645 650 655

Gly Ala Asp Gly Gln Pro Gly Ala Lys Gly Glu Pro Gly Asp Ala Gly

660 665 670

Ala Lys Gly Asp Ala Gly Pro Pro Gly Pro Ala Gly Pro Ala Gly Pro

675 680 685

Pro Gly Pro Ile Gly Asn Val Gly Ala Pro Gly Ala Lys Gly Ala Arg

690

695

700

Gly Ser Ala Gly Pro Pro Gly Ala Thr Gly Phe Pro Gly Ala Ala Gly

705

710

715

720

Arg Val Gly Pro Pro Gly Pro Ser Gly Asn Ala Gly Pro Pro Gly Pro

725

730

735

Pro Gly Pro Ala Gly Lys Glu Gly Gly Lys Gly Pro Arg Gly Glu Thr

740

745

750

Gly Pro Ala Gly Arg Pro Gly Glu Val Gly Pro Pro Gly Pro Pro Gly

755

760

765

Pro Ala Gly Glu Lys Gly Ser Pro Gly Ala Asp Gly Pro Ala Gly Ala

770

775

780

Pro Gly Thr Pro Gly Pro Gln Gly Ile Ala Gly Gln Arg Gly Val Val

785

790

795

800

Gly Leu Pro Gly Gln Arg Gly Glu Arg Gly Phe Pro Gly Leu Pro Gly

805

810

815

Pro Ser Gly Glu Pro Gly Lys Gln Gly Pro Ser Gly Ala Ser Gly Glu

820

825

830

Arg Gly Pro Pro Gly Pro Met Gly Pro Pro Gly Leu Ala Gly Pro Pro

835

840

845

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

850

855

860

Arg Asp Gly Ser Pro Gly Ala Lys Gly Asp Arg Gly Glu Thr Gly Pro

865

870

875

880

Ala Gly Pro Pro Gly Ala Xaa Gly Ala Xaa Gly Ala Pro Gly Pro Val

885

890

895

Gly Pro Ala Gly Lys Ser Gly Asp Arg Gly Glu Thr Gly Pro Ala Gly

900

905

910

Pro Ala Gly Pro Val Gly Pro Ala Gly Ala Arg Gly Pro Ala Gly Pro

915

920

925

Gln Gly Pro Arg Gly Asp Lys Gly Glu Thr Gly Glu Gln Gly Asp Arg

930

935

940

Gly Ile Lys Gly His Arg Gly Phe Ser Gly Leu Gln Gly Pro Pro Gly

945

950

955

960

Pro Pro Gly Ser Pro Gly Glu Gln Gly Pro Ser Gly Ala Ser Gly Pro

965

970

975

Ala Gly Pro Arg Gly Pro Pro Gly Ser Ala Gly Ala Pro Gly Lys Asp			
980	985	990	
Gly Leu Asn Gly Leu Pro Gly Pro Ile Gly Pro Pro Gly Pro Arg Gly			
995	1000	1005	
Arg Thr Gly Asp Ala Gly Pro Val Gly Pro Pro Gly Pro Pro Gly Pro			
1010	1015	1020	
Pro Gly Pro Pro Gly Pro Pro Ser Ala Gly Phe Asp Phe Ser Phe Leu			
1025	1030	1035	1040
Pro Gln Pro Pro Gln Glu Lys Ala His Asp Gly Gly Arg Tyr Tyr Arg			
1045	1050	1055	
Ala Arg Ser Gln Arg Ala Arg Lys Lys Asn Lys Asn Cys Arg Arg His			
1060	1065	1070	
Ser Leu Tyr Val Asp Phe Ser Asp Val Gly Trp Asn Asp Trp Ile Val			
1075	1080	1085	
Ala Pro Pro Gly Tyr Gln Ala Phe Tyr Cys His Gly Asp Cys Pro Phe			
1090	1095	1100	
Pro Leu Ala Asp His Leu Asn Ser Thr Asn His Ala Ile Val Gln Thr			
1105	1110	1115	1120
Leu Val Asn Ser Val Asn Ser Ser Ile Pro Lys Ala Cys Cys Val Pro			

1125

1130

1135

Thr Glu Leu Ser Ala Ile Ser Met Leu Tyr Leu Asp Glu Tyr Asp Lys

1140

1145

1150

Val Val Leu Lys Asn Tyr Gln Glu Met Val Val Glu Gly Cys Gly Cys

1155

1160

1165

Arg

<210> 7

<211> 3531

<212> DNA

<213> Künstliche Sequenz

<220>

<223> Beschreibung der künstlichen Sequenz:

Aminosäuresequenz eines Kollagen-I-(alpha1)/

BMP-2B chimären Proteins

<400> 7

```

gggaaggatt tccatttccc agctgtctta tggctatgat gagaaatcaa ccggaggaat 60
ttccgtgcct ggccccatgg gtccctctgg tectcgtggc ctccttggcc cccctggtgc 120
acctgggtccc caaggcttcc aagggtcccc tgggtgagcct ggcgagcctg gagcttcagg 180
tcccatgggt ccccgaggtc ccccgaggtc ccctggaaag aatggagatg atggggaagc 240

```

tggaaaacct ggtcgtcctg gtgagcgtgg gcctcctggg cctcaggggtg ctcgaggatt 300
 gcccggaaaca gctggcctcc ctggaatgaa gggacacaga ggtttcagtg gtttggatgg 360
 tgccaaggga gatgctggtc ctgctgggcc taaggggtgag cctggcagcc ctggtgaaaa 420
 tggagctcct ggtcagatgg gccccgtgg cctgcctggt gagagaggtc gccctggagc 480
 ccctggccct gctggtgctc gtggaaatga tgggtgctact ggtgctgccg ggccccctgg 540
 tcccaccggc cccgctggtc ctctggcctt ccctgggtgct gttggtgcta aggggtgaagc 600
 tgggtcccaa gggccccgag gctctgaagg tcccaggggt gtgctggtg agcctggccc 660
 ccctggccct gctggtgctg ctggccctgc tggaaacct ggtgctgatg gacagcctgg 720
 tgctaaagggt gccaatggtg ctctgggtat tgctggtgct cctggcttcc ctggtgcccc 780
 agggccctct ggaccccagg gccccggcgg ccctcctggt cccaagggta acagcgggtga 840
 acctggtgct cctggcagca aaggagacac tgggtgctaag ggagagcctg gccctggttg 900
 tgttcaagga cccctggcc ctgctggaga ggaaggaaaag cgaggagctc gaggtgaacc 960
 cggaccact ggctgcccc gacccctgg cgagcgtggt ggacctggtg gccgtggttt 1020
 ccctggcgca gatggtgttg ctgggtccaa ggggtcccgt ggtgaacgtg gttctcctgg 1080
 cccgctggc cccaaaggat ctctgggtga agctggtcgt cccggtgaag ctggtctgcc 1140
 tggtgccaag ggtctgactg gaagccctgg cagccctggt cctgatggca aaactggccc 1200
 ccctggtccc gccggtcaag atggtcgccc cggaccccca ggcccacctg gtgcccgtgg 1260
 tcaggctggt gtgatgggat tccctggacc taaagggtgct gctggagagc ccggcaaggc 1320
 tggagagcga ggtgttcccc gacccctgg cgctgtcggc cctgctggca aagatggaga 1380
 ggctggagct cagggacccc ctggccctgc tgggtcccgt ggcgagagag gtgaacaagg 1440
 ccctgctggc tccccggat tccagggtct ccctggtcct gctggtcctc cagggtgaagc 1500
 aggcaaacct ggtgaacagg gtgttctctg agaccttggc gcccctggcc cctctggagc 1560
 aagaggcgag agaggtttcc ctggcgagcg tgggtgcaaa ggtccccctg gtcctgctgg 1 620
 accccgaggg gccaacggtg ctcccggcaa cgatggtgct aagggtgatg ctggtgcccc 1680
 tggagctccc ggtagccagg gcgcccctgg ccttcaggga atgcctggtg aacgtggtgc 1740
 agctggtctt ccagggccta aggggtgacag aggtgatgct ggtcccaaag gtgctgatgg 1800
 ctctcctggc aaagatggcg tccgtggtct gaccggcccc attggtcctc ctggccctgc 1860
 tgggtgcccc ggtgacaagg gtgaaagtgg tcccagcggc cctgctggtc ccactggagc 1920

tcgtggtgcc cccggagacc gtggtgagcc tgggtccccc ggccctgctg gcttbgtctg 1980
 ccccccctggt gctgacggcc aacctggtgc taaaggcgaa cctggtgatg ctggtg ccaa 2040
 aggcgatggg tccccctggg cctgcccggac ccgctggacc ccctggcccc attggtaatg 2100
 ttggtgctcc tggagccaaa ggtgctcgcg gcagcgtctg tccccctggt gctactgggt 2160
 tccttggtgc tgetggccga gtcggtcctc ctggccctc tggaaatgct ggacccccctg 2220
 gccctcctgg tcctgctggc aaagaaggcg gcaaaggctc ccgtggtgag actggccctg 2280
 ctggacgtcc tgggtgaagt ggtccccctg gtccccctgg ccctgctggc gagaaaggat 2340
 cccctggtgc tgatggtcct gctggtgctc ctggtactcc cgggcctcaa ggtattgctg 2400
 gacagcgtgg tgtggctggc ctgcctggtc agagaggaga gagaggcttc cctggtcttc 2460
 ctggccctc tgggtgaacct ggcaaacaag gtccctctgg agcaagtggg gaacgtggtc 2520
 cccccggtcc catgggcccc cctggattgg ctggaccccc tgggtgaatct ggacgtgagg 2580
 gggctcctgc tgccgaaggc tccccctggac gagacggctc tcctggcgcc aagggtgacc 2640
 gtggtgagac cggccccgct ggacccccctg gtgctctggt gctctggtgc ccctggcccc 2700
 gttggccctg ctggcaagag tgggtgatcg ggtgagactg gtccctgctgg tcccgccggg 2760
 ccctgctggc ccgctggcgc ccgtggcccc gccggacccc aaggcccccg tggtgacaag 2820
 ggtgagacag gcgaacaggg cgacagaggc ataaagggtc accgt ggctt ctctggcctc 2880
 cagggtcccc ctggccctcc tggctctcct ggtgaacaag gtccctctgg agcctctggt 2940
 cctgctggtc cccgaggtcc ccctggctct gctggtgctc ctggcaaaga tggactcaac 3000
 ggtctccctg gcccattgg gccccctggt cctcgcggtc gcaactggtga tgctggctcct 3060
 gttggtcccc cggccctcc tggacctcct ggtccccctg gtccctccag cgctggtttc 3120
 gacttcagct tcctccccca gccacctcaa gagaaggctc acgatggtgg ccgctactac 3180
 cgggctagat cccagcgggc caggaagaag aataagaact gccggcgcca ctgctctat 3240
 gtggacttca gcgatgtggg ctggaatgac tggattgtgg cccaccagg ctaccaggcc 3300
 ttctactgcc atggggactg cccctttcca ctggctgacc acctcaactc aaccaacct 3360
 gccattgtgc agaccctggt caattctgtc aattccagta tccccaaagc ctgttgtgtg 3420
 cccactgaac tgagtgccat ctccatgctg tacctggatg agtatgataa ggtggtactg 3480
 aaaaattatc aggagatggt agtagaggga tgtgggtgcc gctaaaagct t 3531

<210> 8
 <211> 1171
 <212> PRT
 <213> Künstliche Sequenz

<220>

<223> Beschreibung der künstlichen Sequenz:
 Aminosäuresequenz eines Kollagen-I-(alpha1)/
 TGF-beta1 chimären Proteins

<400> 8

Gln Leu Ser Tyr Gly Tyr Asp Glu Lys Ser Thr Gly Gly Ile Ser Val
 1 5 10 15

Pro Gly Pro Met Gly Pro Ser Gly Pro Arg Gly Leu Pro Gly Pro Pro
 20 25 30

Gly Ala Pro Gly Pro Gln Gly Phe Gln Gly Pro Pro Gly Glu Pro Gly
 35 40 45

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

Pro Gly Lys Asn Gly Asp Asp Gly Glu Ala Gly Lys Pro Gly Arg Pro
 65 70 75 80

Gly Glu Arg Gly Pro Pro Gly Pro Gln Gly Ala Arg Gly Leu Pro Gly

85

90

95

Thr Ala Gly Leu Pro Gly Met Lys Gly His Arg Gly Phe Ser Gly Leu

100

105

110

Asp Gly Ala Lys Gly Asp Ala Gly Pro Ala Gly Pro Lys Gly Gl u Pro

115

120

125

Gly Ser Pro Gly Glu Asn Gly Ala Pro Gly Gln Met Gly Pro Arg Gly

130

135

140

Leu Pro Gly Glu Arg Gly Arg Pro Gly Ala Pro Gl y Pro Ala Gly Ala

145

150

155

160

Arg Gly Asn Asp Gly Ala Thr Gly Ala Ala Gly Pro Pro Gly Pro Thr

165

170

175

Gly Pro Ala Gly Pro Pro Gly Phe Pr o Gly Ala Val Gly Ala Lys Gly

180

185

190

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

195

200

205

Arg Gly Glu Pro Gly Pro Pro Gly Pro Ala Gly Ala Ala Gly Pro Ala

210

215

220

Gly Asn Pro Gly Ala Asp Gly Gln Pro Gly Ala Lys Gly Ala Asn Gly

225

230

235

240

Ala Pro Gly Ile Ala Gly Ala Pro Gly Phe Pro Gly Ala Arg Gly Pro

245

250

255

Ser Gly Pro Gln Gly Pro Gly Gly Pro Pro Gly Pro Lys Gly Asn Ser

260

265

270

Gly Glu Pro Gly Ala Pro Gly Ser Lys Gly Asp Thr Gly Ala Lys Gly

275

280

285

Glu Pro Gly Pro Val Gly Val Gln Gly Pro Pro Gly Pro Ala Gly Glu

290

295

300

Glu Gly Lys Arg Gly Ala Arg Gly Glu Pro Gly Pro Thr Gly Leu Pro

305

310

315

320

Gly Pro Pro Gly Glu Arg Gly Gly Pro Gly Ser Arg Gly Phe Pro Gly

325

330

335

Ala Asp Gly Val Ala Gly Pro Lys Gly Pro Ala Gly Glu Arg Gly Ser

340

345

350

Pro Gly Pro Ala Gly Pro Lys Gly Ser Pro Gly Glu Ala Gly Arg Pro

355

360

365

Gly Glu Ala Gly Leu Pro Gly Ala Lys Gly Leu Thr Gly Ser Pro Gly

370

375

380

Ser Pro Gly Pro Asp Gly Lys Thr Gly Pro Pro Gly Pro Ala Gly Gln

385

390

395

400

Asp Gly Arg Pro Gly Pro Pro Gly Pro Pro Gly Ala Arg Gly Gln Ala

405

410

415

Gly Val Met Gly Phe Pro Gly Pro Lys Gly Ala Ala Gly Glu Pro Gly

420

425

430

Lys Ala Gly Glu Arg Gly Val Pro Gly Pro Pro Gly Ala Val Gly Pro

435

440

445

Ala Gly Lys Asp Gly Glu Ala Gly Ala Gln Gly Pro Pro Gly Pro Ala

450

455

460

Gly Pro Ala Gly Glu Arg Gly Glu Gln Gly Pro Ala Gly Ser Pro Gly

465

470

475

480

Phe Gln Gly Leu Pro Gly Pro Ala Gly Pro Pro Gly Glu Ala Gly Lys

485

490

495

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

500

505

510

Gly Ala Arg Gly Glu Arg Gly Phe Pro Gly Glu Arg Gly Val Gln Gly

515

520

525

Pro Pro Gly Pro Ala Gly Pro Arg Gly Ala Asn Gly Ala Pro Gly Asn

530

535

540

Asp Gly Ala Lys Gly Asp Ala Gly Ala Pro Gly Ala Pro Gly Ser Gln

545

550

555

560

Gly Ala Pro Gly Leu Gln Gly Met Pro Gly Glu Arg Gly Ala Ala Gly

565

570

575

Leu Pro Gly Pro Lys Gly Asp Arg Gly Asp Ala Gly Pro Lys Gly Ala

580

585

590

Asp Gly Ser Pro Gly Lys Asp Gly Val Arg Gly Leu Thr Gly Pro Ile

595

600

605

Gly Pro Pro Gly Pro Ala Gly Ala Pro Gly Asp Lys Gly Glu Ser Gly

610

615

620

Pro Ser Gly Pro Ala Gly Pro Thr Gly Ala Arg Gly Ala Pro Gly Asp

625

630

635

640

Arg Gly Glu Pro Gly Pro Pro Gly Pro Ala Gly Phe Ala Gly Pro Pro

645

650

655

Gly Ala Asp Gly Gln Pro Gly Ala Lys Gly Glu Pro Gly Asp Ala Gly

660

665

670

Ala Lys Gly Asp Ala Gly Pro Pro Gly Pro Ala Gly Pro Ala Gly Pro

675

680

685

Pro Gly Pro Ile Gly Asn Val Gly Ala Pro Gly Ala Lys Gly Ala Arg

690

695

700

Gly Ser Ala Gly Pro Pro Gly Ala Thr Gly Phe Pro Gly Ala Ala Gly

705

710

715

720

Arg Val Gly Pro Pro Gly Pro Ser Gly Asn Ala Gly Pro Pro Gly Pro

725

730

735

Pro Gly Pro Ala Gly Lys Glu Gly Gly Lys Gly Pro Arg Gly Glu Thr

740

745

750

Gly Pro Ala Gly Arg Pro Gly Glu Val Gly Pro Pro Gly Pro Pro Gly

755

760

765

Pro Ala Gly Glu Lys Gly Ser Pro Gly Ala Asp Gly Pro Ala Gly Ala

770

775

780

Pro Gly Thr Pro Gly Pro Gln Gly Ile Ala Gly Gln Arg Gly Val Val

785

790

795

800

Gly Leu Pro Gly Gln Arg Gly Glu Arg Gly Phe Pro Gly Leu Pro Gly

805

810

815

Pro Ser Gly Glu Pro Gly Lys Gln Gly Pro Ser Gly Ala Ser Gly Glu

820

825

830

Arg Gly Pro Pro Gly Pro Met Gly Pro Pro Gly Leu Ala Gly Pro Pro

835

840

845

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

850

855

860

Arg Asp Gly Ser Pro Gly Ala Lys Gly Asp Arg Gly Glu Thr Gly Pro

865

870

875

880

Ala Gly Pro Pro Gly Ala Pro Gly Ala Pro Gly Ala Pro Gly Pro Val

885

890

895

Gly Pro Ala Gly Lys Ser Gly Asp Arg Gly Glu Thr Gly Pro Ala Gly

900

905

910

Pro Ala Gly Pro Val Gly Pro Ala Gly Ala Arg Gly Pro Ala Gly Pro

915

920

925

Gln Gly Pro Arg Gly Asp Lys Gly Glu Thr Gly Glu Gln Gly Asp Arg

930

935

940

Gly Ile Lys Gly His Arg Gly Phe Ser Gly Leu Gln Gly Pro Pro Gly

945

950

955

960

Pro Pro Gly Ser Pro Gly Glu Gln Gly Pro Ser Gly Ala Ser Gly Pro

965

970

975

Ala Gly Pro Arg Gly Pro Pro Gly Ser Ala Gly Ala Pro Gly Lys Asp

980

985

990

Gly Leu Asn Gly Leu Pro Gly Pro Ile Gly Pro Pro Gly Pro Arg Gly

995

1000

1005

Arg Thr Gly Asp Ala Gly Pro Val Gly Pro Pro Gly Pro Pro Gly Pro

1010

1015

1020

Pro Gly Pro Pro Gly Pro Pro Ser Ala Gly Phe Asp Phe Ser Phe Leu

1025

1030

1035

1040

Pro Gln Pro Pro Gln Glu Lys Ala His Asp Gly Gly Arg Tyr Tyr Arg

1045

1050

1055

Ala Arg Ser Ala Leu Asp Thr Asn Tyr Cys Phe Ser Ser Thr Glu Lys

1060

1065

1070

Asn Cys Cys Val Arg Gln Leu Tyr Ile Asp Phe Arg Lys Asp Leu Gly

1075

1080

1085

Trp Lys Trp Ile His Glu Pro Lys Gly Tyr His Ala Asn Phe Cys Leu

1090

1095

1100

Gly Pro Cys Pro Tyr Ile Trp Ser Leu Asp Thr Gln Tyr Ser Lys Val

1105

1110

1115

1120

Leu Ala Leu Tyr Asn Gln His Asn Pro Gly Ala Ser Ala Ala Pro Cys

1125

1130

1135

Cys Val Pro Gln Ala Leu Glu Pro Leu Pro Ile Val Tyr Tyr Val Gly

1140

1145

1150

Arg Lys Pro Lys Val Glu Gln Leu Ser Asn Met Ile Val Arg Ser Cys

1155

1160

1165

Lys Cys Ser

1170

<210> 9

<211> 3541

<212> DNA

<213> Künstliche Sequenz

<220>

<223> Beschreibung der künstlichen Sequenz:

DNA-Sequenz codierend ein Kollagen I (alpha1)/

TGF-beta1 chimäres Protein

<400> 9

gggaaggatt tccatttccc agctgtctta tggctatgat gagaaatcaa ccggaggaat 60

ttccgtgcct ggccccatgg gtccctctgg tccctcgtgg ctccctggcc cccctggtg c 120
 acctggtccc caaggcttcc aagggtcccc tggtagacct ggagagacct gagcttcagg 180
 tcccatgggt ccccgaggtc cccaggtcc cctcggaaag aatggagatg atggggaagc 240
 tggaaaacct ggtcgtcctg gtgagcgtgg gcctcctggg cctcaggggtg ctcgaggatt 300
 gcccggaaca gctggcctcc ctggaatgaa gggacacaga ggtttcagtg gtttggatgg 360
 tgccaaggga gatgctgggt ctgctgggtcc taagggtgag cctggcagcc ctggtgaaaa 420
 tggagctcct ggtcagatgg gccccctggg cctgcctggg gagagagggt gccctggagc 480
 ccctggccct gctggtgctc gtggaaatga tgggtctact ggtgctgccc ggc cccctgg 540
 tcccaccggc cccgctgggt ctccctggctt cctggtgct gttggtgcta aggggaagc 600
 tgggtcccaa gggccccgag gctctgaagg tcccaggggt gtgctggtg agcctggccc 660
 ccctggccct gctggtgctg ctggccctgc tggaaacct ggtgctgatg gacagcctgg 720
 tgctaaagggt gccaatggtg ctccctggtat tgctggtgct cctggcttcc ctggtgccc 780
 agggccctct ggaccccagg gccccggcgg cctcctgggt cccaagggtg acagcgggtg 840
 acctggtgct cctggcagca aaggagacac tgggtgctaag ggagagacct gccctggtgg 900
 tgttcaagga cccctggcc ctgctggaga ggaaggaaag cgaggagc tc gaggtgaacc 960
 cggaccact ggctgcccc gacccctgg cgagcgtgggt ggacctggtg gccgtggttt 1020
 ccctggcgca gatggtgttg ctgggtccaa ggggtccgct ggtgaacctg gttctcctgg 1080
 cccgctggc cccaaaggat ctccctggtg agctggtcgt cccggtgaag ctggtctgcc 1140
 tgggtccaag ggtctgactg gaagccctgg cagccctgggt cctgatggca aaactggccc 1200
 ccctggtccc gccggtcaag atggtcgccc cggaccccca ggccacctg gtgcccgtgg 1260
 tcaggctgggt gtgatgggat tccctggacc taaagggtgct gctggagagc ccggcaaggc 1320
 tggagagcga ggtgttcccg gacccctgg cgctgtcggt cc tgctggca aagatggaga 1380
 ggctggagct cagggaacccc ctggccctgc tgggtccgct ggagagagag gtgaacaagg 1440
 cctgctggc tcccccgat tccagggtct cctggtcct gctggtcctc caggtgaagc 1500
 aggcaaacct ggtgaacagg gtgttcctgg agacctggc gccctggcc cctctggagc 1560
 aagaggcgag agaggtttcc ctggcgagcg tgggtgtcaa ggtccccctg gtcctgctgg 1620
 accccgaggg gccaacggtg ctccccgcaa cgatggtgct aagggtgatg ctggtgcccc 1680

tggagctccc ggtagccagg gcgcccctgg ccttcaggga atgcctggtg aacgtggtgc 1740
 agctggtcct ccagggccta aggggtgacag aggtgat.gct ggtcccaaag gtgctgatgg 1800
 ctctcctggc aaagatggcg tccgtggtct gaccggcccc attggtcctc ctggccctgc 1860
 tggtgcccct ggtgacaagg gtgaaagtgg tcccagcggc cctgctggtc ccaactggagc 1920
 tcgtggtgcc cccggagacc gtggtgagcc tgggtcccccc ggccctgctg gctttgctgg 1980
 cccccctggt gctgacggcc aacctggtgc taaaggcgaa cctggtgatg ctggtgccaa 2040
 aggcgatgct ggtccccctg ggccctgcgg acccgctgga cccccctggc ccattggtaa 2100
 tgttggtgct cctggagcca aagggtgctcg cggcagcgtt ggtccccctg gtgctactgg 2160
 ttccctggt gctgctggcc gagtcggctc t cctggcccc tctggaaatg ctggaccccc 2220
 tggccctcct ggtcctgctg gcaaagaagg cggcaaaggc ccccggtgtg agactggccc 2280
 tgctggacgt cctggtgaag ttggtcccc tgggtccccct ggccctgctg gcgagaaagg 2340
 atccccctgt gctgatggtc ctgctggtgc tcctggtact cccgggcctc aaggtattgc 2400
 tggacagcgt ggtgtggtcg gcctgcctgg tcagagagga gagagaggct tcctggtct 2460
 tcctggcccc tctggtgaac ctggcaaaca aggtccctct ggagcaagtg gtgaacgtgg 2520
 tccccccggt cccatgggccc cccctggatt ggctggaccc cctggtgaat ctggacgtga 2580
 gggggctcct gctgccgaag gttccc ctgg acgagacggc tctcctggcg ccaagggtga 2640
 ccgtggtgag accggccccg ctggaccccc tgggtgctcct ggtgctcctg gtgcccctgg 2700
 ccccgttggc cctgctggca agagtgggtga tcgtggtgag actggtcctg ctggtcccg 2760
 cggccccgtc ggccccgctg gcgcccgtgg ccccgccgga ccccaaggcc cccgtggtga 2820
 caagggtgag acaggcgaa acgggacag aggcataaag ggtcacctg gcttctctgg 2880
 cctccagggt cccccctggc ctctggctc tcctggtgaa caaggctcct ctggagcctc 2940
 tggctcctgt ggtccccgag gtccccctgg ctctgctggt gctcctggca aagatggact 3000
 caacgggtct cctggcccca ttgggcccc tggctcctgc ggtcgactg gtgatgctgg 3060
 tcctgttggt cccccgggccc ctctggacc tcctggtccc cctggtcctc ccagcgtgg 3120
 ttctgacttc agcttctctc cccagccacc tcaagagaag gctcacgatg gtggccgcta 3180
 ctaccgggct agatctgccc tggacaccaa ctattgcttc agctccacgg agaagaactg 3240
 ctgcgtgcgg cagctgtaca ttgacttccg caaggacctc ggctggaagt ggatccacga 3300
 gcccaggggc taccatgcca acttctgcct cgggcccctg ccctacattt ggagcctgga 3360

cacgcagtac agcaagggtcc tggccctgta caaccagcat aaccggggcg cctegggcggc 3420
 gccgtgctgc gtgccgcagg cgctggagcc gctgcccatc gtgtactacg tgggcccga 3480
 gcccaagggtg gagcagctgt ccaacatgat cgtgcgctcc tgcaagtgc gctgatctag 3540
 a 3541

<210> 10

<211> 1388

<212> PRT

<213> Künstliche Sequenz

<220>

<223> Beschreibung der künstlichen Sequenz:
 Aminosäuresequenz eines Kollagen-I-(alpha1)/
 Decorin-chimären Proteins

<400> 10

Gln Leu Ser Tyr Gly Tyr Asp Glu Lys Ser Thr Gly Gly Ile Ser Val

1 5 10 15

Pro Gly Pro Met Gly Pro Ser Gly Pro Arg Gly Leu Pro Gly Pro Pro

20 25 30

Gly Ala Pro Gly Pro Gln Gly Phe Gln Gly Pro Pro Gly Glu Pro Gly

35 40 45

Glu Pro Gly Ala Ser Gly Pro Met Gly Pro Arg Gly Pro Pro Gly Pro

50

55

60

Pro Gly Lys Asn Gly Asp Asp Gly Glu Ala Gly Lys Pro Gly Arg Pro

65

70

75

80

Gly Glu Arg Gly Pro Pro Gly Pro Gln Gly Ala Arg Gly Leu Pro Gly

85

90

95

Thr Ala Gly Leu Pro Gly Met Lys Gly His Arg Gly Phe Ser Gly Leu

100

105

110

Asp Gly Ala Lys Gly Asp Ala Gly Pro Ala Gly Pro Lys Gly Glu Pro

115

120

125

Gly Ser Pro Gly Glu Asn Gly Ala Pro Gly Gln Met Gly Pro Arg Gly

130

135

140

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

145

150

155

160

Arg Gly Asn Asp Gly Ala Thr Gly Ala Ala Gly Pro Pro Gly Pro Thr

165

170

175

Gly Pro Ala Gly Pro Pro Gly Phe Pro Gly Ala Val Gly Ala Lys Gly

180

185

190

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

195

200

205

Arg Gly Glu Pro Gly Pro Pro Gly Pro Ala Gly Ala Ala Gly Pro Ala

210

215

220

Gly Asn Pro Gly Ala Asp Gly Gln Pro Gly Ala Lys Gly Ala Asn Gly

225

230

235

240

Ala Pro Gly Ile Ala Gly Ala Pro Gly Phe Pro Gly Ala Arg Gly Pro

245

250

255

Ser Gly Pro Gln Gly Pro Gly Gly Pro Pro Gly Pro Lys Gly Asn Ser

260

265

270

Gly Glu Pro Gly Ala Pro Gly Ser Lys Gly Asp Thr Gly Ala Lys Gly

275

280

285

Glu Pro Gly Pro Val Gly Val Gln Gly Pro Pro Gly Pro Ala Gly Glu

290

295

300

Glu Gly Lys Arg Gly Ala Arg Gly Glu Pro Gly Pro Thr Gly Leu Pro

305

310

315

320

Gly Pro Pro Gly Glu Arg Gly Gly Pro Gly Ser Arg Gly Phe Pro Gly

325

330

335

Ala Asp Gly Val Ala Gly Pro Lys Gly Pro Ala Gly Glu Arg Gly Ser

340

345

350

Pro Gly Pro Ala Gly Pro Lys Gly Ser Pro Gly Glu Ala Gly Arg Pro

355

360

365

Gly Glu Ala Gly Leu Pro Gly Ala Lys Gly Leu Thr Gly Ser Pro Gly

370

375

380

Ser Pro Gly Pro Asp Gly Lys Thr Gly Pro Pro Gly Pro Ala Gly Gln

385

390

395

400

Asp Gly Arg Pro Gly Pro Pro Gly Pro Pro Gly Ala Arg Gly Gln Ala

405

410

415

Gly Val Met Gly Phe Pro Gly Pro Lys Gly Ala Ala Gly Glu Pro Gly

420

425

430

Lys Ala Gly Glu Arg Gly Val Pro Gly Pro Pro Gly Ala Val Gly Pro

435

440

445

Ala Gly Lys Asp Gly Glu Ala Gly Ala Gln Gly Pro Pro Gly Pro Ala

450

455

460

Gly Pro Ala Gly Glu Arg Gly Glu Gln Gly Pro Ala Gly Ser Pro Gly

465

470

475

480

Phe Gln Gly Leu Pro Gly Pro Ala Gly Pro Pro Gly Glu Ala Gly Lys

485

490

495

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

500

505

510

Gly Ala Arg Gly Glu Arg Gly Phe Pro Gly Glu Arg Gly Val Gln Gly

515

520

525

Pro Pro Gly Pro Ala Gly Pro Arg Gly Ala Asn Gly Ala Pro Gly Asn

530

535

540

Asp Gly Ala Lys Gly Asp Ala Gly Ala Pro Gly Ala Pro Gly Ser Gln

545

550

555

560

Gly Ala Pro Gly Leu Gln Gly Met Pro Gly Glu Arg Gly Ala Ala Gly

565

570

575

Leu Pro Gly Pro Lys Gly Asp Arg Gly Asp Ala Gly Pro Lys Gly Ala

580

585

590

Asp Gly Ser Pro Gly Lys Asp Gly Val Arg Gly Leu Thr Gly Pro Ile

595

600

605

Gly Pro Pro Gly Pro Ala Gly Ala Pro Gly Asp Lys Gly Glu Ser Gly

610

615

620

Pro Ser Gly Pro Ala Gly Pro Thr Gly Ala Arg Gly Ala Pro Gly Asp

625

630

635

640

Arg Gly Glu Pro Gly Pro Pro Gly Pro Ala Gly Phe Ala Gly Pro Pro

645

650

655

Gly Ala Asp Gly Gln Pro Gly Ala Lys Gly Glu Pro Gly Asp Ala Gly

660

665

670

Ala Lys Gly Asp Ala Gly Pro Pro Gly Pro Ala Gly Pro Ala Gly Pro

675

680

685

Pro Gly Pro Ile Gly Asn Val Gly Ala Pro Gly Ala Lys Gly Ala Arg

690

695

700

Gly Ser Ala Gly Pro Pro Gly Ala Thr Gly Phe Pro Gly Ala Ala Gly

705

710

715

720

Arg Val Gly Pro Pro Gly Pro Ser Gly Asn Ala Gly Pro Pro Gly Pro

725

730

735

Pro Gly Pro Ala Gly Lys Glu Gly Gly Lys Gly Pro Arg Gly Glu Thr

740

745

750

Gly Pro Ala Gly Arg Pro Gly Glu Val Gly Pro Pro Gly Pro Pro Gly

755

760

765

Pro Ala Gly Glu Lys Gly Ser Pro Gly Ala Asp Gly Pro Ala Gly A la

770

775

780

Pro Gly Thr Pro Gly Pro Gln Gly Ile Ala Gly Gln Arg Gly Val Val

785

790

795

800

Gly Leu Pro Gly Gln Arg Gly Glu Arg Gly Phe Pro Gly Leu Pro Gly

805

810

815

Pro Ser Gly Glu Pro Gly Lys Gln Gly Pro Ser Gly Ala Ser Gly Glu

820

825

830

Arg Gly Pro Pro Gly Pro Met Gly Pro Pro Gly Leu Ala Gly Pro Pro

835

840

845

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

850

855

860

Arg Asp Gly Ser Pro Gly Ala Lys Gly Asp Arg Gly Glu Thr Gly Pro

865

870

875

880

Ala Gly Pro Pro Gly Ala Pro Gly Ala Pro Gly Ala Pro Gly Pro Val

885

890

895

Gly Pro Ala Gly Lys Ser Gly Asp Arg Gly Glu Thr Gly Pro Ala Gly

900

905

910

Pro Ala Gly Pro Val Gly Pro Ala Gly Ala Arg Gly Pro Ala Gly Pro

915

920

925

Gln Gly Pro Arg Gly Asp Lys Gly Glu Thr Gly Glu Gln Gly Asp Arg

930

935

940

Gly Ile Lys Gly His Arg Gly Phe Ser Gly Leu Gln Gly Pro Pro Gly

945

950

955

960

Pro Pro Gly Ser Pro Gly Glu Gln Gly Pro Ser Gly Ala Ser Gly Pro

965

970

975

Ala Gly Pro Arg Gly Pro Pro Gly Ser Ala Gly Ala Pro Gly Lys Asp

980

985

990

Gly Leu Asn Gly Leu Pro Gly Pro Ile Gly Pro Pro Gly Pro Arg Gly

995

1000

1005

Arg Thr Gly Asp Ala Gly Pro Val Gly Pro Pro Gly Pro Pro Gly Pro

1010

1015

1020

Pro Gly Pro Pro Gly Pro Pro Ser Ala Gly Phe Asp Phe Ser Phe Leu

1025

1030

1035

1040

Pro Gln Pro Pro Gln Glu Lys Ala His Asp Gly Gly Arg Tyr Tyr Arg

1045

1050

1055

Ala Arg Ser Asp Glu Ala Ser Gly Ile Gly Pro Glu Val Pro Asp Asp

1060

1065

1070

Arg Asp Phe Glu Pro Ser Leu Gly Pro Val Cys Pro Phe Arg Cys Gln

1075

1080

1085

Cys His Leu Arg Val Val Gln Cys Ser Asp Leu Gly Leu Asp Lys Val

1090

1095

1100

Pro Lys Asp Leu Pro Pro Asp Thr Thr Leu Leu Asp Leu Gln Asn Asn

1105

1110

1115

1120

Lys Ile Thr Glu Ile Lys Asp Gly Asp Phe Lys Asn Leu Lys Asn Leu

1125

1130

1135

His Ala Leu Ile Leu Val Asn Asn Lys Ile Ser Lys Val Ser Pro Gly

1140

1145

1150

Ala Phe Thr Pro Leu Val Lys Leu Glu Arg Leu Tyr Leu Ser Lys Asn

1155

1160

1165

Gln Leu Lys Glu Leu Pro Glu Lys Met Pro Lys Thr Leu Gln Glu Leu

1170

1175

1180

Arg Ala His Glu Asn Glu Ile Thr Lys Val Arg Lys Val Thr Phe Asn

1185

1190

1195

1200

Gly Leu Asn Gln Met Ile Val Ile Glu Leu Gly Thr Asn Pro Leu Lys

1205

1210

1215

Ser Ser Gly Ile Glu Asn Gly Ala Phe Gln Gly Met Lys Lys Leu Ser

1220

1225

1230

Tyr Ile Arg Ile Ala Asp Thr Asn Ile Thr Ser Ile Pro Gln Gly Leu

1235

1240

1245

Pro Pro Ser Leu Thr Glu Leu His Leu Asp Gly Asn Lys Ile Ser Arg

1250

1255

1260

Val Asp Ala Ala Ser Leu Lys Gly Leu Asn Asn Leu Ala Lys Leu Gly

1265

1270

1275

1280

Leu Ser Phe Asn Ser Ile Ser Ala Val Asp Asn Gly Ser Leu Ala Asn

1285

1290

129 5

Thr Pro His Leu Arg Glu Leu His Leu Asp Asn Asn Lys Leu Thr Arg

1300

1305

1310

Val Pro Gly Gly Leu Ala Glu His Lys Tyr Ile Gln Val Val Tyr Leu

1315

1320

1325

His Asn Asn Asn Ile Ser Val Val Gly Ser Ser Asp Phe Cys Pro Pro

1330

1335

1340

Gly His Asn Thr Lys Lys Ala Ser Tyr Ser Gly Val Ser Leu Phe Ser

1345

1350

1355

1360

Asn Pro Val Gln Tyr Trp Glu Ile Gln Pro Ser Thr Phe Arg Cys Val

1365

1370

1375

Tyr Val Arg Ser Ala Ile Gln Leu Gly Asn Tyr Lys

1380

1385

<210> 11

<211> 1107

<212> PRT

<213> Künstliche Sequenz

<220>

<223> Beschreibung der künstlichen Sequenz:

Aminosäuresequenz eines chimären Proteins für
ein Kollagen-I-(alpha1)/Decorin-Peptid

<400> 11

Gln Leu Ser Tyr Gly Tyr Asp Glu Lys Ser Thr Gly Gly Ile Ser Val

1

5

10

15

Pro Gly Pro Met Gly Pro Ser Gly Pro Arg Gly Leu Pro Gly Pro Pro

20

25

30

Gly Ala Pro Gly Pro Gln Gly Phe Gln Gly Pro Pro Gly Glu Pro Gly

35

40

45

Glu Pro Gly Ala Ser Gly Pro Met Gly Pro Arg Gly Pro Pro Gly Pro

50

55

60

Pro Gly Lys Asn Gly Asp Asp Gly Glu Ala Gly Lys Pro Gly Arg Pro
 65 70 75 80

Gly Glu Arg Gly Pro Pro Gly Pro Gln Gly Ala Arg Gly Leu Pro Gly
 85 90 95

Thr Ala Gly Leu Pro Gly Met Lys Gly His Arg Gly Phe Ser Gly Leu
 100 105 110

Asp Gly Ala Lys Gly Asp Ala Gly Pro Ala Gly Pro Lys Gly Glu Pro
 115 120 125

Gly Ser Pro Gly Glu Asn Gly Ala Pro Gly Gln Met Gly Pro Arg Gly
 130 135 140

Leu Pro Gly Glu Arg Gly Arg Pro Gly Ala Pro Gly Pro Ala Gly Ala
 145 150 155 160

Arg Gly Asn Asp Gly Ala Thr Gly Ala Ala Gly Pro Pro Gly Pro Thr
 165 170 175

Gly Pro Ala Gly Pro Pro Gly Phe Pro Gly Ala Val Gly Ala Lys Gly
 180 185 190

Glu Ala Gly Pro Gln Gly Pro Arg Gly Ser Glu Gly Pro Gln Gly Val
 195 200 205

Arg Gly Glu Pro Gly Pro Pro Gly Pro Ala Gly Ala Ala Gly Pro Ala

210

215

220

Gly Asn Pro Gly Ala Asp Gly Gln Pro Gly Ala Lys Gly Ala Asn Gly

225

230

235

240

Ala Pro Gly Ile Ala Gly Ala Pro Gly Phe Pro Gly Ala Arg Gly Pro

245

250

255

Ser Gly Pro Gln Gly Pro Gly Gly Pro Pro Gly Pro Lys Gly Asn Ser

260

265

270

Gly Glu Pro Gly Ala Pro Gly Ser Lys Gly Asp Thr Gly Ala Lys Gly

275

280

285

Glu Pro Gly Pro Val Gly Val Gln Gly Pro Pro Gly Pro Ala Gly Glu

290

295

300

Glu Gly Lys Arg Gly Ala Arg Gly Glu Pro Gly Pro Thr Gly Leu Pro

305

310

315

320

Gly Pro Pro Gly Glu Arg Gly Gly Pro Gly Ser Arg Gly Phe Pro Gly

325

330

335

Ala Asp Gly Val Ala Gly Pro Lys Gly Pro Ala Gly Glu Arg Gly Ser

340

345

350

Pro Gly Pro Ala Gly Pro Lys Gly Ser Pro Gly Glu Ala Gly Arg Pro

355	360	365	
Gly Glu Ala Gly Leu Pro Gly Ala Lys Gly Leu Thr Gly Ser Pro Gly			
370	375	380	
Ser Pro Gly Pro Asp Gly Lys Thr Gly Pro Pro Gly Pro Ala Gly Gln			
385	390	395	400
Asp Gly Arg Pro Gly Pro Pro Gly Pro Pro Gly Ala Arg Gly Gln Ala			
	405	410	415
Gly Val Met Gly Phe Pro Gly Pro Lys Gly Ala Ala Gly Glu Pro Gly			
	420	425	430
Lys Ala Gly Glu Arg Gly Val Pro Gly Pro Pro Gly Ala Val Gly Pro			
435	440	445	
Ala Gly Lys Asp Gly Glu Ala Gly Ala Gln Gly Pro Pro Gly Pro Ala			
450	455	460	
Gly Pro Ala Gly Glu Arg Gly Glu Gln Gly Pro Ala Gly Ser Pro Gly			
465	470	475	480
Phe Gln Gly Leu Pro Gly Pro Ala Gly Pro Pro Gly Glu Ala Gly Lys			
	485	490	495
Pro Gly Glu Gln Gly Val Pro Gly Asp Leu Gly Ala Pro Gly Pro Ser			
500	505	510	

Gly Ala Arg Gly Glu Arg Gly Phe Pro Gly Glu Arg Gly Val Gln Gly

515

520

525

Pro Pro Gly Pro Ala Gly Pro Arg Gly Ala Asn Gly Ala Pro Gly Asn

530

535

540

Asp Gly Ala Lys Gly Asp Ala Gly Ala Pro Gly Ala Pro Gly Ser G ln

545

550

555

560

Gly Ala Pro Gly Leu Gln Gly Met Pro Gly Glu Arg Gly Ala Ala Gly

565

570

575

Leu Pro Gly Pro Lys Gly Asp Arg Gly Asp Ala Gly P ro Lys Gly Ala

580

585

590

Asp Gly Ser Pro Gly Lys Asp Gly Val Arg Gly Leu Thr Gly Pro Ile

595

600

605

Gly Pro Pro Gly Pro Ala Gly Ala Pro Gly Asp Lys Gly Glu Ser Gly

610

615

620

Pro Ser Gly Pro Ala Gly Pro Thr Gly Ala Arg Gly Ala Pro Gly Asp

625

630

635

640

Arg Gly Glu Pro Gly Pro P ro Gly Pro Ala Gly Phe Ala Gly Pro Pro

645

650

655

Gly Ala Asp Gly Gln Pro Gly Ala Lys Gly Glu Pro Gly Asp Ala Gly

660

665

670

Ala Lys Gly Asp Ala Gly Pro Pro Gly Pro Ala Gly Pro Ala Gly Pro

675

680

685

Pro Gly Pro Ile Gly Asn Val Gly Ala Pro Gly Ala Lys Gly Ala Arg

690

695

700

Gly Ser Ala Gly Pro Pro Gly Ala Thr Gly Phe Pro Gly Ala Ala Gly

705

710

715

720

Arg Val Gly Pro Pro Gly Pro Ser Gly Asn Ala Gly Pro Pro Gly Pro

725

730

735

Pro Gly Pro Ala Gly Lys Glu Gly Gly Lys Gly Pro Arg Gly Glu Thr

740

745

750

Gly Pro Ala Gly Arg Pro Gly Glu Val Gly Pro Pro Gly Pro Pro Gly

755

760

765

Pro Ala Gly Glu Lys Gly Ser Pro Gly Ala Asp Gly Pro Ala Gly Ala

770

775

780

Pro Gly Thr Pro Gly Pro Gln Gly Ile Ala Gly Gln Arg Gly Val Val

785

790

795

800

Gly Leu Pro Gly Gln Arg Gly Glu Arg Gly Phe Pro Gly Leu Pro Gly

805

810

815

Pro Ser Gly Glu Pro Gly Lys Gln Gly Pro Ser Gly Ala Ser Gly Glu

820

825

830

Arg Gly Pro Pro Gly Pro Met Gly Pro Pro Gly Leu Ala Gly Pro Pro

835

840

845

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

850

855

860

Arg Asp Gly Ser Pro Gly Ala Lys Gly Asp Arg Gly Glu Thr Gly Pro

865

870

875

880

Ala Gly Pro Pro Gly Ala Pro Gly Ala Pro Gly Ala Pro Gly Pro Val

885

890

895

Gly Pro Ala Gly Lys Ser Gly Asp Arg Gly Glu Thr Gly Pro Ala Gly

900

905

910

Pro Ala Gly Pro Val Gly Pro Ala Gly Ala Arg Gly Pro Ala Gly Pro

915

920

925

Gln Gly Pro Arg Gly Asp Lys Gly Glu Thr Gly Glu Gln Gly Asp Arg

930

935

940

Gly Ile Lys Gly His Arg Gly Phe Ser Gly Leu Gln Gly Pro Pro Gly

945

950

955

960

Pro Pro Gly Ser Pro Gly Glu Gln Gly Pro Ser Gly Ala Ser Gly Pro

965

970

975

Ala Gly Pro Arg Gly Pro Pro Gly Ser Ala Gly Ala Pro Gly Lys Asp

980

985

990

Gly Leu Asn Gly Leu Pro Gly Pro Ile Gly Pro Pro Gly Pro Arg Gly

995

1000

1005

Arg Thr Gly Asp Ala Gly Pro Val Gly Pro Pro Gly Pro Pro Gly Pro

1010

1015

1020

Pro Gly Pro Pro Gly Pro Pro Ser Ala Gly Phe Asp Phe Ser Phe Leu

1025

1030

1035

1040

Pro Gln Pro Pro Gln Glu Lys Ala His Asp Gly Gly Arg Tyr Tyr Arg

1045

1050

1055

Ala Arg Ser Pro Lys Asp Leu Pro Pro Asp Thr Thr Leu Leu Asp Leu

1060

1065

1070

Gln Asn Asn Lys Ile Thr Glu Ile Lys Asp Gly Asp Phe Lys Asn Leu

1075

1080

1085

Lys Asn Leu His Ala Leu Ile Leu Val Asn Asn Lys Ile Ser Lys Val

1090

1095

110 0

Ser Pro Gly

1105

<210> 12

<211> 4167

<212> DNA

<213> Künstliche Sequenz

<220>

<223> Beschreibung der künstlichen Sequenz:
 DNA-Sequenz eines chimären Proteins für
 ein Kollagen-I-(alpha1)/Decorin-Peptid

<400> 12

cagctgtctt atggctatga tgagaaatca accggaggaa ttcccggtgcc tggcccatg 60
 ggtccctctg gtccctcgtgg tctccctggc cccctcgttg cacctgggtcc ccaaggcttc 120
 caagggtccc ctggtgagcc tggcgagcct ggagcttcag gtcccatggg tccccgaggt 180
 cccccaggtc cccctggaaa gaatggagat gatggggaag ctggaaaacc tggtcgtcct 240
 ggtgagcgtg ggctcctgc gcctcagggt gctcgaggat tgcccggaac agctggcctc 300
 cctggaatga agggacacag aggtttcagt ggtttggatg gtgccaaggg agatgctggt 360
 cctgctggtc ctaagggatga gcctggcagc cctggtgaaa atggagctcc tggtcagatg 420
 ggcccccggtg gcctgcctgg tgagagaggt cgcctggag cccctggccc tgctggtgct 480
 cgtggaaatg atggtgctac tgggtgctgcc gggccccctg gtcccaccgg ccccgctggt 540
 cctcctggct tccctggtgc tgttggtgct aagggtgaag ctgggtccca agggccccga 600

ggctctgaag gtccccaggg tgtgctggt gagcctggcc cccctggccc tgctggtgct 660
 gctggccctg ctggaaaccc tgggtctgat ggacagcctg gtgctaaagg tgccaatggt 720
 gctcctggta ttgctggtgc tcctggcttc cctggtgccc gagggccctc tggaccccag 780
 ggccccggcg gccctcctgg tcccaagggg aacagcgggtg aacctgggtgc tcctggcagc 840
 aaaggagaca ctgggtgctaa gggagagcct ggccctgttg gtgttcaagg accccctggc 900
 cctgctggag agcaaggaaa gcgaggagct cgaggagaac ccggacccac tggcctgccc 960
 ggacccctg gcgagcgtgg tggacctggt agccgtggtt tccctggcgc agatggtggt 1020
 gctggtccca aggggtccgc tgggtgaacgt ggttctcctg gcccgcctgg ccccaaagga 1080
 tctcctcgtg aagctgggtc tcccgggtgaa gctggtctgc ctggtgcaa ggggtctgact 1140
 ggaagccctg gcagccctgg tcctgatggc aaaactggcc cccctggctc cgccggtcaa 1200
 gatggtcgcc ccggaccccc agggccacct ggtgcccgtg gtcaggctgg tgtgatggga 1260
 ttccctggac ctaaagggtgc tgctcgagag cccggcaagg ctggagagcg aggtgttccc 1320
 ggacccctc gcgctgtcgg tcctgctggc aaagatggag aggctggagc tcagggaccc 1380
 cctggccctg ctgggtccgc tggcgagaga ggtgaacaag gccctgctgg ctccccgga 1440
 ttccagggtc tccctggtcc tgctggtcct ccagggtgaag caggcaaacc tgggtgaacag 1500
 ggtgttctcgt gagaccttgg cggccctggc ccctctggag caagaggcga gagagggttc 1560
 cctggcgagc gtggtgtgca aggtccccc ggtcctgctg gaacccgagg ggccaacggt 1 620
 gctcccgcca acgatgctgc taagggtgat gctggtgccc ctggagctcc cggtagccag 1680
 ggcgccccctg gccttcaggg aatgcctggt gaacgtggtg cagctggtct tccagggcct 1740
 aagggtgaca gaggtgatgc tgggtccaaa ggtgctgatg gctctcctgg caaagatggc 1800
 gtccgtggtc tgaccgaccc cattggtcct cctggccctg ctggtgcccc tggtgacaag 1860
 ggtgaaagtg gtcccagcgg ccctgctggt cccactggag ctggtggtgc cccggagac 1920
 cgtggtgagc ctggtcccc cggccctgct ggctttgctg gccccctgg tgctgacggc 1980
 caacctggtg ctaaaggcga acctggtgat gctggtgcca aaggcgatgc tgggtcc ccct 2040
 gggcctgccc gacccgctgg accccctggc cccattggta atgttgggtgc tcctggagcc 2100
 aaacgtgctc gcggcagcgc tgggtccccc ggtgctactg gtttccctgg tgctgctggc 2160
 cgagtgggtc ctccctggccc ctctggaaat gctggacccc ctggccctcc tggctcctgct 2220
 ggcaaagaag gcggcaaagg tccccgtggt gagactggcc ctgctggacg tcctggtgaa 2280

gttggtcccc ctggtccccc tggccctgct ggcgagaaag gatcccctgg tgctgatggt 2340
 cctgctggtg ctccctggtac tcccgggcct caaggtattg ctggacagcg tgggtgtggtc 2400
 ggccctgcctg gtcagagagg agagagaggc ttccctgggtc ttcttggccc ctctggtgaa 2460
 cctggcaaac aaggtccctc tggagcaagt ggtgaacgtg gtccccccgg tcccatgggc 2520
 cccctgggat tggctggacc ccctggtgaa tctggacgtg agggggctcc tgctgccgaa 2580
 ggttccccctg gacgagacgg ttctcctggc gccaaagggtg accgtgggtg gaccggcccc 2640
 gctggacccc ctggtgctcc tgggtgctcct ggtgcccctg gccccttgg ccctgctggc 2700
 aagagtgggtg atcgtgggtg gactggctcct gctgggtcccg ccggtcccggt cggccccgct 2760
 ggcgcccgtg gccccgccgg accccaaggc ccccggtgtg acaagggtga gacaggcgaa 2820
 caggcgacaa gaggcataaa ggggtcacgt ggcttctctg gcctc caggg tccccctggc 2880
 cctcctgggt ctccctggtga acaagggtccc tctggagcct ctgggtcctgc tgggtccccga 2940
 ggtccccctg gctctgctgg tgctcctggc aaagatggac tcaacgggtc ccctggcccc 3000
 attgggcccc ctggtcctcg cggctgcact ggtgatgctg gtccctgttg tccccccggc 3060
 cctcctggac ctccctggtc ccctgggtcct ccagcgctg gtttcgactt cagcttcctc 3120
 cccagccac ctcaagagaa gggtcacgat ggtggccgct actaccgggc tagatccgat 3180
 gaggtctctg ggatagcccc agaagttcct gatgaccgag acttcgagcc ctccctaggc 3240
 ccagtgtgct ccttcgctg tcaatgccat cttcgagtgg tccagtgttc tgatttgggt 3300
 ctggacaaaag tgccaaagga tcttccccct gacacaactc tgctagacct gcaaaacaac 3360
 aaaataaccg aaatcaaaga tggagacttt aagaacctga agaaccttca cgcattgatt 3420
 cttgtcaaca ataaaattag caaagttagt cctggagcat ttacaccttt ggtgaagttg 3480
 gaacgacttt atctgtccaa gaatcagctg aaggaattgc cagaaaaaat gcccaaaact 3540
 cttcaggagc tgcgtgcccc tgagaatgag atcaccaaag tgcgaaaagt tactttcaat 3600
 ggactgaacc agatgattgt catagaactg ggcaccaatc cgctgaagag ctcaggaatt 3660
 gaaaatgggg ctttccaggg aatgaagaag ctct cctaca tccgcattgc tgataccaat 3720
 atcaccagca ttctcaagg tcttctcctc tcccttacgg aattacatct tgatggcaac 3780
 aaaatcagca gagttgatgc agctagcctg aaaggactga ataatttggc taagttggga 3840
 ttgagtttca acagcatctc tgctgttgac aatggctctc tggccaacac gcctcatctg 3900


```

aggagcttc acttgacaa caacaagctt accagagtac ctggtgggct ggcagagcat 3960
aagtacatcc aggttgtcta ccttcataac aacaatatct ctgtagttgg atcaagtac 4020
ttctgccac ctggacacaa caccaaaaag gcttcttatt cgggtgtgag tcttttcagc 4080
aaccgggtcc agtactggga gatacagcc a tccaccttca gatgtgtcta cgtgcgctct 4140
gccattcaac tgggaaacta taagtaa 4167

```

<210> 13

<211> 3349

<212> DNA

<213> Künstliche Sequenz

<220>

<223> Beschreibung der künstlichen Sequenz:
DNA-Sequenz codierend ein chimäres Kollagen/
Decorin-Protein

<400> 13

```

gggaaggatt tccatttccc agctgtctta tggctatgat gagaaatcaa ccggaggaat 60
ttccgtgcct ggcccatgg gtccctctgg tctcgtgggt ctccctggcc cccctgggtgc 120
acctgggtccc caaggcttcc aagggtcccc tgggtgagcct ggcgagcctg gagcttcagg 180
tcccatgggt ccccgaggtc cccaggtcc ccctggaaag aatggagatg atggggaagc 240
tggaacacct ggtcgtcctg gtgagcgtgg gcctcctggg cctcaggggtg ctcgaggatt 300
gcccggaaca gctggcctcc ctggaatgaa gggacacaga ggtttcagtg gtttggtatgg 360
tgccaaggga gatgctggtc ctgctgggtcc taagggtgag cctggcagcc ctggtgaaaa 420
tgagctcct ggtcagatgg gccccctgg cctgcctgggt gagagaggtc gccctggagc 480
ccctggccct gctgggtgct gtggaaatga tgggtgtact ggtgctgccg ggccccctgg 540
tcccaccggc ccgctggtc ctctggctt ccctgggtgct gttggtgcta aggggtgaagc 600
tggtcccaa gggccccgag gctctgaagg tccccagggt gtgcgtgggt agcctggccc 660

```

ccctggccct gctggtgctg ctggccctgc tggaaaccct ggtgctgatg gacagcctgg 720
 tgctaaaggt gccaatggtg ctccctggtat tgctggtgct cctggcttcc ctggtgcccg 780
 agggccctct ggaccccagg gccccggcgg ccctcctggt cccaagggtg acagcgggtg 840
 acctggtgct cctggcagca aaggagacac tggtgctaag ggagagcctg gccctggttg 900
 tgttcaagga cccctggcc ctgctggaga ggaaggaaag cgaggagctc gaggtgaacc 960
 cggaccact ggcctgcccg gacccctgg cgagcgtggt ggacctggtg gccgtggttt 1020
 ccctggcgca gatggtgttg ctggtcccaa ggtcccgt ggtgaacgtg gttctcctgg 1080
 ccccgtggc ccaaaaggat ctccctggtg agctggtcgt ccggtgaag ctggtctgcc 1140
 tggtgccaag ggtctgactg gaagccctgg cagccctggt cctgatggca aaactggccc 1200
 ccctggtccc gccggtcaag atggtcgccc cggaccccca ggcccacctg gtgcccgtgg 1260
 tcaggctggt gtgatgggat tccctggacc taaagggtgct gctggagagc ccggcaaggc 1320
 tggagagcga ggtgttcccg gacccctgg cgctgtcgt cctgctggca aagatggaga 1380
 ggctggagct cagggacccc ctggccctgc tgggtcccgt ggcgagagag gtgaacaagg 1440
 ccctgctggc tccccggat tccagggtct ccctggtcct gctggtcctc caggtgaagc 1500
 aggcaaacct ggtgaacagg gtgttcctgg agaccttggc gccctggcc cctctggagc 1560
 aagaggcgag agaggtttcc ctggcgagcg tgggtgtgca ggtcccctg gtccctgctgg 1620
 accccgaggg gccaacggtg ctcccggcaa cgatggtgct aagggtgatg ctggtgcccc 1680
 tggagctccc ggtagccagg gcgcccctgg ccttcaggga atgcctggtg aacgtggtgc 1740
 agctggtctt ccagggccta aggggtgacag aggtgatgct ggtcccaaag gtgctgatgg 1800
 ctctcctggc aaagatggcg tccgtggtct gaccggcccc attggtcctc ctggccctgc 1860
 tggtgcccct ggtgacaagg gtgaaagtgg tcccagcggc cctgctggtc ccactggagc 1920
 tcgtggtgcc ccgggagacc gtggtgagcc tgggtcccc ggccctgctg gctttgctgg 1980
 cccccctggt gctgacggcc aacctggtgc taaaggcgaa cctggtgatg ctggtgccaa 2040
 aggcgatgct ggtcccctg ggcctgccgg acccgctgga cccctggcc ccattggtaa 2100
 tgttggtgct cctggagcca aagggtgctc cggcagcgt ggtcccctg gtgctactgg 2160
 tttccctggt gctgctggcc gagtcggtcc tcctggcccc tctggaaatg ctggaccccc 2220
 tggccctcct ggtcctgctg gcaaagaagg cggcaaagggt cccctggtg agactggccc 2280

```

tgctggacgt cctggtgaag ttggtccccc tgggtcccct ggccctgctg gcgagaaagg 2340
atcccttggg gctgatgggc ctgctgggtgc tcctgggtact cccgggcctc aaggtattgc 2400
tggacagcgt ggtgtgggtcg gcctgcctgg tcagagagga gagagaggct tccttgggtct 2460
tcctggcccc tctggtgaac ctggcaaaca aggtccctct ggagcaagtg gtgaacgtgg 2520
tccccccggg cccatggggc cccctggatt ggctggacct cctggtgaat ctggacgtga 2580
ggggggtcct gctgccgaag gttcccttgg acgagacggg tctcctggcg ccaaggggtga 2640
ccgtggtgag accggccccg ctggaccccc tgggtgctcct ggtgctcctg gtgccccctg 2700
ccccgttggc cctgctggca agagtgggtga togtgggtgag actggtcctg ctggtcccg 2760
cgggtcccgtc ggccccgctg gggcccggtg ccccgccgga cccaaggcc cccgtgggtga 2820
caaggggtgag acaggcgaac agggcgacag aggcataaag ggtcacctg gcttctcttg 2880
cctccagggt ccccttggcc ctcttggctc tcctgggtgaa caaggctcct ctggagcctc 2940
tggctcctgct ggtccccgag gtccccctgg ctctgctggt gctcctggca aagatggact 3000
caacgggtctc cctggcccca ttggggcccc tgggtcctgc ggtcgcaactg gtgatgctgg 3060
tcctgttggg cccccgggct ctcttggacc tcctgggtccc cctgggtcctc ccagcgctgg 3120
tttcgacttc agcttcctcc ccagccacc tcaagagaag gctcacgatg gtggccgcta 3180
ctaccgggt agatctccaa aggatcttcc cctgacaca actctgctag acctgcaaaa 3240
caacaaaata accgaaatca aagatggaga cttaagaac ctgaagaacc ttcacgcatt 3300
gattcttgtc aacaataaaa ttagcaaagt tagtcctgga taactgcag 3349

```

<210> 14

<211> 57

<212> DNA

<213> Künstliche Sequenz

<220>

<223> Beschreibung der künstlichen Sequenz:

DNA-Sequenz einer Polylinker-Klonierungsstelle,
die in dem pMaI-Klonierungsvektor enthalten ist.

<400> 14

atcgagggaa ggatttcaga attcggatcc tctagagtcg acctgcaggc aagcttg 57

<210> 15

<211> 3171

<212> DNA

<213> Homo sapiens

<400> 15

cagctgtctt atggctatga tgagaaatca accggaggaa ttcccgtagc tggcccatg 60
 ggtccctctg gtccctcgtg tctccctggc cccctcgtg cacctggtec ccaaggcttc 120
 caaggteccc ctggtgagcc tggcgagcct ggagcttcag gtcccatggg tccccgaggt 180
 cccccaggtc cccctggaaa gaatggagat gatggggaag ctggaaaacc tggtcgtcct 240
 ggtgagcgtg ggcctcctgg gcctcagggt gtcgaggat tgcccggaac agctggcctc 300
 cctggaatga agggacacag aggtttcagt ggtttggatg gtgccaaggg agatgctggt 360
 cctgctggtc ctaagggtga gcctggcagc cctggtgaaa atggagctcc tggtcagatg 420
 ggcccccgtag gcctgcctgg tgagagaggt cggcctggag cccctggccc tgctggtgct 480
 cgtggaaatg atgggtgctac tgggtgctgcc gggccccctg gtcccaccgg ccccgctggt 540
 cctcctggct tcctcgtgct tgttggtgct aagggtgaag ctggtcccca agggccccga 600
 ggctctgaag gtccccaggg tgtgcgtggt gagcctggcc cccctggccc tgctggtgct 660
 gctggccctg ctggaaaccc tgggtgctgat ggacagcctg gtgctaaagg tgccaatggt 720
 gtcctcgtga ttgctggtgc tcctggcttc cctggtgccc gagggccctc tggacccag 780
 ggccccggcg gccctcctgg tcccaagggt aacagcgggt aacctggtgc tcctggcagc 840
 aaaggagaca ctggtgctaa gggagagcct ggccctggtg gtgttcaagg accccctggc 900

cctgctggag aggaaggaaa gcgaggagct cgaggtgaac ccggacccac tggcctgccc 960
 ggacccctg gcgagcgtgg tggacctggt agccgtggtt tccctggcgc agatggtggt 1020
 gctggtccca aggggtccgc tggatgaacgt ggttctcctg gcccgcctgg ccccaaagga 1080
 tctcctggtg aagctggtcg tcccgggtgaa gctggtctgc ctggtgcca gggctctgact 1140
 ggaagccctg gcagccctgg tctgatggc aaaactggcc cccctggtcc cgccgggtcaa 1200
 gatggtcgcc ccggaccccc agggccacct ggtgcccgtg gtcaggctgg tgtgatggga 1260
 ttccctggac ctaaagggtg tgctggagag cccggcaagg ctggagagcg aggtgttccc 1320
 ggacccctg gcgctgtcgg tctgctggc aaagatggag aggctggagc tcagggaccc 1380
 cctggccctg ctggtccgc tggcgagaga ggtgaacaag gccctgctgg ctcccccgga 1440
 ttccagggtc tccctggtcc tgctggtcct ccaggtgaag caggcaaacc tggatgaaca g 1500
 ggtgttctg gagacctg gcgccctggc ccctctggag caagaggcga gagaggtttc 1560
 cctggcgagc gtggtgtgca aggtccccct ggtcctgctg gaccccgagg ggccaacggt 1620
 gctcccggca acgatggtg taagggtgat gctggtgccc ctggagctcc cggtagccag 1680
 ggcgccctg gccttcaggg aatgcctggt gaacgtggtg cagctggtct tccagggcct 1740
 aagggtgaca gaggtgatg tggccccaa ggtgctgatg gctctcctgg caaagatggc 1800
 gtccgtggtc tgaccggccc cattggtcct cctggccctg ctggtgcccc tggatgacaag 1860
 ggtgaaagtg gtcccagcgg ccctgctggt cccactggag ctctggtgct ccc cggagac 1920
 cgtggtgagc ctggtcccc cggccctgct ggctttgctg gccccctgg tgctgacggc 1980
 caacctggtg ctaaaggcga acctggtgat gctggtgcca aaggcgatg tggteccccct 2040
 gggcctgccc gacccgctgg accccctggc ccattggtg atgttggtg tcttggagcc 2100
 aaagggtgctc gcggcagcgc tgggtccccct ggtgctactg gtttccctgg tgctgctggc 2160
 cgagtcggtc ctctggccc ctctggaaat gctggacccc ctggccctcc tggctcctgct 2220
 ggcaaagaag gcggcaaagg tccccgtggt gagactggcc ctgctggacg tccctggtgaa 2280
 gttggtcccc ctggtcccc tggccctgct ggcgagaaag gatccccct gg tgctgatggt 2340
 cctgctggtg ctctggtac tcccgggcct caaggattg ctggacagcg tgggtgtggtc 2400
 ggcctgcctg gtcagagagg agagagaggc ttccctggtc ttctggccc ctctggtgaa 2460
 cctggcaaac aaggctccctc tgagcaagt ggtgaacgtg gtccccccgg tcccatgggc 2520
 cccctggat tggctggacc ccctggtgaa tctggacgtg agggggctcc tgctgccgaa 2580

gggtccctg gacgagacgg ttctcctggc gccaaagggg accgtgggtga gaccggcccc 2640
 gctggacccc ctggtgctcc tgggtgctcct ggtgcccctg gcccgttg cctgctggc 2700
 aagagtgggtg atcgtgggtga gactggctcct gctgggtcccg cc ggtcccgt cggccccgct 2760
 ggcgcccctg gccccgccgg accccaaggg ccccggtggg acaaggggtga gacaggcgaa 2820
 cagggcgaca gaggcataaa gggtcaccgt ggcttctctg gcctccaggg tccccctggc 2880
 cctcctggct ctctggtga acaagggtccc tctggagcct ctggctcctgc tgggtcccgga 2940
 ggtccccctg gctctgctgg tgcctcctggc aaagatggac tcaacggtct cctggcccc 3000
 attgggcccc ctgggtcctcg cggtcgcact ggtgatgctg gtcctggttg tccccccggc 3060
 cctcctggac ctctggtcc cctgggtcct cccagcgctg gtttcgactt cagcttcctc 3120
 cccagccac ctcaagagaa gggtcacgat ggtggcc gct actaccgggc t 3171

<210> 16

<211> 1057

<212> PRT

<213> Homo sapiens

<400> 16

Gln Leu Ser Tyr Gly Tyr Asp Glu Lys Ser Thr Gly Gly Ile Ser Val

1

5

10

15

Pro Gly Pro Met Gly Pro Ser Gly Pro Arg Gly Leu Pro Gly Pro Pro

20

25

30

Gly Ala Pro Gly Pro Gln Gly Phe Gln Gly Pro Pro Gly Glu Pro Gly

35

40

45

Glu Pro Gly Ala Ser Gly Pro Met Gly Pro Arg Gly Pro Pro Gly Pro

50

55

60

Pro Gly Lys Asn Gly Asp Asp Gly Glu Ala Gly Lys Pro Gly Arg Pro

65

70

75

80

Gly Glu Arg Gly Pro Pro Gly Pro Gln Gly Ala Arg Gly Leu Pro Gly

85

90

95

Thr Ala Gly Leu Pro Gly Met Lys Gly His Arg Gly Phe Ser Gly Leu

100

105

110

Asp Gly Ala Lys Gly Asp Ala Gly Pro Ala Gly Pro Lys Gly Glu Pro

115

120

125

Gly Ser Pro Gly Glu Asn Gly Ala Pro Gly Gln Met Gly Pro Arg Gly

130

135

140

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

145

150

155

160

Arg Gly Asn Asp Gly Ala Thr Gly Ala Ala Gly Pro Pro Gly Pro Thr

165

170

175

Gly Pro Ala Gly Pro Pro Gly Phe Pro Gly Ala Val Gly Ala Lys Gly

180

185

190

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

195

200

205

Arg Gly Glu Pro Gly Pro Pro Gly Pro Ala Gly Ala Ala Gly Pro Ala

210

215

220

Gly Asn Pro Gly Ala Asp Gly Gln Pro Gly Ala Lys Gly Ala Asn Gly

225

230

235

240

Ala Pro Gly Ile Ala Gly Ala Pro Gly Phe Pro Gly Ala Arg Gly Pro

245

250

255

Ser Gly Pro Gln Gly Pro Gly Gly Pro Pro Gly Pro Lys Gly Asn Ser

260

265

270

Gly Glu Pro Gly Ala Pro Gly Ser Lys Gly Asp Thr Gly Ala Lys Gly

275

280

285

Glu Pro Gly Pro Val Gly Val Gln Gly Pro Pro Gly Pro Ala Gly Glu

290

295

300

Glu Gly Lys Arg Gly Ala Arg Gly Glu Pro Gly Pro Thr Gly Leu Pro

305

310

315

320

Gly Pro Pro Gly Glu Arg Gly Gly Pro Gly Ser Arg Gly Phe Pro Gly

325

330

335

Ala Asp Gly Val Ala Gly Pro Lys Gly Pro Ala Gly Glu Arg Gly Ser

340

345

350

Pro Gly Pro Ala Gly Pro Lys Gly Ser Pro Gly Glu Ala Gly Arg Pro

355

360

365

Gly Glu Ala Gly Leu Pro Gly Ala Lys Gly Leu Thr Gly Ser Pro Gly

370

375

380

Ser Pro Gly Pro Asp Gly Lys Thr Gly Pro Pro Gly Pro Ala Gly Gln

385

390

395

400

Asp Gly Arg Pro Gly Pro Pro Gly Pro Pro Gly Ala Arg Gly Gln Ala

405

410

415

Gly Val Met Gly Phe Pro Gly Pro Lys Gly Ala Ala Gly Glu Pro Gly

420

425

430

Lys Ala Gly Glu Arg Gly Val Pro Gly Pro Pro Gly Ala Val Gly Pro

435

440

445

Ala Gly Lys Asp Gly Glu Ala Gly Ala Gln Gly Pro Pro Gly Pro Ala

450

455

460

Gly Pro Ala Gly Glu Arg Gly Glu Gln Gly Pro Ala Gly Ser Pro Gly

465

470

475

480

Phe Gln Gly Leu Pro Gly Pro Ala Gly Pro Pro Gly Glu Ala Gly Lys

485

490

495

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

500

505

510

Gly Ala Arg Gly Glu Arg Gly Phe Pro Gly Glu Arg Gly Val Gln Gly

515

520

525

Pro Pro Gly Pro Ala Gly Pro Arg Gly Ala Asn Gly Ala Pro Gly Asn

530

535

540

Asp Gly Ala Lys Gly Asp Ala Gly Ala Pro Gly Ala Pro Gly Ser Gln

545

550

555

560

Gly Ala Pro Gly Leu Gln Gly Met Pro Gly Glu Arg Gly Ala Ala Gly

565

570

575

Leu Pro Gly Pro Lys Gly Asp Arg Gly Asp Ala Gly Pro Lys Gly Ala

580

585

590

Asp Gly Ser Pro Gly Lys Asp Gly Val Arg Gly Leu Thr Gly Pro Ile

595

600

605

Gly Pro Pro Gly Pro Ala Gly Ala Pro Gly Asp Lys Gly Glu Ser Gly

610

615

620

Pro Ser Gly Pro Ala Gly Pro Thr Gly Ala Arg Gly Ala Pro Gly Asp

625

630

635

640

Arg Gly Glu Pro Gly Pro Pro Gly Pro Ala Gly Phe Ala Gly Pro Pro

645

650

655

Gly Ala Asp Gly Gln Pro Gly Ala Lys Gly Glu Pro Gly Asp Ala Gly

660

665

670

Ala Lys Gly Asp Ala Gly Pro Pro Gly Pro Ala Gly Pro Ala Gly Pro

675

680

685

Pro Gly Pro Ile Gly Asn Val Gly Ala Pro Gly Ala Lys Gly Ala Arg

690

695

700

Gly Ser Ala Gly Pro Pro Gly Ala Thr Gly Phe Pro Gly Ala Ala Gly

705

710

715

720

Arg Val Gly Pro Pro Gly Pro Ser Gly Asn Ala Gly Pro Pro Gly Pro

725

730

735

Pro Gly Pro Ala Gly Lys Glu Gly Gly Lys Gly Pro Arg Gly Glu Thr

740

745

750

Gly Pro Ala Gly Arg Pro Gly Glu Val Gly Pro Pro Gly Pro Pro Gly

755

760

765

Pro Ala Gly Glu Lys Gly Ser Pro Gly Ala Asp Gly Pro Ala Gly Ala

770

775

780

Pro Gly Thr Pro Gly Pro Gln Gly Ile Ala Gly Gln Arg Gly Val Val

785

790

795

800

Gly Leu Pro Gly Gln Arg Gly Glu Arg Gly Phe Pro Gly Leu Pro Gly

805

810

815

Pro Ser Gly Glu Pro Gly Lys Gln Gly Pro Ser Gly Ala Ser Gly Glu

820

825

830

Arg Gly Pro Pro Gly Pro Met Gly Pro Pro Gly Leu Ala Gly Pro Pro

835

840

845

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

850

855

860

Arg Asp Gly Ser Pro Gly Ala Lys Gly Asp Arg Gly Glu Thr Gly Pro

865

870

875

880

Ala Gly Pro Pro Gly Ala Pro Gly Ala Pro Gly Ala Pro Gly Pro Val

885

890

895

Gly Pro Ala Gly Lys Ser Gly Asp Arg Gly Glu Thr Gly Pro Ala Gly

900

905

910

Pro Ala Gly Pro Val Gly Pro Ala Gly Ala Arg Gly Pro Ala Gly Pro

915

920

925

Gln Gly Pro Arg Gly Asp Lys Gly Glu Thr Gly Glu Gln Gly Asp A rg

930

935

940

Gly Ile Lys Gly His Arg Gly Phe Ser Gly Leu Gln Gly Pro Pro Gly

945

950

955

960

Pro Pro Gly Ser Pro Gly Glu Gln Gly Pro Ser Gly A la Ser Gly Pro

965

970

975

Ala Gly Pro Arg Gly Pro Pro Gly Ser Ala Gly Ala Pro Gly Lys Asp

980

985

990

Gly Leu Asn Gly Leu Pro Gly Pro Ile G ly Pro Pro Gly Pro Arg Gly

995

1000

1005

Arg Thr Gly Asp Ala Gly Pro Val Gly Pro Pro Gly Pro Pro Gly Pro

1010

1015

1020

Pro Gly Pro Pro Gly Pro Pro Ser Ala Gly Phe Asp Phe Ser Phe Leu

1025

1030

1035

1040

Pro Gln Pro Pro Gln Glu Lys Ala His Asp Gly Gly Arg Tyr Tyr Arg

1045

1050

1055

Ala

<210> 17
 <211> 46
 <212> PRT
 <213> Künstliche Sequenz

<220>

<223> Beschreibung der künstlichen Sequenz:
 Aminosäuresequenz eines chimären GST-ColECol-Proteins

<220> Xaa an Position 1 ist die Aminosäuresequenz für
 Glutathion-S-Transferase

<220>

<223> Xaa an Position 20 sind 338 Wiederholungen der
 folgenden Tripletts Gyl-X-Y, worin etwa 35 % der
 X- und Y-Positionen bei Prolin und 4-Hydroxyprolin
 enthalten sind.

<400> 17

Xaa Met Gln Leu Ser Tyr Gly Tyr Asp Glu Lys Ser Thr Gly Gly Ile

1

5

10

15

Ser Val Pro Xaa Ser Ala Gly Phe Asp Phe Ser Phe Leu Pro Gln Pro

20

25

30

Pro Gln Glu Lys Ala His Asp Gly Gly Arg Tyr Tyr Arg Ala

35

40

45

<210> 18

<211> 31

<212> PRT

<213> Künstliche Sequenz

<220>

<223> Xaa an Position 1 ist die Aminosäuresequenz für
Glutathion-S-Transferase

<220>

<223> Xaa an Position 4 sind 338 Wiederholungen der
folgenden Tripletts Gyl-X-Y, worin etwa 35 % der
X- und Y-Positionen bei Prolin und 4-Hydroxyprolin
enthalten sind.

<220>

<223> Beschreibung der künstlichen Sequenz:
Aminosäuresequenz des chimären Proteins GST-D4

<400> 18

Xaa Met Gly Xaa Tyr Ser Ala Gly Phe Asp Phe Ser Phe Leu Pro Gln

1

5

10

15

Pro Pro Gln Glu Lys Ala His Asp Gly Gly Arg Tyr Tyr Arg Ala

20

25

30

<210> 19
 <211> 3171
 <212> DNA
 <213> Künstliche Sequenz

<220>

<223> Beschreibung der künstlichen Sequenz:
 Humanes Protein vom Kollagen-I-alpha-Typ
 mit optimierter E. coli-Codonverwendung

<400> 19

```

cagctgagct atggctatga tgaaaaaagc accggcgggca tcagcgtgcc gggcccgatg 60
ggtcgagcgg gccctcgtgg cctgcccgggc ccgccagggtg cggccgggtcc gcagggccttt 120
caggggtccgc cgggcgaacc gggcgaaacct ggtgcgagcgg gcccgatggg cccgcgcggc 180
ccgccgggtc cgccaggcaa aaacggcgat gatggcggaag cgggcaaacc gggacgtccg 240
ggtgaacgtg gccccccggg cccgcagggc gcgcgcggac tgccgggtac tgccgggactg 300
ccgggcatga aaggccaccg cggtttctct ggtctggatg gtgcgaaagg tga tgccgggt 360
ccggcgggtc cgaaagggtga gccgggcagc ccgggcgaaa acggcgcgcc gggtcagatg 420
ggcccgcggtg gcctgcctgg tgaacgcggt cggccggggc ccccgggccc agctggcgca 480
cgtggcaacg atggtgcgac cggtgccggc ggtccaccgg gcccgacggg cccggcgggt 540
cccccgggct ttccgggtgc ggtgggtgcg aaaggcggaag caggtccgca ggggcgcgcg 600
gggagcgagg gtcctcaggg cgttcgtggt gaaccgggccc cggcggggccc ggcgggtgcg 660
gcgggcccgg ctggtaaccc tggcgcggaac ggtcagccag gtgcgaaagg tgccaacggc 720
gcgccgggta ttgcaggtgc accgggcttc ccgggtgccc gcggccccg tc cggcccgag 780
ggcccgggcg gcccgcccgg cccgaaaggg aacagcggtg aaccgggtgc gccaggcagc 840

```


aaaggcgaca ccggtgcgaa aggtgaaccg ggcccagtgg gtgttcaagg cccgccccgc 900
 ccggcgggcg aggaaggcaa acgcggtgct cgcggtgaac cgggcccgac cggcctgcct 960
 ggcccccggg gagaacgtgg tggccccgggt agccgcggtt ttccggggcg ggatggtgtg 1020
 gggggccccga aaggtccggc ggggtgaacgt ggtagccccg gcccgggcggg cccaaaaggc 1080
 agccccggcg aggcaggacg tccgggtgaa gcggggtctcc cgggcgcaa aggtctgacc 1140
 ggctctccgg gcagccccgg tccggatggc aaaacgggccc cg cctggtcc ggccggccag 1200
 gatggtcgcc cgggcccgcc gggccccgcg ggtgccccgt gtcaggcggg tgtcatgggc 1260
 tttccaggcc ccaaagggtg ggggggtgaa ccgggcaaag cgggcgaacg cgggtgtccg 1320
 ggtccgcccg gcgctgtcgg gccggcgggc aaagatggcg aagcgggcgc gcaaggcccc 1380
 ccgggaccag cgggtccggc gggcgagcgc ggtgaacagg gcccggcagg cagcccggggt 1440
 ttccagggtc tgccggggccc tgcgggtcca ccgggtgaag cgggcaaacc gggggaacaa 1500
 ggtgtgccgg gcgacctggg cggcccaggc ccgagcggcg cgcgcggcga acgcggtttc 1560
 ccgggcgaac gtggtgtgca gggccccgcc ggccccg ctg gtccgcgcgg cgccaacggc 1620
 gcgccccgca acgatggtgc gaaaggatgat gggggtgccc caggtgcgcc gggcagccag 1680
 ggcgcccccg ggctgcaagg catgcgggt gaacgtggtg ccgcgggtct accgggtccg 1740
 aaaggcgacc gcggtgatgc gggtcacaaa ggtgcggatg gctccccctg caaagatggc 1800
 gttcgtggtc tgaccggccc gatcggcccc ccgggccccg caggtgcccc gggtgacaaa 1860
 ggtgaaagcg gtccgagcgg ccagcgggc cccactggtg cgcgtggtgc cccgggcgac 1920
 cgtggtgaac cgggtccgcc gggccccggc ggctttgcgg gcccgccagg cgctgacggc 1980
 cagccgggtg cgaaaggcga accgggggat g cgggtgcta aaggcgacgc gggtcgccg 2040
 ggccctgccg gcccggcggg cccgccaggc ccgattggca acgtgggtgc gccgggtgcc 2100
 aaagggtgcgc gcggcagcgc tgggtccgcc ggcgcgaccg gtttcccccg tgcggcgggg 2160
 cgcgtgggtc cgccaggccc gagcggtaac ggggtccgc caggtccgcc tggccccgct 2220
 ggcaaagagg gcggcaagg tccgcgtggt gaaaccggcc ctgcgggacg tccaggtgaa 2280
 gtgggtccgc cgggccccgc gggccccggc ggcgaaaaag gtagccccgg tgcggatggt 2340
 cccgccggtg cgccaggcac gccgggtccg caaggtatcg ctggccagcg tgggtgtcgtc 2400
 gggctgccgg gtcagcgcgg cgaacg cggc ttcccggtc tgccggggccc gagcgggtgag 2460
 ccgggcaaac agggtcctc tggcgcgagc ggtgaacgtg gcccgccggg tcccatgggc 2520

```

ccgccgggtc tggcgggccc tccgggtgaa agcggtcgtg aaggcgcgcc gggtgccgaa 2580
ggcagcccag gccgcgacgg tagcccgggg gccaaagggg atcgtggtga aaccggcccg 2640
gcggggcccc cgggtgcacc gggcgcgccg ggtgccccag gcccggtggg cccggcgggc 2700
aaaagcgggtg atcgtggtga gaccgggtccg gcgggcccgg ccggtccggt gggcccagcg 2760
ggcgcccgtg gcccggccgg tccgcagggc ccgcgggggtg acaaaggtga aacggggcga 2820
cagggcgacc gtggcattaa aggccaccgt ggcttcagcg gcctgcaggg tccaccgggc 2880
ccgcggggca gtccgggtga acaggggtccg tccggagcca gcggggccggc gggcccacgc 2940
ggtccgcccgg gcagcgcggg cgcgccgggc aaagacggtc tgaacgggtct gccgggccccg 3000
atcgggccgc cgggcccacg cggccgcacc ggtgatgcgg gtccgggtggg tccccgggc 3060
ccgcggggcc cggcaggccc gccgggaccg ccgagcgcggt gtttcgactt cagcttcctg 3120
ccgcagccgc cgcaggagaa agcgcacgac ggcggtcgct actaccgtgc g 3171

```

<210> 20

<211> 1057

<212> PRT

<213> Künstliche Sequenz

<220>

<223> Beschreibung der künstlichen Sequenz:

Menschliches Protein vom Kollagen-I-(alpha)-Typ
mit optimierter E. coli-Codonverwendung

<400> 20

Gln Leu Ser Tyr Gly Tyr Asp Glu Lys Ser Thr Gly Gly Ile Ser Val

1

5

10

15

Pro Gly Pro Met Gly Pro Ser Gly Pro Arg Gly Leu Pro Gly Pro Pro
 20 25 30

Gly Ala Pro Gly Pro Gln Gly Phe Gln Gly Pro Pro Gly Glu Pro Gly
 35 40 45

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

Pro Gly Lys Asn Gly Asp Asp Gly Glu Ala Gly Lys Pro Gly Arg Pro
 65 70 75 80

Gly Glu Arg Gly Pro Pro Gly Pro Gln Gly Ala Arg Gly Leu Pro Gly
 85 90 95

Thr Ala Gly Leu Pro Gly Met Lys Gly His Arg Gly Phe Ser Gly Leu
 100 105 110

Asp Gly Ala Lys Gly Asp Ala Gly Pro Ala Gly Pro Lys Gly Glu Pro
 115 120 125

Gly Ser Pro Gly Glu Asn Gly Ala Pro Gly Gln Met Gly Pro Arg Gly
 130 135 140

Leu Pro Gly Glu Arg Gly Arg Pro Gly Ala Pro Gly Pro Ala Gly Ala
 145 150 155 160

Arg Gly Asn Asp Gly Ala Thr Gly Ala Ala Gly Pro Pro Gly Pro Thr

165

170

175

Gly Pro Ala Gly Pro Pro Gly Phe Pro Gly Ala Val Gly Ala Lys Gly

180

185

190

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

195

200

205

Arg Gly Glu Pro Gly Pro Pro Gly Pro Ala Gly Ala Ala Gly Pro Ala

210

215

220

Gly Asn Pro Gly Ala Asp Gly Gln Pro Gly Ala Lys Gly Ala Asn Gly

225

230

235

240

Ala Pro Gly Ile Ala Gly Ala Pro Gly Phe Pro Gly Ala Arg Gly Pro

245

250

255

Ser Gly Pro Gln Gly Pro Gly Gly Pro Pro Gly Pro Lys Gly Asn Ser

260

265

270

Gly Glu Pro Gly Ala Pro Gly Ser Lys Gly Asp Thr Gly Ala Lys Gly

275

280

285

Glu Pro Gly Pro Val Gly Val Gln Gly Pro Pro Gly Pro Ala Gly Glu

290

295

300

Glu Gly Lys Arg Gly Ala Arg Gly Glu Pro Gly Pro Thr Gly Leu Pro

305	310	315	320
Gly Pro Pro Gly Glu Arg Gly Gly Pro Gly Ser Arg Gly Phe Pro Gly			
325	330	335	
Ala Asp Gly Val Ala Gly Pro Lys Gly Pro Ala Gly Glu Arg Gly Ser			
340	345	350	
Pro Gly Pro Ala Gly Pro Lys Gly Ser Pro Gly Glu Ala Gly Arg Pro			
355	360	365	
Gly Glu Ala Gly Leu Pro Gly Ala Lys Gly Leu Thr Gly Ser Pro Gly			
370	375	380	
Ser Pro Gly Pro Asp Gly Lys Thr Gly Pro Pro Gly Pro Ala Gly Gln			
385	390	395	400
Asp Gly Arg Pro Gly Pro Pro Gly Pro Pro Gly Ala Arg Gly Gln Ala			
405	410	415	
Gly Val Met Gly Phe Pro Gly Pro Lys Gly Ala Ala Gly Glu Pro Gly			
420	425	430	
Lys Ala Gly Glu Arg Gly Val Pro Gly Pro Pro Gly Ala Val Gly Pro			
435	440	445	
Ala Gly Lys Asp Gly Glu Ala Gly Ala Gln Gly Pro Pro Gly Pro Ala			
450	455	460	

Gly Pro Ala Gly Glu Arg Gly Glu Gln Gly Pro Ala Gly Ser Pro Gly

465 470 475 480

Phe Gln Gly Leu Pro Gly Pro Ala Gly Pro Pro Gly Glu Ala Gly Lys

485 490 495

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

500 505 510

Gly Ala Arg Gly Glu Arg Gly Phe Pro Gly Glu Arg Gly Val Gln Gly

515 520 525

Pro Pro Gly Pro Ala Gly Pro Arg Gly Ala Asn Gly Ala Pro Gly A sn

530 535 540

Asp Gly Ala Lys Gly Asp Ala Gly Ala Pro Gly Ala Pro Gly Ser Gln

545 550 555 560

Gly Ala Pro Gly Leu Gln Gly Met Pro Gly Glu Arg G ly Ala Ala Gly

565 570 575

Leu Pro Gly Pro Lys Gly Asp Arg Gly Asp Ala Gly Pro Lys Gly Ala

580 585 590

Asp Gly Ser Pro Gly Lys Asp Gly Val A rg Gly Leu Thr Gly Pro Ile

595 600 605

Gly Pro Pro Gly Pro Ala Gly Ala Pro Gly Asp Lys Gly Glu Ser Gly

610

615

620

Pro Ser Gly Pro Ala Gly Pro Thr Gly Ala Arg Gly Ala Pro Gly Asp

625

630

635

640

Arg Gly Glu Pro Gly Pro Pro Gly Pro Ala Gly Phe Ala Gly Pro Pro

645

650

655

Gly Ala Asp Gly Gln Pro Gly Ala Lys Gly Glu Pro Gly Asp Ala Gly

660

665

670

Ala Lys Gly Asp Ala Gly Pro Pro Gly Pro Ala Gly Pro Ala Gly Pro

675

680

685

Pro Gly Pro Ile Gly Asn Val Gly Ala Pro Gly Ala Lys Gly Ala Arg

690

695

700

Gly Ser Ala Gly Pro Pro Gly Ala Thr Gly Phe Pro Gly Ala Ala Gly

705

710

715

720

Arg Val Gly Pro Pro Gly Pro Ser Gly Asn Ala Gly Pro Pro Gly Pro

725

730

735

Pro Gly Pro Ala Gly Lys Glu Gly Gly Lys Gly Pro Arg Gly Glu Thr

740

745

750

Gly Pro Ala Gly Arg Pro Gly Glu Val Gly Pro Pro Gly Pro Pro Gly

755

760

765

Pro Ala Gly Glu Lys Gly Ser Pro Gly Ala Asp Gly Pro Ala Gly Ala

770

775

780

Pro Gly Thr Pro Gly Pro Gln Gly Ile Ala Gly Gln Arg Gly Val Val

785

790

795

800

Gly Leu Pro Gly Gln Arg Gly Glu Arg Gly Phe Pro Gly Leu Pro Gly

805

810

815

Pro Ser Gly Glu Pro Gly Lys Gln Gly Pro Ser Gly Ala Ser Gly Glu

820

825

830

Arg Gly Pro Pro Gly Pro Met Gly Pro Pro Gly Leu Ala Gly Pro Pro

835

840

845

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

850

855

860

Arg Asp Gly Ser Pro Gly Ala Lys Gly Asp Arg Gly Glu Thr Gly Pro

865

870

875

880

Ala Gly Pro Pro Gly Ala Pro Gly Ala Pro Gly Ala Pro Gly Pro Val

885

890

895

Gly Pro Ala Gly Lys Ser Gly Asp Arg Gly Glu Thr Gly Pro Ala Gly
 900 905 910

Pro Ala Gly Pro Val Gly Pro Ala Gly Ala Arg Gly Pro Ala Gly Pro
 915 920 925

Gln Gly Pro Arg Gly Asp Lys Gly Glu Thr Gly Glu Gln Gly Asp Arg
 930 935 940

Gly Ile Lys Gly His Arg Gly Phe Ser Gly Leu Gln Gly Pro Pro Gly
 945 950 955 960

Pro Pro Gly Ser Pro Gly Glu Gln Gly Pro Ser Gly Ala Ser Gly Pro
 965 970 975

Ala Gly Pro Arg Gly Pro Pro Gly Ser Ala Gly Ala Pro Gly Lys Asp
 980 985 990

Gly Leu Asn Gly Leu Pro Gly Pro Ile Gly Pro Pro Gly Pro Arg Gly
 995 1000 1005

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

Pro Gly Pro Pro Gly Pro Pro Ser Ala Gly Phe Asp Phe Ser Phe Leu
 1025 1030 1035 1040

Pro Gln Pro Pro Gln Glu Lys Ala His Asp Gly Gly Arg Tyr Tyr Arg

1045

1050

105 5

Ala

<210> 21

<211> 79

<212> DNA

<213> Künstliche Sequenz

<220>

<223> Beschreibung der künstlichen Sequenz:
Rekonstituierte synthetische Oligonukleotide
vom menschlichen Gen-Kollagen-I-(alpha)-Typ
mit optimierter E. coli-Codonverwendung

<400> 21

ggaattcatg cagctgagct atggctatga tgaaaaaagc accggcggca tcagcgtgcc 60

gggcccgatg ggtccgagc

79

<210> 22

<211> 75

<212> DNA

<213> Künstliche Sequenz

<220>

<223> Beschreibung der künstlichen Sequenz:
 Rekonstituierte synthetische Oligonukleotide
 vom menschlichen Gen-Kollagen-I-(alpha)-Typ
 mit optimierter E. coli-Codonverwendung

<400> 22

ggccccgggct acccaggctc gccggg cgca ccggacggcc cgggcgggtcc agcggggcca 60
gcattattecg aaccc 75

<210> 23

<211> 81

<212> DNA

<213> Künstliche Sequenz

<220>

<223> Beschreibung der künstlichen Sequenz:
 Rekonstituierte synthetische Oligonukleotide
 vom menschlichen Gen-Kollagen-I-(alpha)-Typ
 mit optimierter E. coli-Codonverwendung

<400> 23

ggaattccgg gtccgcaggg ctttcagggt ccgccgggcg aacctggtgc gagcgggccc 60
atggggcccgc gcggcccgc c 81

<210> 24
 <211> 87
 <212> DNA
 <213> Künstliche Sequenz

<220>

<223> Beschreibung der künstlichen Sequenz:
 Rekonstituierte synthetische Oligonukleotide
 vom menschlichen Gen-Kollagen-I-(alpha)-Typ
 mit optimierter E. coli-Codonverwendung

<400> 24

taccggggcg cgccggggcg gccaggcggt ccgtttttgc cgctactacc gttcgcccgt 60
ttggccctgc aggcattatt cgaaccc 87

<210> 25
 <211> 111
 <212> DNA
 <213> Künstliche Sequenz

<220>

<223> Beschreibung der künstlichen Sequenz:
 DNA-Sequenz eines Fragments vom menschlichen
 Gen-Kollagen-I-(alpha1)-Typ mit optimierter
 E. coli-Codonverwendung codierend bei
 Plasmid pBSN-1

<400> 25

cagctgagct atggctatga tgaaaaaagc accggcggca tcagcgtgcc gggcccgatg 60
 ggtccgagcg gccctcgtgg cctgccgggc ccgccaggtg cgcccggtec g 111

<210> 26

<211> 37

<212> DNA

<213> Künstliche Sequenz

<220>

<223> Beschreibung der künstlichen Sequenz:
 Aminosäuresequenz-Sequenz eines menschlichen
 Gen-Kollagen-I-(alpha1)-Typs mit optimierter
 E. coli-Codonverwendung codierend bei
 Plasmid pBSN-1

<400> 26

Gln Leu Ser Tyr Gly Tyr Asp Glu Lys Ser Thr Gly Gly Ile Ser Val
 1 5 10 15

Pro Gly Pro Met Gly Pro Ser Gly Pro Arg Gly Leu Pro Gly Pro Pro
 20 25 30

Gly Ala Pro Gly Pro
 35

<210> 27
 <211> 240
 <212> DNA
 <213> Künstliche Sequenz

<220>

<223> Beschreibung der künstlichen Sequenz:
 DNA-Sequenz eines Fragments eines menschlichen
 Gen-Kollagen-I-(alpha1)-Typs mit optimierter
 E. coli-Codonverwendung codierend bei
 Plasmid pBSN-2

<400> 27

cagctgagct atggctatga tgaaaaaagc accggcggca tcagcgtgcc gggcccgatg 60
 ggtccgagcg gccctcgtgg cctgccgggc ccgccagggtg cgcccgggtcc gcagggcttt 120
 caggggtccgc cgggcgaacc gggcgaacct ggtgcgagcg gcccgatggg ccgcgcgggc 180
 ccgccgggtc cgccaggcaa aaacggcgat gatggcgaag cgggcaaacc gggacgtccg 240

<210> 28
 <211> 80
 <212> PRT
 <213> Künstliche Sequenz

<220>

<223> Beschreibung der künstlichen Sequenz:
 Aminosäuresequenz-Sequenz eines Fragments
 eines menschlichen Gen-Kollagen-I-(alpha1)-Typs
 mit optimierter E. coli-Codonverwendung

<400> 28

Gln Leu Ser Tyr Gly Tyr Asp Glu Lys Ser Thr Gly Gly Ile Ser Val

1

5

10

15

Pro Gly Pro Met Gly Pro Ser Gly Pro Arg Gly Leu Pro Gly Pro Pro

20

25

30

Gly Ala Pro Gly Pro Gln Gly Phe Gln Gly Pro Pro Gly Glu Pro Gly

35

40

45

Glu Pro Gly Ala Ser Gly Pro Met Gly Pro Arg Gly Pro Pro Gly Pro

50

55

60

Pro Gly Lys Asn Gly Asp Asp Gly Glu Ala Gly Lys Pro Gly Arg Pro

65

70

75

80

<210> 29

<211> 3120

<212> DNA

<213> Künstliche Sequenz

<220>

<223> Beschreibung der künstlichen Sequenz:
 DNA-Sequenz eines helikalen Bereichs des
 menschlichen Typ 1 (alpha2)-Kollagens plus
 11 Amino-terminaler extrahelikaler Aminosäuren
 und 12 Carboxy-terminaler extrahelikaler
 Aminosäuren

cagtatgatg gaaaaggagt tggacttggc cctggaccaa tgggcttaat gggacctaga 60
 ggcccacctg gtgcagctgg agccccaggc cctcaagggt tccaaggacc tgctggtgag 120
 cctggtgaac ctggtcaaac tggtcctgca ggtgctcgtg gtccagctgg ccctcctggc 180
 aaggctggtg aagatggtca ccctggaaaa ccgggacgac ctggtgagag aggagttggt 240
 ggaccacagg gtgctcgtgg tttccctgga actcctggac ttcttggtt caaaggcatt 300
 aggggacaca atggtctgga tggattgaag ggacagcccg gtgctcctgg tgtgaagggt 360
 gaacctggtg cccctggtga aaatggaact ccagggtcaaa caggagcccg tgggcttcct 420
 ggtgagagag gacgtgttgg tgcccctggc ccagctggtg cccgtggcag tgatggaagt 480
 gtgggtcccg tgggtcctgc tgggtccatt gggctctgctg gccctccagg ctcccagggt 540
 gcccttggtc ccaagggtga aattggagct gttggtaacg ctggtcctgc tgggtccgcc 600
 ggtccccgtg gtgaagtggg tcttccaggc ctctccggcc ccgttggacc tcttggtaat 660
 cctggagcaa acggccttac tggtgccaag ggtgctgctg gccttcccgg cgttgctggg 720
 gctcccggcc tccctggacc ccgcggtatt cctggccctg ttggtgctgc cggtgctact 780
 ggtgccagag gacttggttg tgagcctggt ccagctggct ccaaaggaga gagcggtaac 840
 aagggtgagc ccggtctctg tgggccccaa ggtcctcctg gtcccagtgg tgaagaagga 900
 aagagaggcc ctaatgggga agctggatct gccggccctc caggacctcc tgggctgaga 960
 ggtagtcctg gttctcgtgg tcttccctgga gctgatggca gagctggcgt catgggccct 1020
 cctggtagtc gtggtgcaag tggccctgct ggagtcogag gacctaatgg agatgctggt 1080
 cgccctgggg agcctggtct catgggaccc agaggtcttc ctggttcccc tggaaatata 1140
 ggccccgctg gaaaagaagg tctgtcggc ctccctggca tcgacggcag gcctggccca 1200
 attggcccag ctggagcaag aggagagcct ggcaacattg gattccctgg acccaaagg c 1260
 cccactggtg atcctggcaa aaacggtgat aaaggctcatg ctggtcttgc tgggtgctcg 1320
 ggtgctccag gtctgatgg aaacaatggt gctcaggac ctctggacc acagggtggt 1380

caaggtggaa aaggtgaaca ggggtcccgt ggtcctccag gcttccaggg tctgcctggc 1440
 ccctcaggtc ccgctggtga agttggcaaa ccaggagaaa ggggtctcca tggtaggttt 1500
 ggtctccctg gtcctgctgg tccaagaggg gaacgcggtc cccaggtga gagtgggtgct 1560
 gccggtccta ctggctcctat tggaagccga ggtccttctg gacccccagg gcctgatgga 1620
 aacaaggggtg aacctgggtg ggttgggtgct gtgggcactg ctgggtccatc tgg tcctagt 1680
 ggactcccag gagagagggg tgctgctggc atacctggag gcaagggaga aaagggtgaa 1740
 cctgggtctca gaggtgaaat tggtaacctt ggcagagatg gtgctcgtgg tgctcatggt 1800
 gctgtaggtg cccctgggtc tgctggagcc acaggtgacc ggggcgaagc tggggctgct 1860
 ggtcctgctg gtcctgctgg tctcgggga agccctgggtg aacgtggcga ggtcgggtcct 1920
 gctggcccca acggatttgc tggtcgggtt ggtgctgctg gtcaaccggg tgctaaagga 1980
 gaaagaggag ccaaagggcc taagggtgaa aacggtgttg ttggtccac agggcccgtt 2040
 ggagctgctg gccagctgg tccaaatggt ccccccggtc ctgctgga ag tcgtgggtgat 2100
 ggaggccccc ctggtatgac tggtttccct ggtgctgctg gacggactgg tccccagga 2160
 ccctctggtg tttctggccc tccctgggtc cctgggtcctg ctgggaaaga agggcttcgt 2220
 ggtcctcgtg gtgaccaagg tccagttggc cgaactggag aagtaggtgc agttgggtccc 2280
 cctggcttcg ctggtgagaa gggtcctctt ggagaggctg gtactgctgg acctcctggc 2340
 actccaggtc ctcagggtct tcttgggtgct cctgggtattc tgggtctccc tggctcgaga 2400
 ggtgaacgtg gtctacctgg tgttgctggt gctgtgggtg aacctgggtc tcttggcatt 2460
 gccggccctc ctggggcccg tggctcctct ggtgctgctg gt agtccctgg agtcaacggt 2520
 gctcctgggtg aagctggctg tgatggcaac cctgggaacg atgggtcccc aggtcgcgat 2580
 ggtcaacccg gacacaaggg agagcgcggt taccctggca atattgggtc cgttgggtgct 2640
 gcaggtgcac ctgggtcctc tggccccgtg ggtcctgctg gcaaaccatg aaaccgtggt 2700
 gaaaactggtc cttctgggtc tgttgggtcct gctgggtgctg ttggcccaag aggtcctagt 2760
 ggcccacaag gcattcgtgg cgataagggg gagcccgggtg aaaagggggc cagaggtctt 2820
 cctggcttaa agggacacaa tggattgcaa ggtctgcctg gtatcgctgg tcaccatggt 2880
 gatcaaggtg ctcctgggtc cgtgggtcct gctgggtc cta ggggccctgc tgggtccttct 2940
 ggccctgctg gaaaagatgg tcgcactgga catcctggta cggttggacc tgctggcatt 3000
 cgaggccctc agggtcacca aggcctgct gggccccctg gtccccctgg cctcctgga 3060

cctccaggtg taagcgggtgg tggttatgac ttggttacg atggagactt ctacagggct 3120

<210> 30

<211> 1040

<212> PRT

<213> Künstliche Sequenz

<220>

<223> Beschreibung der künstlichen Sequenz:

Aminosäuresequenz eines helikalen Bereichs des menschlichen Typ 1 (alpha2)-Kollagens plus 11 Amino-terminaler extrahelikaler Aminosäuren und 12 Carboxy-terminaler extrahelikaler Aminosäuren

<400> 30

Gln Tyr Asp Gly Lys Gly Val Gly Leu Gly Pro Gly Pro Met Gly Leu

1

5

10

15

Met Gly Pro Arg Gly Pro Pro Gly Ala Ala Gly Ala Pro Gly Pro G ln

20

25

30

Gly Phe Gln Gly Pro Ala Gly Glu Pro Gly Glu Pro Gly Gln Thr Gly

35

40

45

Pro Ala Gly Ala Arg Gly Pro Ala Gly Pro Pro Gly L ys Ala Gly Glu

50

55

60

Asp Gly His Pro Gly Lys Pro Gly Arg Pro Gly Glu Arg Gly Val Val

65

70

75

80

Gly Pro Gln Gly Ala Arg Gly Phe Pro Gly Thr Pro Gly Leu Pro Gly

85

90

95

Phe Lys Gly Ile Arg Gly His Asn Gly Leu Asp Gly Leu Lys Gly Gln

100

105

110

Pro Gly Ala Pro Gly Val Lys Gly Glu Pro Gly Ala Pro Gly Glu Asn

115

120

125

Gly Thr Pro Gly Gln Thr Gly Ala Arg Gly Leu Pro Gly Glu Arg Gly

130

135

140

Arg Val Gly Ala Pro Gly Pro Ala Gly Ala Arg Gly Ser Asp Gly Ser

145

150

155

160

Val Gly Pro Val Gly Pro Ala Gly Pro Ile Gly Ser Ala Gly Pro Pro

165

170

175

Gly Phe Pro Gly Ala Pro Gly Pro Lys Gly Glu Ile Gly Ala Val Gly

180

185

190

Asn Ala Gly Pro Ala Gly Pro Ala Gly Pro Arg Gly Glu Val Gly Leu

195

200

205

Pro Gly Leu Ser Gly Pro Val Gly Pro Pro Gly Asn Pro Gly Ala Asn

210

215

220

Gly Leu Thr Gly Ala Lys Gly Ala Ala Gly Leu Pro Gly Val Ala Gly

225

230

235

240

Ala Pro Gly Leu Pro Gly Pro Arg Gly Ile Pro Gly Pro Val Gly Ala

245

250

255

Ala Gly Ala Thr Gly Ala Arg Gly Leu Val Gly Glu Pro Gly Pro Ala

260

265

270

Gly Ser Lys Gly Glu Ser Gly Asn Lys Gly Glu Pro Gly Ser Ala Gly

275

280

285

Pro Gln Gly Pro Pro Gly Pro Ser Gly Glu Glu Gly Lys Arg Gly Pro

290

295

300

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

305

310

315

320

Gly Ser Pro Gly Ser Arg Gly Leu Pro Gly Ala Asp Gly Arg Ala Gly

325

330

335

Val Met Gly Pro Pro Gly Ser Arg Gly Ala Ser Gly Pro Ala Gly Val

340

345

350

Arg Gly Pro Asn Gly Asp Ala Gly Arg Pro Gly Glu Pro Gly Leu Met

355

360

365

Gly Pro Arg Gly Leu Pro Gly Ser Pro Gly Asn Ile Gly Pro Ala Gly

370

375

380

Lys Glu Gly Pro Val Gly Leu Pro Gly Ile Asp Gly Arg Pro Gly Pro

385

390

395

400

Ile Gly Pro Ala Gly Ala Arg Gly Glu Pro Gly Asn Ile Gly Phe Pro

405

410

415

Gly Pro Lys Gly Pro Thr Gly Asp Pro Gly Lys Asn Gly Asp Lys Gly

420

425

430

His Ala Gly Leu Ala Gly Ala Arg Gly Ala Pro Gly Pro Asp Gly Asn

435

440

445

Asn Gly Ala Gln Gly Pro Pro Gly Pro Gln Gly Val Gln Gly Gly Lys

450

455

460

Gly Glu Gln Gly Pro Ala Gly Pro Pro Gly Phe Gln Gly Leu Pro Gly

465

470

475

480

Pro Ser Gly Pro Ala Gly Glu Val Gly Lys Pro Gly Glu Arg Gly Leu

485

490

495

His Gly Glu Phe Gly Leu Pro Gly Pro Ala Gly Pro Arg Gly Glu Arg

500

505

510

Gly Pro Pro Gly Glu Ser Gly Ala Ala Gly Pro Thr Gly Pro Ile Gly

515

520

525

Ser Arg Gly Pro Ser Gly Pro Pro Gly Pro Asp Gly Asn Lys Gly Glu

530

535

540

Pro Gly Val Val Gly Ala Val Gly Thr Ala Gly Pro Ser Gly Pro Ser

545

550

555

560

Gly Leu Pro Gly Glu Arg Gly Ala Ala Gly Ile Pro Gly Gly Lys Gly

565

570

575

Glu Lys Gly Glu Pro Gly Leu Arg Gly Glu Ile Gly Asn Pro Gly Arg

580

585

590

Asp Gly Ala Arg Gly Ala His Gly Ala Val Gly Ala Pro Gly Pro Ala

595

600

605

Gly Ala Thr Gly Asp Arg Gly Glu Ala Gly Ala Ala Gly Pro Ala Gly

610

615

620

Pro Ala Gly Pro Arg Gly Ser Pro Gly Glu Arg Gly Glu Val Gly Pro

625

630

635

640

Ala Gly Pro Asn Gly Phe Ala Gly Pro Ala Gly Ala Ala Gly Gln Pro

645

650

655

Gly Ala Lys Gly Glu Arg Gly Ala Lys Gly Pro Lys Gly Glu Asn Gly

660

665

670

Val Val Gly Pro Thr Gly Pro Val Gly Ala Ala Gly Pro Ala Gly Pro

675

680

685

Asn Gly Pro Pro Gly Pro Ala Gly Ser Arg Gly Asp Gly Gly Pro Pro

690

695

700

Gly Met Thr Gly Phe Pro Gly Ala Ala Gly Arg Thr Gly Pro Pro Gly

705

710

715

720

Pro Ser Gly Ile Ser Gly Pro Pro Gly Pro Pro Gly Pro Ala Gly Lys

725

730

735

Glu Gly Leu Arg Gly Pro Arg Gly Asp Gln Gly Pro Val Gly Arg Thr

740

745

750

Gly Glu Val Gly Ala Val Gly Pro Pro Gly Phe Ala Gly Glu Lys Gly

755

760

765

Pro Ser Gly Glu Ala Gly Thr Ala Gly Pro Pro Gly Thr Pro Gly Pro

770

775

780

Gln Gly Leu Leu Gly Ala Pro Gly Ile Leu Gly Leu Pro Gly Ser Arg

785

790

795

800

Gly Glu Arg Gly Leu Pro Gly Val Ala Gly Ala Val Gly Glu Pro Gly

805

810

815

Pro Leu Gly Ile Ala Gly Pro Pro Gly Ala Arg Gly Pro Pro Gly Ala

820

825

830

Val Gly Ser Pro Gly Val Asn Gly Ala Pro Gly Glu Ala Gly Arg Asp

835

840

845

Gly Asn Pro Gly Asn Asp Gly Pro Pro Gly Arg Asp Gly Gln Pro Gly

850

855

860

His Lys Gly Glu Arg Gly Tyr Pro Gly Asn Ile Gly Pro Val Gly Ala

865

870

875

880

Ala Gly Ala Pro Gly Pro His Gly Pro Val Gly Pro Ala Gly Lys His

885

890

895

Gly Asn Arg Gly Glu Thr Gly Pro Ser Gly Pro Val Gly Pro Ala Gly

900

905

910

Ala Val Gly Pro Arg Gly Pro Ser Gly Pro Gln Gly Ile Arg Gly Asp

915

920

925

Lys Gly Glu Pro Gly Glu Lys Gly Pro Arg Gly Leu Pro Gly Leu Lys

930

935

940

Gly His Asn Gly Leu Gln Gly Leu Pro Gly Ile Ala Gly His His Gly
 945 950 955 960

Asp Gln Gly Ala Pro Gly Ser Val Gly Pro Ala Gly Pro Arg Gly Pro
 965 970 975

Ala Gly Pro Ser Gly Pro Ala Gly Lys Asp Gly Arg Thr Gly His Pro
 980 985 990

Gly Thr Val Gly Pro Ala Gly Ile Arg Gly Pro Gln Gly His Gln Gly
 995 1000 1005

Pro Ala Gly Pro Pro Gly Pro Pro Gly Pro Pro Gly Pro Pro Gly Val
 1010 1015 1020

Ser Gly Gly Gly Tyr Asp Phe Gly Tyr Asp Gly Asp Phe Tyr Arg Ala
 1025 1030 1035 1040

<210> 31

<211> 3120

<212> DNA

<213> Künstliche Sequenz

<220>

<223> Beschreibung der künstlichen Sequenz:

DNA-Sequenz von HuCol (alpha2)-E. coli der
helikalen Region des menschlichen Typ 1 (alpha2)
Kollagens plus 11 Amino-terminaler
extrahelikaler Aminosäuren und 12 Carboxy-
terminaler extrahelikaler Aminosäuren

<400> 31

```

cagtacgacg gtaaaggcgt aggcctgggt ccgggtccga tgggcctgat ggggccacgt 60
ggccccaccgg gtgcagcagg tgcgccgggt ccgcagggct tccaagggtcc ggcgggtgaa 120
ccgggcgaac cgggtcagac gggccggcg ggtgctcgcg gtccggctgg cccaccgggc 180
aaagctggcg aagacgggtca cccgggtaag ccaggccgcc cgggcgaacg tggcgctcgtg 240
gggtccgcaag gtgcgcgtgg ttccccgggc acgccgggtc tgccgggttt caaaggcatt 300
cgtggtcaca acggtctgga cggctctgaa ggccaaccgg gtgctccggg cgtcaaaggc 360
gaaccgggtg cccagggcga aaacgggtac ccgggccaga ctgggtgcgcg tgggtctgccg 420
gggtgaacgcg gccgtgttgg cgctccgggt ccggctggcg cgcgtggcag cgatgggtcc 480
gtcgggtccgg ttggccctgc ggggtccgatt ggttccgctg gccctccggg ttccccgggt 540
gcgccgggtc cgaagggtga gatcggcgcg gttggcaacg caggccccgc tgggtccagcc 600
ggccctcgtg gcgaagtcgg tctgccgggt ctgagcggtc cggtaggccc accgggtaac 660
ccgggcgcaa acggcctgac gggtgcaaaa ggtgcggctg gcctgcgggg cgttgccggt 720
gccccggggc tgccgggtcc gcgcggtatt ccgggtccgg taggcgcagc cggtgcaact 780
ggtgcccgtg gcctggttgg cgaaccgggt ccggcggggt ctaaaggcga aagcggtaac 840
aaaggtgagc cgggttccgc gggccgcag ggtccgccgg gtccgagcgg cgaagaagggt 900
aaacgtggtc cgaacggcga ggctgggtcc gcaggccctc cgggtccgcc ggggtctgcgt 960
ggcagccccg gtagccgtgg cctgccgggc gcggacggcc gtgcgggcgt gatgggtccg 1020
ccgggttccc gtggtgcctc tgggtccggt ggtgtcgtg gtccgaatgg cgacgcgggc 1080
cgtccgggtg aaccgggcct gatgggtccg cgtggcctgc cgggtagccc gggtaacatt 1140

```

ggtccggcgg gtaaggaggg tccggtaggt ctgccgggta ttgatggtcg tccgggtccg 1200
 atcggccctg cgggcgctcg tggcgagccg ggtaacatcg gttttccggg tccgaagggg 1260
 ccgacggggc acccgggcaa gaacggtgat aaaggccatg caggctctggc aggtgcccgt 1320
 ggtgcaccgg gtccggatgg taacaatggg gcgcagggtc cgccgggtcc gcagggcgta 1380
 cagggtggca aaggtgaaca ggggtccggca ggcccaccgg gcttccaggg tctgccgggt 1440
 ccgagcggcc cggctgggta agtgggcaaa ccgggcgaac gtggcctcca tggcgagttt 1500
 ggcctgccgg gtccggcccg tccgcgtggg gagcgcggcc ctccgggcga atccggcgcg 1560
 gcaggtccga cgggcccgat tggttcccgt ggtccgagcg gccaccggg tccggacggc 1620
 aacaaaggcg agccgggtgt tgttggtgct gttggtaccg ccggcccgtc tgggtccgagc 1680
 ggtctgccgg gcgaacgcgg tgccgctggg attccgggcg gcaaagggtga aaaagggtgaa 1740
 ccgggtctgc gcggtgagat tggcaaccgg ggccgtgacg gtgctcgcgg tgacacggc 1800
 gcggttggcg caccgggtcc ggcaggcgcg actggtgatc gtggcgaagc tgggtgcagcg 1860
 ggtccggcgg gtccggcccg ccctcgcggg tcccggggcg aacgcggcga agtcggccccg 1 920
 gctggcccga atggctttgc tggcccagcg ggcgctgcgg gccaacggg tgcgaaagg 1980
 gagcgcggtg ccaaaggccc gaaagggtgaa aatgggtgtag ttggtccgac ggggtccggtt 2040
 ggtgcggctg gtccggctgg cccgaatggg ccgcgggtc cggcaggcag ccgtggcgat 2100
 ggtggcccac cgggcatgac cggtttccct ggcgcggccg gtcgcaccgg ccgcccgggt 2160
 ccgtctggca tttctggccc accgggtccg ccgggtccgg cgggcaaaga aggtctgcgt 2220
 ggcccacggc gcgaccaggg tccggtgggc cgtaccggcg aagtcggtgc tgttggccct 2280
 ccgggctttg cgggtgagaa aggtccgagc ggtgaagctg gcaccgcagg cccgcc gggg 2340
 acgcggggtc cgcaaggctt gctgggtgct ccgggtatcc tgggcctgcc ggggtcccgt 2400
 ggcgaacggc gtctgccggg cgttgcaggc gctgtaggcg aaccgggtcc gctgggtatc 2460
 gcgggtccgc cgggtgcgcg tgggtccggc ggtgccgtgg gctctccggg tgttaacggc 2520
 gcccctgggtg aagcgggccc gcacggcaat ccgggcaacg atgggtccgc gggctcgtgat 2580
 ggtcagccgg gtcacaaagg tgagcgtggc taccgggta acatcggtcc ggttgggtgcg 2640
 gccggcgctc cgggtccgca cgggtccgta ggcccagccg gcaaacacgg taaccgtggg 2700
 gaaacgggtc cgtccgggtc ggtaggtccg gcgggtgctg ttggtccaag cggcccgtcc 2760
 ggcccgcagg gtattcgcgg tgacaaaggc gaaccgggcg aaaaagggtcc gcgtgggtctg 2820

ccgggcctta agggccacaa cgggtctgcaa ggtctgccgg gtatcgcggg tcaccacggt 2880
gatcaggggtg ctccgggttc cgttgggtccg gccgggtccgc gtggcccggc tgggtccgtct 2940
ggtccgggccg gtaaagacgg ccgtacgggc caccgggta cggtaggggtcc ggccggcatt 3000
cgcggtccgc aaggcacca ggggtccggcg ggtccgccgg gtccgccggg tccgccgggt 3060
ccgccgggtg ttagcggtgg cggttatgat tttggttatg acggtgattt ctatcgtgcg 3120

<210> 32

<211> 1040

<212> PRT

<213> Künstliche Sequenz

<220>

<223> Beschreibung der künstlichen Sequenz:

Aminosäuresequenz von HuCol (alpha2)-E. coli der
helikalen Region des menschlichen Typ 1 (alpha2)-
Kollagens mit optimierter E. coli-Codonverwendung

<400> 32

Gln Tyr Asp Gly Lys Gly Val Gly Leu Gly Pro Gly Pro Met Gly Leu

1 5 10 15

Met Gly Pro Arg Gly Pro Pro Gly Ala Ala Gly Ala Pro Gly Pro Gln

20 25 30

Gly Phe Gln Gly Pro Ala Gly Glu Pro Gly Glu Pro Gly Gln Thr Gly

35 40 45

Pro Ala Gly Ala Arg Gly Pro Ala Gly Pro Pro Gly Lys Ala Gly Glu

50

55

60

Asp Gly His Pro Gly Lys Pro Gly Arg Pro Gly Glu Arg Gly Val Val

65

70

75

80

Gly Pro Gln Gly Ala Arg Gly Phe Pro Gly Thr Pro Gly Leu Pro Gly

85

90

95

Phe Lys Gly Ile Arg Gly His Asn Gly Leu Asp Gly Leu Lys Gly Gln

100

105

110

Pro Gly Ala Pro Gly Val Lys Gly Glu Pro Gly Ala Pro Gly Glu Asn

115

120

125

Gly Thr Pro Gly Gln Thr Gly Ala Arg Gly Leu Pro Gly Glu Arg Gly

130

135

140

Arg Val Gly Ala Pro Gly Pro Ala Gly Ala Arg Gly Ser Asp Gly Ser

145

150

155

160

Val Gly Pro Val Gly Pro Ala Gly Pro Ile Gly Ser Ala Gly Pro Pro

165

170

175

Gly Phe Pro Gly Ala Pro Gly Pro Lys Gly Glu Ile Gly Ala Val Gly

180

185

190

Asn Ala Gly Pro Ala Gly Pro Ala Gly Pro Arg Gly Glu Val Gly Leu

195

200

205

Pro Gly Leu Ser Gly Pro Val Gly Pro Pro Gly Asn Pro Gly Ala Asn

210

215

220

Gly Leu Thr Gly Ala Lys Gly Ala Ala Gly Leu Pro Gly Val Ala Gly

225

230

235

240

Ala Pro Gly Leu Pro Gly Pro Arg Gly Ile Pro Gly Pro Val Gly Ala

245

250

255

Ala Gly Ala Thr Gly Ala Arg Gly Leu Val Gly Glu Pro Gly Pro Ala

260

265

270

Gly Ser Lys Gly Glu Ser Gly Asn Lys Gly Glu Pro Gly Ser Ala Gly

275

280

285

Pro Gln Gly Pro Pro Gly Pro Ser Gly Glu Glu Gly Lys Arg Gly Pro

290

295

300

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

305

310

315

320

Gly Ser Pro Gly Ser Arg Gly Leu Pro Gly Ala Asp Gly Arg Ala Gly

325

330

335

Val Met Gly Pro Pro Gly Ser Arg Gly Ala Ser Gly Pro Ala Gly Val

340

345

350

Arg Gly Pro Asn Gly Asp Ala Gly Arg Pro Gly Glu Pro Gly Leu Met

355

360

365

Gly Pro Arg Gly Leu Pro Gly Ser Pro Gly Asn Ile Gly Pro Ala Gly

370

375

380

Lys Glu Gly Pro Val Gly Leu Pro Gly Ile Asp Gly Arg Pro Gly Pro

385

390

395

400

Ile Gly Pro Ala Gly Ala Arg Gly Glu Pro Gly Asn Ile Gly Phe Pro

405

410

415

Gly Pro Lys Gly Pro Thr Gly Asp Pro Gly Lys Asn Gly Asp Lys Gly

420

425

430

His Ala Gly Leu Ala Gly Ala Arg Gly Ala Pro Gly Pro Asp Gly Asn

435

440

445

Asn Gly Ala Gln Gly Pro Pro Gly Pro Gln Gly Val Gln Gly Gly Lys

450

455

460

Gly Glu Gln Gly Pro Ala Gly Pro Pro Gly Phe Gln Gly Leu Pro Gly

465

470

475

480

Pro Ser Gly Pro Ala Gly Glu Val Gly Lys Pro Gly Glu Arg Gly Leu

485

490

495

His Gly Glu Phe Gly Leu Pro Gly Pro Ala Gly Pro Arg Gly Glu Arg

500

505

510

Gly Pro Pro Gly Glu Ser Gly Ala Ala Gly Pro Thr Gly Pro Ile Gly

515

520

525

Ser Arg Gly Pro Ser Gly Pro Pro Gly Pro Asp Gly Asn Lys Gly Glu

530

535

540

Pro Gly Val Val Gly Ala Val Gly Thr Ala Gly Pro Ser Gly Pro Ser

545

550

555

560

Gly Leu Pro Gly Glu Arg Gly Ala Ala Gly Ile Pro Gly Gly Lys Gly

565

570

575

Glu Lys Gly Glu Pro Gly Leu Arg Gly Glu Ile Gly Asn Pro Gly Arg

580

585

590

Asp Gly Ala Arg Gly Ala His Gly Ala Val Gly Ala Pro Gly Pro Ala

595

600

605

Gly Ala Thr Gly Asp Arg Gly Glu Ala Gly Ala Ala Gly Pro Ala Gly

610

615

620

Pro Ala Gly Pro Arg Gly Ser Pro Gly Glu Arg Gly Glu Val Gly Pro

625

630

635

640

Ala Gly Pro Asn Gly Phe Ala Gly Pro Ala Gly Ala Ala Gly Gln Pro

645

650

655

Gly Ala Lys Gly Glu Arg Gly Ala Lys Gly Pro Lys Gly Glu Asn Gly

660

665

670

Val Val Gly Pro Thr Gly Pro Val Gly Ala Ala Gly Pro Ala Gly Pro

675

680

685

Asn Gly Pro Pro Gly Pro Ala Gly Ser Arg Gly Asp Gly Gly Pro Pro

690

695

700

Gly Met Thr Gly Phe Pro Gly Ala Ala Gly Arg Thr Gly Pro Pro Gly

705

710

715

720

Pro Ser Gly Ile Ser Gly Pro Pro Gly Pro Pro Gly Pro Ala Gly Lys

725

730

735

Glu Gly Leu Arg Gly Pro Arg Gly Asp Gln Gly Pro Val Gly Arg Thr

740

745

750

Gly Glu Val Gly Ala Val Gly Pro Pro Gly Phe Ala Gly Glu Lys Gly

755

760

765

Pro Ser Gly Glu Ala Gly Thr Ala Gly Pro Pro Gly Thr Pro Gly Pro

770

775

780

Gln Gly Leu Leu Gly Ala Pro Gly Ile Leu Gly Leu Pro Gly Ser Arg

785

790

795

800

Gly Glu Arg Gly Leu Pro Gly Val Ala Gly Ala Val Gly Glu Pro Gly

805

810

815

Pro Leu Gly Ile Ala Gly Pro Pro Gly Ala Arg Gly Pro Pro Gly Ala

820

825

830

Val Gly Ser Pro Gly Val Asn Gly Ala Pro Gly Glu Ala Gly Arg Asp

835

840

845

Gly Asn Pro Gly Asn Asp Gly Pro Pro Gly Arg Asp Gly Gln Pro Gly

850

855

860

His Lys Gly Glu Arg Gly Tyr Pro Gly Asn Ile Gly Pro Val Gly Ala

865

870

875

880

Ala Gly Ala Pro Gly Pro His Gly Pro Val Gly Pro Ala Gly Lys His

885

890

895

Gly Asn Arg Gly Glu Thr Gly Pro Ser Gly Pro Val Gly Pro Ala Gly

900

905

910

Ala Val Gly Pro Arg Gly Pro Ser Gly Pro Gln Gly Ile Arg Gly Asp

915

920

925

Lys Gly Glu Pro Gly Glu Lys Gly Pro Arg Gly Leu Pro Gly Leu Lys

930

935

940

Gly His Asn Gly Leu Gln Gly Leu Pro Gly Ile Ala Gly His His Gly

945

950

955

960

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

965

970

975

Ala Gly Pro Ser Gly Pro Ala Gly Lys Asp Gly Arg Thr Gly His Pro

980

985

990

Gly Thr Val Gly Pro Ala Gly Ile Arg Gly Pro Gln Gly His Gln Gly

995

1000

1005

Pro Ala Gly Pro Pro Gly Pro Pro Gly Pro Pro Gly Pro Pro Gly Val

1010

1015

1020

Ser Gly Gly Gly Tyr Asp Phe Gly Tyr Asp Gly Asp Phe Tyr Arg Ala

1025

1030

1035

1040

<210> 33

<211> 76

<212> DNA

<213> Künstliche Sequenz

<220>

<223> Beschreibung der künstlichen Sequenz:
 Synthetische Oligonukleotide für die
 Rekonstruktion der ersten 240 Basenpaare
 des menschlichen Typ 1 (alpha2)-Kollagen-Gens
 mit optimierter E. coli-Codonverwendung

<400> 33

ggaattcatg cagtatgatg gcaaagcggt cggcctcggc ccggggcccaa tgggcctcat 60
gggcccgcgc ggccca **76**

<210> 34

<211> 79

<212> DNA

<213> Künstliche Sequenz

<220>

<223> Beschreibung der künstlichen Sequenz:
 Synthetische Oligonukleotide für die
 Rekonstruktion der ersten 240 Basenpaare
 des menschlichen Typ 1 (alpha2)-Kollagen-Gens
 mit optimierter E. coli-Codonverwendung

<400> 34

ccggggcgcgc cgggtggccc acgtcgaccg cgggggtccgg gcgttccaaa ggtcccggga 60
cggccaatta ttccaaccc **79**

<210> 35
 <211> 82
 <212> DNA
 <213> Künstliche Sequenz

<220>

<223> Beschreibung der künstlichen Sequenz:
 Synthetische Oligonukleotide für die
 Rekonstruktion der ersten 240 Basenpaare
 des menschlichen Typ 1 (alpha2)-Kollagen-Gens
 mit optimierter E. coli-Codonverwendung

<400> 35

ggaattcgcc ggtgagccgg gtgaaccggg ccaaaccggg ccggcaggtc cacgtgggtcc 60
agcggggcccg cctggcaagg cg 82

<210> 36
 <211> 84
 <212> DNA
 <213> Künstliche Sequenz

<220>

<223> Beschreibung der künstlichen Sequenz:
 Synthetische Oligonukleotide für die
 Rekonstruktion der ersten 240 Basenpaare
 des menschlichen Typ 1 (alpha2)-Kollagen-Gens
 mit optimierter E. coli-Codonverwendung

<400> 36
ccggggcggac cgttccgccc acttctaccg gtgggaccgt ttggcccggc gggccactcg 60
caccgcatca cattattcga accc 84

<210> 37

<211> 240

<212> DNA

<213> Künstliche Sequenz

<220>

<223> Beschreibung der künstlichen Sequenz:
 DNA-Sequenz eines Fragments des menschlichen
 Kollagen-Gens vom Typ I (alpha2) mit
 optimierter E. coli-Verwendung

<400> 37
cagtatgatg gcaaaggcgt cggcctcggc ccggggccaa tgggcctcat gggcccgcgc 60
ggcccaccgg gtgcagctgg cgccccaggc ccgcaagggt tccaggggcc tgccggtgag 120
ccgggtgaac cgggccaac gggtcggca ggtgcacgtg gtccagcggg cccgcctggc 180
aaggcgggtg aagatggcca ccctggcaaa ccggggccgc cgggtgagcg tggcgtagtg 240

<210> 38

<211> 80

<212> PRT

<213> Künstliche Sequenz

<220>

<223> Beschreibung der künstlichen Sequenz:
 Aminosäuresequenz eines Fragments des menschlichen
 Kollagen-Gens vom Typ I (alpha2) mit
 optimierter E. coli-Verwendung

<400> 38

Gln Tyr Asp Gly Lys Gly Val Gly Leu Gly Pro Gly Pro Met Gly Leu

1

5

10

15

Met Gly Pro Arg Gly Pro Pro Gly Ala Ala Gly Ala Pro Gly Pro Gln

20

25

30

Gly Phe Gln Gly Pro Ala Gly Glu Pro Gly Glu Pro Gly Gln Thr Gly

35

40

45

Pro Ala Gly Ala Arg Gly Pro Ala Gly Pro Pro Gly Lys Ala Gly Glu

50

55

60

Asp Gly His Pro Gly Lys Pro Gly Arg Pro Gly Glu Arg Gly Val Val

65

70

75

80

<210> 39

<211> 276

<212> DNA

<213> Künstliche Sequenz

<220>

<223> Beschreibung der künstlichen Sequenz:

DNA-Sequenz von dem Kollagen-Mimetikum 4

<400> 39

```

atggggctcg ctggcccacc gggcgaaccg ggtccgccag gcccgaaagg tccgcgtggc 60
gatagcgggc tcgctggccc accgggcgaa ccgggtccgc caggcccgaa aggtccgcgt 120
ggcgatagcg ggctcgctgg ccacccgggc gaaccgggtc cgccaggccc gaaagggtccg 180
cgtggcgata gcgggctcgc tggcccaccg ggcgaaccgg gtccgccagg cccgaaaggt 240
ccgcgtggcg atagcgggct cccgggcgat tcctaa 276

```

<210> 40

<211> 91

<212> PRT

<213> Künstliche Sequenz

<220>

<223> Beschreibung der künstlichen Sequenz:

Aminosäuresequenz von dem Kollagen-Mimetikum 4

<400> 40

Met Gly Leu Ala Gly Pro Pro Gly Glu Pro Gly Pro P ro Gly Pro Lys

1

5

10

15

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

20

25

30

Pro Pro Gly Pro Lys Gly Pro Arg Gly A sp Ser Gly Leu Ala Gly Pro

35

40

45

Pro Gly Glu Pro Gly Pro Pro Gly Pro Lys Gly Pro Arg Gly Asp Ser

50

55

60

Gly Leu Ala Gly Pro Pro Gly Glu Pro Gly Pro Pro Gly Pro Lys Gly

65

70

75

80

Pro Arg Gly Asp Ser Gly Leu Pro Gly Asp Ser

85

90

<210> 41

<211> 13

<212> PRT

<213> Künstliche Sequenz

<220>

<223> Beschreibung der künstlichen Sequenz:

Vorhergesagte Aminosäuresequenz der DNA-

Sequenz der ersten 13 Aminosäuren des Proteins

D4-alpha1

<400> 41

Gly Pro Pro Gly Leu Ala Gly Pro Pro Gly Glu Ser G ly

1

5

10

<210> 42
<211> 13
<212> PRT
<213> Künstliche Sequenz

<220>
<223> Xaa an den Positionen 2 und 3 ist 4-Hydroxyprolin

<220>
>223> Xaa an den Positionen 8 und 9 ist 4-Hydroxyprolin

<220>
<223> Beschreibung der künstlichen Sequenz:
Experimentell bestimmte Aminosäuresequenz
der DNA-Sequenz der ersten 13 Aminosäuren
des Proteins D4-alpha1

<400> 42

Gly Xaa Xaa Gly Leu Ala Gly Xaa Xaa Gly Glu Ser Gly

1 5 10

<210> 43
<211> 660
<212> DNA
<213> Künstliche Sequenz

<220>

<223> Beschreibung der künstlichen Sequenz:
 DNA-Sequenz eines 657 Nukleotide langen
 Fragments vom menschlichen Kollagen-Typ-I-
 (alpha1)3' mit optimierter E. coli-
 Codonverwendung, die als D4 bezeichnet wird

<400> 43

```

atggggcccg cgggtctggc gggccctccg ggtgaaagcg gtcgtgaagg cgcgccgggt 60
gccgaaggca gccagggccg cgacggtagc ccggggggcca aaggggatcg tggtgaaacc 120
ggcccgggcg gccccccggg tgcaccgggc gcgccgggtg cccagggccc ggtggggcccg 180
gcgggcaaaa gcggtgatcg tggtgagacc ggtccggcgg gcccgggccg tccggtgggc 240
ccagcggggc cccgtggccc ggccgggtccg cagggcccg cgggtgacaa aggtgaaacg 300
ggcgaacagg gcgaccgtgg cattaaaggc caccgtggct tcagcggcct gcaggggtcca 360
ccggggcccg cgggcagtcg ggggtgaacag ggtccgtccg gagccagcgg gccggcgggc 420
ccacgaggtc cgcggggcag cgcggggcgc ccggggcaaag acggtctgaa cgggtctgccg 480
ggcccgatcg gcccgccggg cccacgcggc cgcaccgggt atgcgggtcc ggtgggtccc 540
ccggggcccg cgggc ccgcc agggccgcgg ggaccgccga gcgcggggtt cgacttcagc 600
ttcctgccgc agccgccgca ggagaaagcg caccgaggcg gtcgctacta ccgtgcgtaa 660

```

<210> 44

<211> 219

<212> PRT

<213> Künstliche Sequenz

<220>

<223> Beschreibung der künstlichen Sequenz:
 Aminosäuresequenz eines C-terminalen
 219 Aminosäure langen Fragments des
 menschliche4n Kollagen Typ I (alpha1), das
 als D4 bezeichnet wird

<400> 44

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

1 5 10 15

Gly Ala Pro Gly Ala Glu Gly Ser Pro Gly Arg Asp Gly Ser Pro Gly

20 25 30

Ala Lys Gly Asp Arg Gly Glu Thr Gly Pro Ala Gly Pro Pro Gly Ala

35 40 45

Pro Gly Ala Pro Gly Ala Pro Gly Pro Val Gly Pro Ala Gly Lys Ser

50 55 60

Gly Asp Arg Gly Glu Thr Gly Pro Ala Gly Pro Ala Gly Pro Val Gly

65 70 75 80

Pro Ala Gly Ala Arg Gly Pro Ala Gly Pro Gln Gly Pro Arg Gly Asp

85 90 95

Lys Gly Glu Thr Gly Glu Gln Gly Asp Arg Gly Ile Lys Gly His Arg

100

105

110

Gly Phe Ser Gly Leu Gln Gly Pro Pro Gly Pro Pro Gly Ser Pro Gly

115

120

125

Glu Gln Gly Pro Ser Gly Ala Ser Gly Pro Ala Gly Pro Arg Gly Pro

130

135

140

Pro Gly Ser Ala Gly Ala Pro Gly Lys Asp Gly Leu Asn Gly Leu Pro

145

150

155

160

Gly Pro Ile Gly Pro Pro Gly Pro Arg Gly Arg Thr Gly Asp Ala Gly

165

170

175

Pro Val Gly Pro Pro Gly Pro Pro Gly Pro Pro Gly Pro Pro Gly Pro

180

185

190

Pro Ser Ala Gly Phe Asp Phe Ser Phe Leu Pro Gln Pro Pro Gln Glu

195

200

205

Lys Ala His Asp Gly Gly Arg Tyr Tyr Arg Ala

210

215

<210> 45

<211> 627

<212> DNA

<213> Künstliche Sequenz

<220>

<223> Beschreibung der künstlichen Sequenz:

DNA-Sequenz eines 3'-Fragments aus

627 Nukleotiden des menschlichen Kollagens

vom Typ 1 (alpha2)3' mit optimierter

E. coli-Codonverwendung

<400> 45

```

atgggctctc cgggtgttaa cggcgccctt ggtgaagcgg gccgcgacgg caatccgggc 60
aacgatggtc cgccgggtcg tgatggtcag cggggtcaca aaggtgagcg tggctaccgc 120
ggtaacatcg gtccgggttg tgccggccgg gtcgggggtc cgacgggtcc ggtaggccca 180
gccggcaaac acggtaaccg tggtgaaacg ggtccgtccg gtccggtagg tccggcgggt 240
gctgttggtc caccgggccc gtccggcccc cagggtattc gcggtgacaa aggcgaaccg 300
ggcgaaaaag gtccgcgtgg tctgccgggc ctttaaggcc acaacgggtc gcaagggtcg 360
ccgggtatcg cgggtcacca cggtgatcag ggtgctccgg gttccgttgg tccggcgggt 420
ccgcgtggcc cggctgggtc gtctgggtcc gccggtaaag acggccgtac gggccaccgc 4 80
ggtaacgggtg gtccggccgg cattcgcggt ccgcaaggtc accagggtcc ggcgggtccg 540
ccgggtccgc cgggtccgcc gggtcggccg ggtgttagcg gtggcggtta tgattttggg 600
tatgacggtg atttctatcg tgcgtaa 627

```

<210> 46

<211> 219

<212> DNA

<213> Künstliche Sequenz

<220>

<223> Beschreibung der künstlichen Sequenz:
 Aminosäuresequenz eines C-terminalen
 209 Aminosäure langen Fragments des
 menschlichen Kollagens vom Typ I (alpha2)

<400> 46

<210> 47

<211> 95

<212> PRT

<213> Künstliche Sequenz

<220>

<223> Beschreibung der künstlichen Sequenz:
 Synthetische Oligonukleotide

<400> 47

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

1 5 10 15

Gly Ala Pro Gly Ala Glu Gly Ser Pro Gly Arg Asp Gly Ser Pro Gly

20 25 30

Ala Lys Gly Asp Arg Gly Glu Thr Gly Pro Ala Gly Pro Pro Gly Ala

35 40 45

Pro Gly Ala Pro Gly Ala Pro Gly Pro Val Gly Pro Ala Gly Lys Ser

50 55 60

Gly Asp Arg Gly Glu Thr Gly Pro Ala Gly Pro Ala Gly Pro Val Gly

65

70

75

80

Pro Ala Gly Ala Arg Gly Pro Ala Gly Pro Gln Gly Pro Arg Gly Asp

85

90

95

Lys Gly Glu Thr Gly Glu Gln Gly Asp Arg Gly Ile Lys Gly His Arg

100

105

110

Gly Phe Ser Gly Leu Gln Gly Pro Pro Gly Pro Pro Gly Ser Pro Gly

115

120

125

Glu Gln Gly Pro Ser Gly Ala Ser Gly Pro Ala Gly Pro Arg Gly Pro

130

135

140

Pro Gly Ser Ala Gly Ala Pro Gly Lys Asp Gly Leu Asn Gly Leu Pro

145

150

155

160

Gly Pro Ile Gly Pro Pro Gly Pro Arg Gly Arg Thr Gly Asp Ala Gly

165

170

175

Pro Val Gly Pro Pro Gly Pro Pro Gly Pro Pro Gly Pro Pro Gly Pro

180

185

190

Pro Ser Ala Gly Phe Asp Phe Ser Phe Leu Pro Gln Pro Pro Gln Glu

195

200

205

Lys Ala His Asp Gly Gly Arg Tyr Tyr Arg Ala

210

215

<210> 48

<211> 97

<212> PRT

<213> Künstliche Sequenz

<220>

<223> Beschreibung der künstlichen Sequenz:
Synthetische Oligonukleotide

<400> 48

cttcggtcgg gtcggcgct gccatcgggc ccccggtttc ccctagcacc actttggccg 60
ggcgcgccgg ggggcccacg tggcattatt cgaaacc 97

<210> 49

<211> 91

<212> DNA

<213> Künstliche Sequenz

<220>

<223> Beschreibung der künstlichen Sequenz:
Synthetische Oligonukleotide

<400> 49

ggaattcggg gcaccggggc cgcggggtgc ccagggcccg gtgggcccgg cgggcaaaag 60
 cggatgatcgt ggcgagaccg gtccggcggg c 91

<210> 50

<211> 91

<212> PRT

<213> Künstliche Sequenz

<220>

<223> Beschreibung der künstlichen Sequenz:

Synthetische Oligonukleotide

<400> 50

ctctggccag gccgcccggg ccggccaggc caccggggtc gccgcggggc accgggcccgg 60
 ccaggcgtcc cgggcgccat tattogaacc c 91

Patentansprüche

1. Verfahren zur Herstellung von extrazellulärem Matrixprotein (EMP) oder einem Fragment davon, umfassend Hydroxyprolin, das in der Lage ist, ein Selbstaggregat in einer Zelle zur Verfügung zu stellen, welche üblicherweise Prolin nicht hydroxyliert, umfassend Zurverfügungstellen einer Nukleinsäuresequenz codierend für das EMP oder Fragment davon, welche optimiert worden ist für die Expression der Zelle durch Substitution von natürlich auftretenden Codons, die von der Zelle nicht bevorzugt werden, durch Codons, die von Zelle bevorzugt werden; Inkorporieren der Nukleinsäuresequenz in die Zelle; Zurverfügungstellen von hypertonischem Wachstumsmedium enthaltend zumindest eine Aminosäure ausgewählt aus der Gruppe bestehend aus trans-4-Hydroxyprolin und 3-Hydroxyprolin; und Kontaktieren der Zelle mit dem Wachstumsmedium, worin die zumindest eine Aminosäure in die Zelle assimiliert wird und in das EMP oder Fragment davon inkorporiert wird.

2. Verfahren zur Herstellung eines extrazellulären Matrixproteins (EMP) oder Fragments davon gemäß Anspruch 1, worin das EMP ausgewählt ist aus der Gruppe bestehend aus menschlichem Kollagen, Fibrinogen, Fibronectin und kollagenähnlichem Peptid.

3. Verfahren zur Herstellung eines extrazellulären Matrixproteins (EMP) oder Fragments davon gemäß Anspruch 1 oder 2, worin die Zelle ein Prokaryont ist.

4. Verfahren zur Herstellung eines extrazellulären Matrixproteins (EMP) oder Fragments davon gemäß Anspruch 3, worin der Prokaryont E. coli ist.

5. Verfahren zur Herstellung eines extrazellulären Matrixproteins (EMP) oder Fragments davon gemäß einem der Ansprüche 2 bis 4, worin das menschliche Kollagen vom Typ I $\alpha 1$ ist.

6. Verfahren zur Herstellung eines extrazellulären Matrixproteins (EMP) oder Fragments davon gemäß Anspruch 5, worin die Nukleinsäure, die menschliches Kollagen Typ I $\alpha 1$ codiert, die Sequenz, die in SEQ ID Nr.

19 gezeigt wird, umfaßt.

7. Verfahren zur Herstellung eines extrazellulären Matrixproteins (EMP) oder Fragments davon gemäß einem der Ansprüche 2 bis 4, worin das menschliche Kollagen vom Typ I $\alpha 2$ ist.

8. Verfahren zur Herstellung eines extrazellulären Matrixproteins (EMP) oder Fragments davon gemäß Anspruch 7, worin die Nukleinsäure, die menschliches Kollagen Typ I $\alpha 2$ codiert, die Sequenz, die in SEQ ID Nr. 31 gezeigt wird, umfaßt.

9. Verfahren zur Herstellung eines extrazellulären Matrixproteins (EMP) oder Fragments davon gemäß einem der Ansprüche 1 bis 8, worin die für das EMP codierende Nukleinsäure die Sequenz, die in SEQ ID Nr. 43 gezeigt wird, umfaßt.

10. Verfahren zur Herstellung eines extrazellulären Matrixproteins (EMP) oder Fragments davon gemäß einem der Ansprüche 1 bis 8, worin die für das EMP codierende Nukleinsäure die Sequenz, die in SEQ ID Nr. 45 gezeigt wird, umfaßt.

11. Verfahren zur Herstellung eines extrazellulären Matrixproteins (EMP) oder Fragments davon gemäß einem der Ansprüche 1 bis 10, worin die Nukleinsäuresequenz eine Nukleinsäure umfaßt, die für ein physiologisch aktives Peptid codiert.

12. Verfahren zur Herstellung eines extrazellulären Matrixproteins (EMP) oder Fragments davon gemäß Anspruch 11, worin das physiologisch aktive Peptid ausgewählt ist aus der Gruppe bestehend aus knochenmorphogenem Protein, transformierendem Wachstumsfaktor- β oder Decorin.

13. Verfahren zur Herstellung eines extrazellulären Matrixproteins (EMP) oder Fragments davon gemäß einem der Ansprüche 1 bis 4, worin das EMP oder Fragment davon ein kollagenähnliches Peptid ist.

14. Verfahren zur Herstellung eines extrazellulären Matrixproteins (EMP) oder Fragments davon gemäß Anspruch 13, worin das EMP oder Fragment davon die Aminosäuresequenz, die in SEQ ID Nr. 4 gezeigt wird, umfaßt.

15. Verfahren zur Herstellung eines extrazellulären Matrixproteins (EMP) oder Fragments davon gemäß Anspruch 13, worin das EMP die Aminosäuresequenz, die in SEQ ID Nr. 40 gezeigt wird, umfaßt.

16. Verfahren zur Herstellung eines extrazellulären Matrixproteins (EMP) oder Fragments davon gemäß Anspruch 1, worin das EMP die Aminosäuresequenz, die in SEQ ID Nr. 44 gezeigt wird, umfaßt.

17. Verfahren zur Herstellung eines extrazellulären Matrixproteins (EMP) oder Fragments davon gemäß Anspruch 1, worin das EMP ein Kollagenfragment umfassend die Aminosäuresequenz, die in SEQ ID Nr. 26 gezeigt wird, ist.

18. Verfahren zur Herstellung eines extrazellulären Matrixproteins (EMP) oder Fragments davon gemäß Anspruch 1, worin das EMP ein Kollagenfragment umfassend die Aminosäuresequenz, die in SEQ ID Nr. 46 gezeigt wird, ist.

Es folgen 109 Blatt Zeichnungen

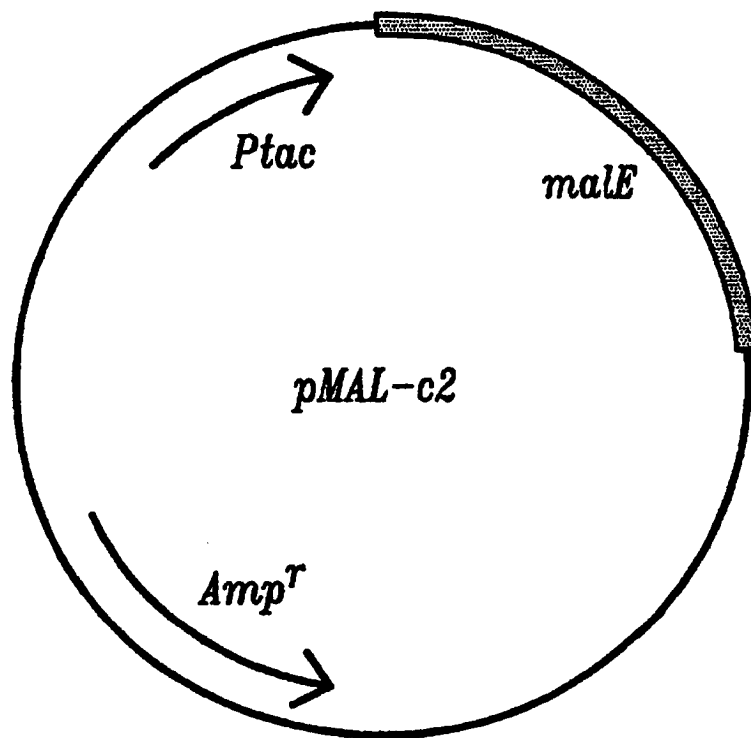
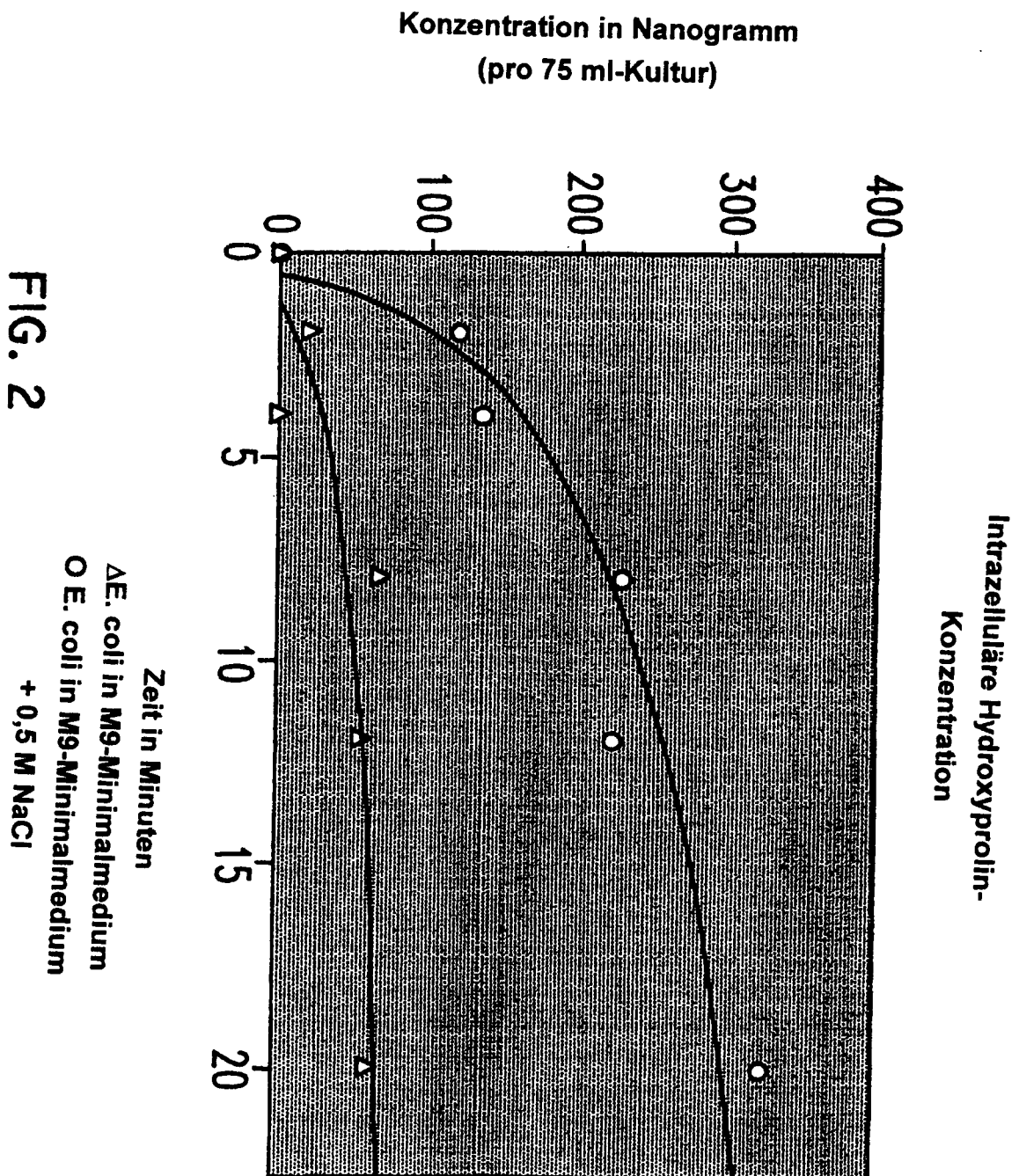


FIG. 1



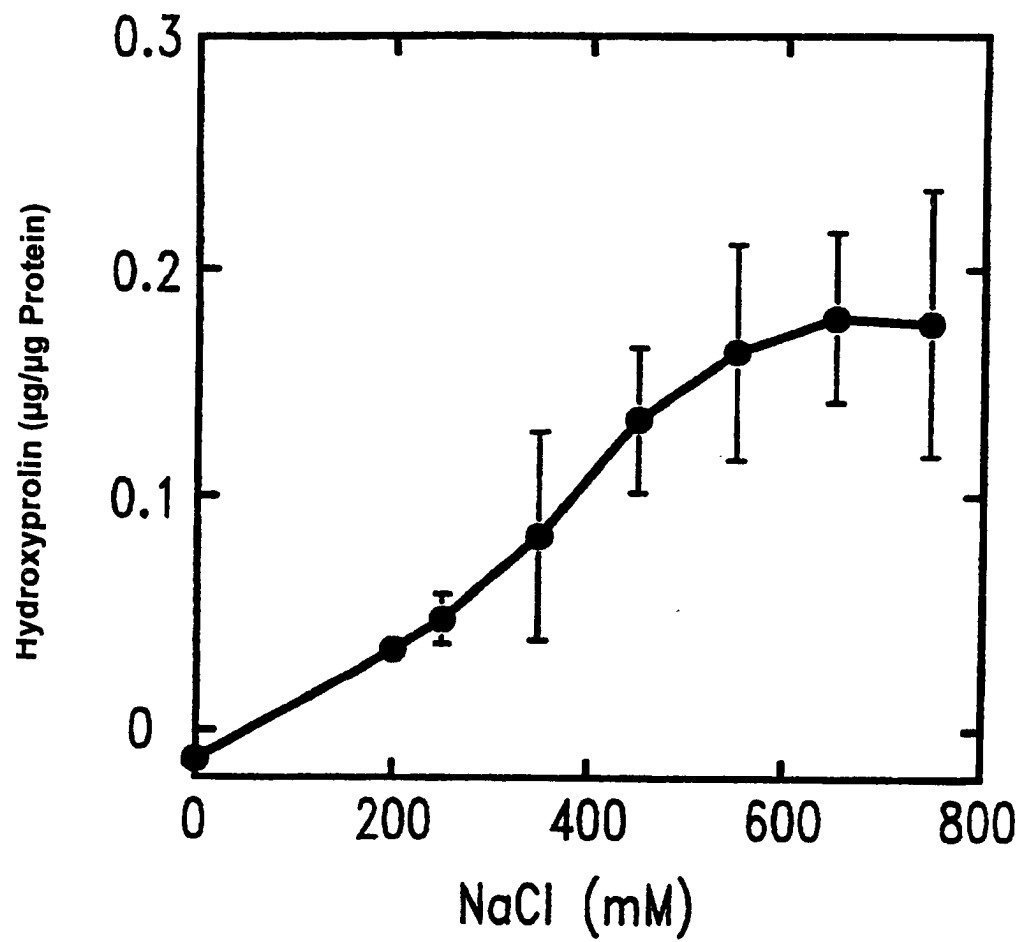


FIG. 2A

5'- CAGCTGTCTT ATGGCTATGA TGAGAAATCA ACCGGAGGAA TTTCOGTGCC
 TGGCCCCATG GGTCCCTCTG GTCTCGTGG TCTOCTGSC CCCCTGGTG
 CACCTGGTCC CCAAGGCTTC CAAGGTCCCC CTGGTGAGCC TGGOGAGCCT
 GGAGCTTCAG GTCCCATGGG TCOC CGAGGT CCCCCAGGTC CCCCTGGAAA
 GAATGGAGAT GATGGGGAAG CTGGA AAAACC TGGTCGTCTT GGTGAGCGTG
 GGCCTCCTGG GCTCAGGGT GCTCGAGGAT TGCCCGGAAC AGCTGGCCTC
 CCTGGAATGA AGGGACACAG AGGTTTCAGT GGTTTGATG GTGCCAAGGG
 AGATGCTGGT CCTGCTGGTC CTAAGGGTGA GCCTGGCAGC CCTGGTGAAA
 ATGGAGCTCC TGGTCAGATG GGCCCCCGTG GCCTGCCTGG TGAGAGAGGT
 CGCCCTGGAG CCCCTGGCCC TGCTGGTGCT CGTGGAATG ATGGTGCTAC
 TGGTGCTGCC GGGCCCCCTG GTCCACCCG CCOGCTGGT CTCCTGGCT
 TCCCTGGTGC TGTGCTGCT AAGGGTGAAG CTGGTCCCCA AGGGCCCCGA
 GGCTCTGAAG GTCCCCAGGG TGTGCTGGT GAGCTGGCC CCCCTGGCCC
 TGCTGGTGCT GCTGGCCCTG CTGGA AAACC TGGTGCTGAT GGACAGCCTG
 GTGCTAAAGG TGCCAATGGT GCTCCTGGTA TTGCTGGTGC TCCTGGCTTC
 CCTGGTGCCC GAGGCCCTC TGGACCCAG GGCCCCGGCG GOCCTCTGG
 TCCCAAGGGT AACAGCGGTG AACCTGGTGC TCCTGGCAGC AAAGGAGACA
 CTGGTGCTAA GGGAGAGCCT GGCCCTGTTG GTGTTCAAGG ACCCCTGGC
 CCTGCTGGAG AGGAAGGAAA GCGAGGAGCT CGAGGTGAAC CCGGACCCAC
 TGGCCTGCCC GGACCCCTG GCGAGCGTGG TGGACCTGGT AGCGTGCTT
 TCCCTGGCGC AGATGGTGTT GCTGGTCCCA AGGGTCCCGC TGGTGAACGT
 GGTCTCCTG GCCCCGCTGG CCCC AAAGGA TCTCTGGTG AAGCTGGTCG
 TCCCGGTGAA GCTGGTCTGC CTGGTGCCAA GGGTCTGACT GGAAGCCCTG
 GCAGCCCTGG TCCTGATGGC AAAACTGGCC CCCCCTGGTC GCGGGTCAA
 GATGGTGCC CCGGACCCOC AGGCCACCT GGTGCCCCG GTCAGGCTGG
 TGTGATGGGA TTCCCTGGAC CTAAAGGTGC TGCTGGAGAG CCGGCAAGG
 CTGGAGAGCG AGGTGTTCC GGACCCCTG GCGCTGTCG TCCTGCTGGC
 AAAGATGGAG AGGCTGGAGC TCAGGGAGCC CCTGGCCCTG CTGGTCCCGC
 TGGCGAGAGA GGTGAACAAG GOCCTGCTGG CTCCCCCGGA TTCCAGGGTC
 TCCCTGGTCC TGCTGGTCTT CCAGGTGAAG CAGGCAACC TGGTGAACAG
 GGTGTTCTCTG GAGACCTTG GCGCCCTGGC CCTCTGGAG CAAGAGGCGA
 GAGAGGTTTC CCTGGCGAGC GTGGTGTCGA AGGTCCCCCT GGTCTGCTG
 GACCCCGAGG GGCCAACGGT GCTCCCGGCA ACGATGGTGC TAAGGGTGAT
 GCTGGTGCCC CTGGAGCTCC CGGTAGCCAG GCGCCCCCTG GCCTTCAGGG
 AATGCCTGGT GAACGTGGTG CAGCTGGTCT TCAGGGCCT AAGGGTGACA
 GAGGTGATGC TGGTCCCAA GGTGCTGATG GCTCTCTGG CAAAGATGGC

FIG. 3A

GTCCG1GGTC TGACCGGCCC CATTGGTCCT CCTGGCCCTG CTGGTGCCCC.
 TGGTGACAAG GGTGAAAGTG GTCCCAGCGG CCCTGCTGGT CCCACTGGAG
 CTCGTGGTGC CCCCCGAGAC CGTGGTGAGC CTGGTCCCCC CGGCCCTGCT
 GGCTTTGCTG GCCCCCTGG TGCTGACGGC CAAOCTGGTG CTAAAGGCGA
 ACCTGGTGAT GCTGGTGCCA AAGGCGATGC TGGTCCCCCT GGGCCTGCCG
 GACCCGCTGG ACCCCCTGGC CCCATTGGTA ATGTTGGTGC TCCTGGAGCC
 AAAGGTGCTC GCGGCAGCGC TGGTCCCCCT GGTGCTACTG GTTTCOCTGG
 TGCTGCTGGC CGAGTCGGTC CTCTGGCCC CTCTGGAAT GCTGGACCCC
 CTGGCCCTCC TGGTCCTGCT GGCAAAGAAG GCGGCAAAGG TCCCCTGGGT
 GAGACTGGCC CTGCTGGACG TCCTGGTGAA GTTGGTCCCC CTGGTCCCCC
 TGGCCCTGCT GGCGAGAAAG GATCCCTGG TGCTGATGGT CCTGCTGGTG
 CTCCTGGTAC TCCCGGGCCT CAAGGTATTG CTGGACAGCG TGGTGTGGTG
 GGCTGCCTG GTCAGAGAGG ACAGAGAGGC TTCCTGGTC TTCCTGGCCC
 CTCTGGTEAA CCTGGCAAAC AAGGTCCCTC TGGAGCAAGT GGTGAACGTG
 GTCCCCCGG TCCCATGGGC CCCCCTGGAT TGGCTGGACC CCCTGGTGAA
 TCTGGACGTG AGGGGGCTCC TGCTGCCGAA GGTTCOCTG GACGAGACGG
 TTCTCCTGGC GCAAGGGTG ACOGTGGTEA GACCGGCCCC GCTGGACCCC
 CTGGTGCTCC TGGTGCTOCT GGTGCCOCTG GCCCCGTTGG CCCTGCTGGC
 AAGAGTGGTG ATCGTGGTGA GACTGGTOCT GCTGGTCCCG CCGGTCCCGT
 CGGCCCCGCT GGCGCCCGTG GCCCCCGCG ACCCCAAGGC CCCCCTGGTG
 ACAAGGGTGA GACAGGCGAA CAGGGCGACA GAGGCATAAA GGGTCACCGT
 GGCTTCTCTG GCTCCAGGG TCCCCCTGGC CCTOCTGGCT CTCCTGGTGA
 ACAAGGTCCC TCTGGAGCCT CTGGTCCTGC TGGTCCCCGA GGTCCCCCTG
 GCTCTGCTGG TGCTCCTGGC AAAGATGGAC TCAACGGTCT CCCTGGCCCC
 ATTGGGCCCC CTGGTCTCTG CGGTGCACT GGTGATGCTG GTCTGTGG
 TCCCCCGGC CCTCCTGGAC CTCTGGTCC CCCTGGTCCT CCAGCGCTG
 GTTTCGACTT CAGCTTCCTC CCCCAGCCAC CTCAAGAGAA GGCTCAGCAT
 GGTGGCCGCT ACTAOCGGC T-3'

FIG. 3B

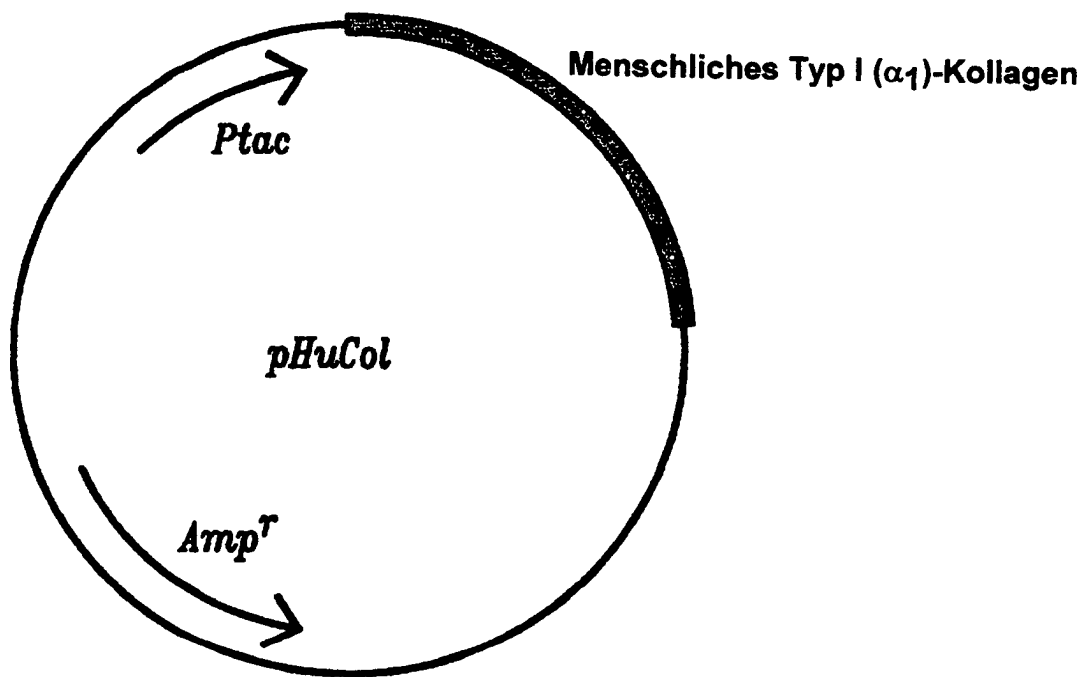


FIG. 4

5'- CAGCTGTCTT ATGGCTATGA TGAGAAATCA ACCGGAGGAA TTTCOGTGCC
TGGCCCCATG GGTOCCTCTG GTOCTCGTGG TCTOOCCTGGC CCCCCCTGGTG
CACCTGGTCC CCAAGGCTTC CAAGGTCCCC CTGGTGAGCC TGGGAGAGCCT
GGAGCTTCAG GTCCCATGGG TCCCCGAGGT CCCCCAGGTC CCCCTGGAAA
GAATGGAGAT GATGGGGAAG CTGGAAAACC TGGTCGTCCT-3'

FIG. 5

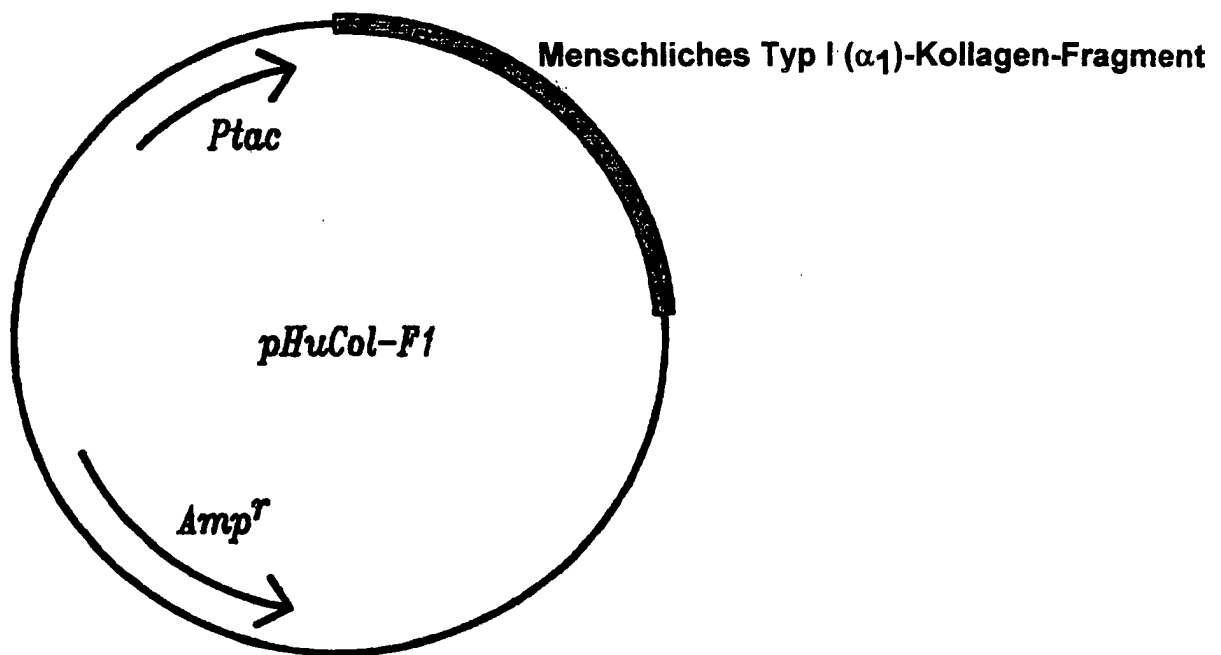


FIG. 6

GGA TCC ATG GGG CTC GCT GGC CCA CCG GGC GAA CCG GGT
CCG CCA GGC CCG AAA GGT CCG CGT GGC GAT AGC GGG CTC
CCG GGC GAT TCC TAA TGG ATC C

FIG. 7

Gly-Leu-Ala-Gly-Pro-Pro-Gly-Glu-Pro-Gly-Pro-Pro-
Gly-Pro-Lys-Gly-Pro-Arg-Gly-Asp-Ser

FIG. 8

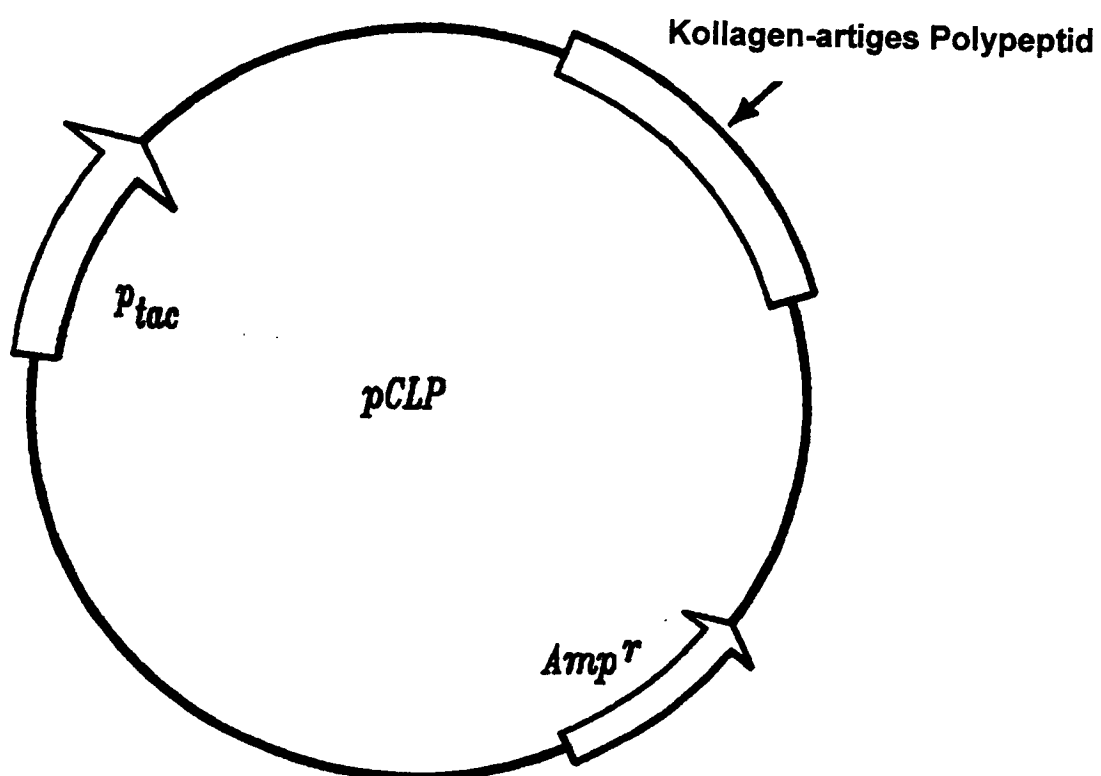


FIG. 9

5' - CAGCGGGCCA GGAAGAAGAA TAAGAACTGC CGGCGCCACT CGCTCTATGT
GGACTTCAGC GATGTGGGCT GGAATGACTG GATTGTGGCC CCACCAGGCT
ACCAGGCCTT CTA CTGCCAT GGGGACTGCC CCTTTCCACT GGCTGACCAC
CTCAACTCAA CCAACCATGC CATTGTGCAG ACCCTGGTCA ATTCTGTCAA
TTCCAGTATC CCCAAAGCCT GTTGTGTGCC CACTGAACTG AGTGCCATCT
CCATGCTGTA CCTGGATGAG TATGATAAGG TGGTACTGAA AAATTATCAG
GAGATGGTAG TAGAGGGATG TGGGTGCCGC -3'

FIG. 10

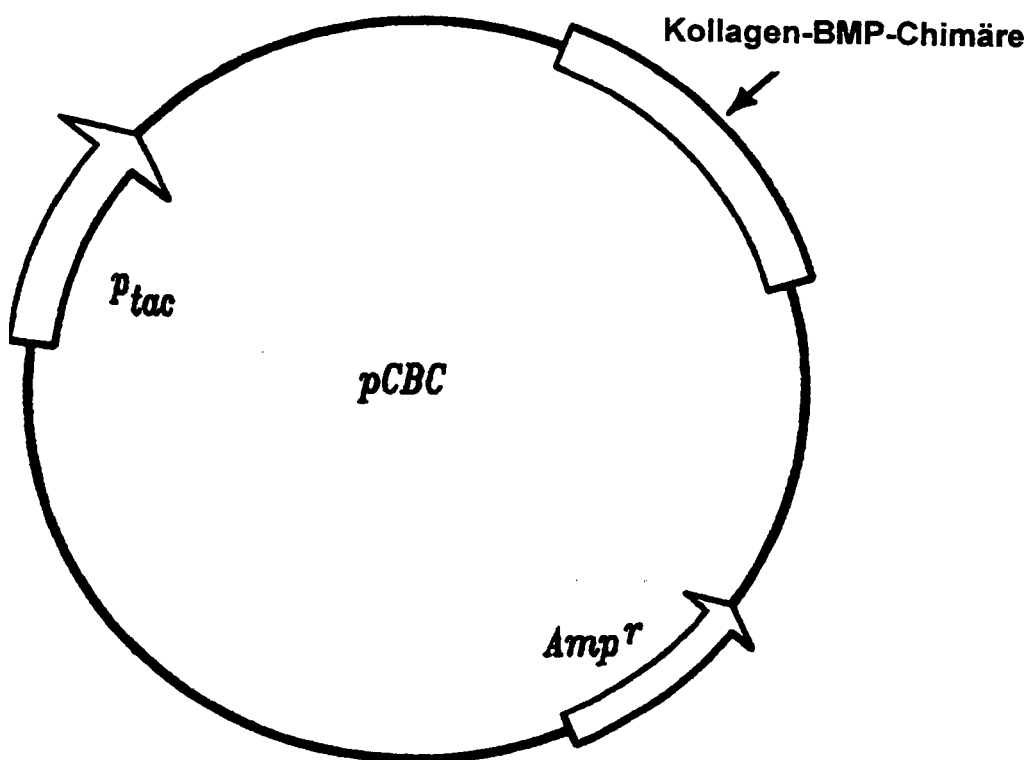


FIG. II

Prozentualer Einbau von Prolin
und Hydroxyprolin in Maltose-
bindendes Protein (MBP) unter
verschiedenen Bedingungen

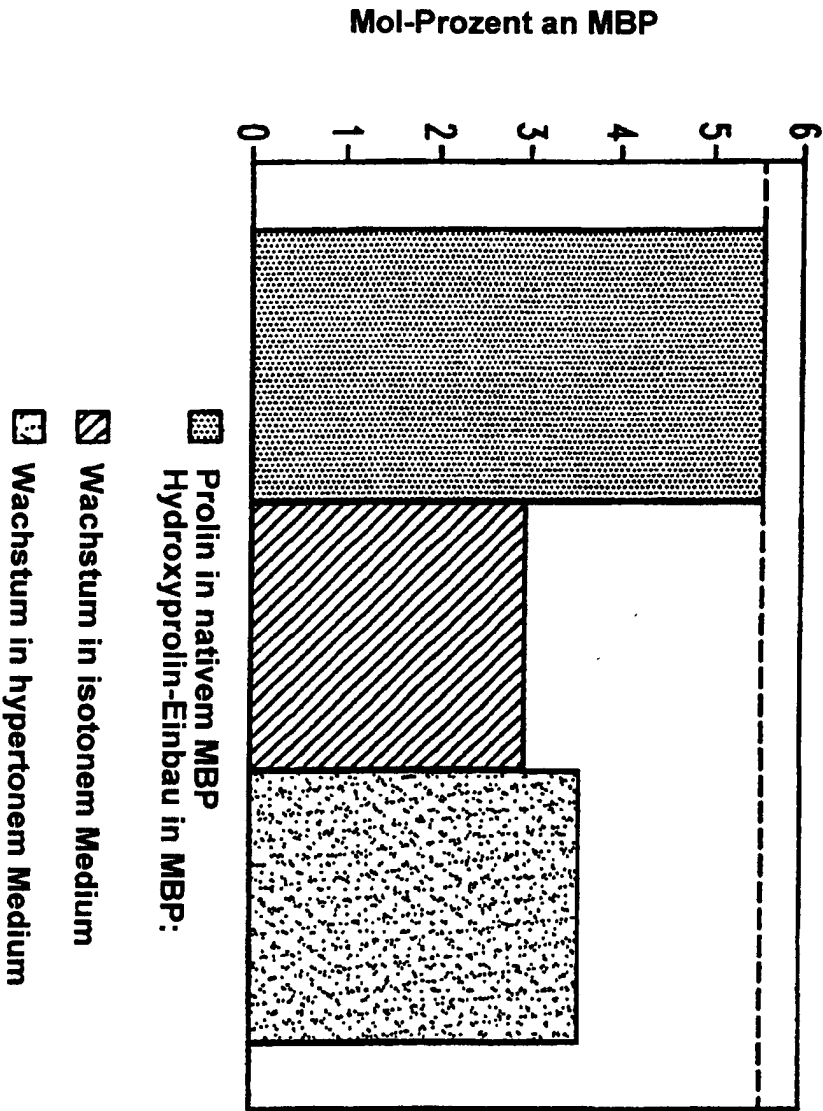


FIG. 12

10	20	30	40	50	60
QLSYGYDEKS	TGGISVPGPM	GPSGPRGLPG	PPGAPGPQGF	QGPPGEPGEP	GASGPMGPRG
70	80	90	100	110	120
PPGPPGXNGD	DGEAGKPGAP	GERGPPGPQG	ARGLPGTAGL	PGMKGHRGFS	GLDGAXGDAG
130	140	150	160	170	180
PAGPKGEPGS	PGENGAPGQM	GPRGLPGERG	RPGAPGPAGA	RQNDGATGAA	GPPGPTGPAG
190	200	210	220	230	240
PPGFPGAVGA	XGEAGPQGPR	GSEGPQGVRG	EPGPPGPAGA	AGPAGNPGAD	GQPGAKGANG
250	260	270	280	290	300
APGLIAGAPGF	PGARGPSGPQ	GPGGPPGPKG	NSGEPGAPGS	XGDTGAXGEP	GPVGVQGPFG
310	320	330	340	350	360
PAGEEGKXGA	RGEPPGTGLP	GPPGERGGPG	SRGFPGADGV	AGPKGPAGER	GSPGPAGPKG
370	380	390	400	410	420
SPGEAGRPE	AGLPKAKGLT	GSPGSPGPDG	XTGPPGPAGQ	DGRPGPPGPP	GARGQAGVMG
430	440	450	460	470	480
FFGPKGAAGE	PGKAGERGVF	GPPGAVGPAG	KDGEAGAQQP	PGPAGPAGER	GEQGPAGSPG
490	500	510	520	530	540
FQGLPGPAGP	PGEAGKPGEQ	GVPGDLGAPG	PSGARGERGF	PGERGVQGPP	GPAGPRGANG
550	560	570	580	590	600
APQNDGAKGD	AGAPGAPGSQ	GAPGLQGMPP	ERGAAGLPGP	KGDRGDAGPK	GADGSPGKDG
610	620	630	640	650	660
VRGLTGPIGP	PGPAGAPGDK	GESGPSGPAG	PTGARGAPGD	RGEPPGPPGA	GFAGPPGADG
670	680	690	700	710	720
QPGAKGEFGD	AGAKGDAGPP	GPAGPAGPPG	PIGNVCAPGA	XGARGSAGPP	GATGFPGAAG
730	740	750	760	770	780
RVGPPGPSGN	AGPPGPPGPA	GKGGGKGPBG	ETGPAGRPGE	VGPPGPPGPA	GEKGSFGADG
790	800	810	820	830	840
PAGAPGTPGP	QGLAQQRGVV	GLPGQRGERG	FPGLPGPSGE	PGKQGPPSGAS	GERGPPGPMG
850	860	870	880	890	900
PPGLAGPPGE	SGREGAPAAE	GSPGRDGSFG	AKGDRGETGP	AGPPGAXGAX	GAPGFVGPAG
910	920	930	940	950	960
KSGDRGETGP	AGPAGFVGPA	GARGPAGPQG	PRGDKGETGE	QGDRGIKGR	GFSGLQGPPG
970	980	990	1000	1010	1020
PPGSPGEQGP	SGASGPAGPR	GPPGSAGAPG	KDGLNGLPGP	IGPPGPRGRT	GDAGFVGPPG
1030	1040	1050	1060	1070	1080
PPGPPGPPGP	PSAGFDFSFL	PQPPQEKAYD	GGRYYRARSQ	RARQQNQNCR	RHSLYVDFSD
1090	1100	1110	1120	1130	1140
VGWNDWIVAP	PGYQAFYCHG	DCPFPLADHL	NSTNHAIIVQT	LVNSVNSSIP	KACCVFTELS
1150	1160	1170	1180	1190	1200
AISHLYLDEY	DKVVLKXNYQE	MVVEGCGCR*			

FIG. 13

```

      10      20      30      40      50      60
gggaaggatt tccatttccc AGCTGTCTTA TGGCTATGAT GAGAAATCPA CCGGAGGAAT

      70      80      90     100     110     120
TTCCGTGCCT GGGCCCATGG GTCCCTCTGG TCCTCGTGGT CTCCCTGGCC CCCCTGGTGC

      130     140     150     160     170     180
ACCTGGTCCC CAAGGCTTCC AAGGTCCCCC TGGTGAGCCT GGCAGCCTG GAGCTTCAGG

      190     200     210     220     230     240
TCCCATGGGT CCCCAGGGTC CCCCAGGTCC CCCTGATAAG AATGGAGATG ATGGGGAAGC

      250     260     270     280     290     300
TGGAAAACCT GGTCTCTCTG GTGAGCGTGG GCCTCTTGGG CCTCAGGGTG CTCGAGGATT

      310     320     330     340     350     360
GCCCCGAACA GCTGGCCTCC CTGGAATGAA GGGACACAGA GGTTCAGTG GTTTGGATGG

      370     380     390     400     410     420
TGCCAAGGGA GATGCTGGTC CTGCTGGTCC TAAGGATGAG CTTGGCAGCC CTGGTGAAAA

      430     440     450     460     470     480
TGGAGCTCCT GGTCAGATGG GCCCCGTGG CCTGCCTGGT GAGAGAGGTC GCCCTGGAGC

      490     500     510     520     530     540
CCCTGGCCCT GCTGGTGCTC GTGGAATGA TGGTGCTACT GGTGCTGCCG GCCCCCTGG

      550     560     570     580     590     600
TCCCACCGGC CCGCTGGTC CTCTGGCTT CCCTGATGCT GTTGGTGCTA AGGGTGAAGC

      610     620     630     640     650     660
TGGTCCCCAA GGGCCCCGAG GCTCTGAAGG TCCCCAGGGT GTGCGTGGTG AGCCTGGCCC

      670     680     690     700     710     720
CCCTGGCCCT GCTGGTGCTG CTGGCCCTGC TGGAAACCTT GGTGCTGATG GACAGCCTGG

      730     740     750     760     770     780
TGCTAAAGGT GCCAATGGTG CTCTGGTAT TGCTGATGCT CCTGGCTTCC CTGGTGCCCC

      790     800     810     820     830     840
AGGCCCTCTT GGACCCAGG GCCCCGGCGG CCCTCCTGGT CCAAGGGTA ACAGCGGTGA

      850     860     870     880     890     900
ACCTGGTGCT CCTGGCAGCA AAGGAGACAC TGGTGATAAG GGAGAGCCTG GCCCTGTTGG

      910     920     930     940     950     960
TGTTCAAGGA CCCCCTGGCC CTGCTGGAGA GGAAGATAAG CGAGGAGCTC GAGGTGAACC

      970     980     990    1000    1010    1020
CGGACCCACT GGCTGCCCCG GACCCCTGG CGAGCTGGGT GGACCTGGTA GCCGTGGTTT

      1030    1040    1050    1060    1070    1080
CCCTGGCGCA GATGGTGTTG CTGGTCCCAA GGGTCTCGCT GTGAACGTG GTTCTCTGG

      1090    1100    1110    1120    1130    1140
CCCCGCTGGC CCAAAGGAT CTCCTGGTGA AGCTGATCGT CCCGGTGAAG CTGGTCTGCC

      1150    1160    1170    1180    1190    1200
TGGTGCCAAG GGTCTGACTG GAAGCCCTGG CAGCCCTGGT CCTGATGGCA AAACCTGGCC

      1210    1220    1230    1240    1250    1260
CCCTGGTCCC GCGGTCAAG ATGGTCGCCC CGGACATCCA GGGCACCTG GTGCCCGTGG

```

FIG. 14A

```

1270      1280      1290      1300      1310      1320
TCAGGCTGGT GTGATGGGAT TCCCTGGACC TAAAGGTGCT GCTGGAGAGC CCGGCAAGGC

1330      1340      1350      1360      1370      1380
TGGAGAGCGA GGTGTTCCCG GACCCCTGGG CGCTGTCGGT CCTGCTGGCA AAGATGGAGA

1390      1400      1410      1420      1430      1440
GGCTGGAGCT CAGGGACCCC CTGGCCCTGC TGGTCCTCGCT GGCGAGAGAG GTGAACAAGG

1450      1460      1470      1480      1490      1500
CCCTGCTGGC TCCCCCGGAT TCCAGGGTCT CCCTGCTCCT GCTGCTCCTC CAGGTGAAGC

1510      1520      1530      1540      1550      1560
AGGCAAACCT GGTGAACAGG GTGTTCTTGG AGACCTTGGC CCCCCCTGGC CCTCTGGAGC

1570      1580      1590      1600      1610      1620
AAGAGGCGAG AGAGGTTTCC CTGGCGAGCG TGGTCTGCAA GGTCCCCCTG GTCTCTGCTG

1630      1640      1650      1660      1670      1680
ACCCCGAGGG GCCAACGGTG CTCCCGCAA CGATGGTGCT AAGGGTGATG CTGGTCCCCC

1690      1700      1710      1720      1730      1740
TGGAGCTCCC GGTAGCCAGG GCGCCCTGG CCTTCAGGGA ATGCCTGGTG AACGTGGTGC

1750      1760      1770      1780      1790      1800
AGCTGGTCTT CCAGGGCCTA AGGGTGACAG AGGTGATGCT GGTCCCAAAG GTGCTGATGG

1810      1820      1830      1840      1850      1860
CTCTCTGGC AAGATGGCG TCCGTGGTCT GACCGCCCC ATTGGTCTC CTGGCCCTGC

1870      1880      1890      1900      1910      1920
TGGTGCCCCC GTGACAAGG GTGAAAGTGG TCCCAACGGC CCTGCTGGTC CCACTGGAGC

1930      1940      1950      1960      1970      1980
TCGTGGTGCC CCCGAGACC GTGGTGAGCC TGGTCCCCCC GGCCCTGCTG GCTTTGCTGG

1990      2000      2010      2020      2030      2040
CCCCCTGGT GCTGACGGC AACCTGGTGC TAAAGGCGAA CCTGGTGATG CTGGTGCCAA

2050      2060      2070      2080      2090      2100
AGGCGATGCT GGTCCCCCTG GGCTGCGCG ACCCGCTGGA CCCCCTGGCC CCATTGGTAA

2110      2120      2130      2140      2150      2160
TGTGGTGCT CCTGGAGCCA AAGGTGCTCG CGGCAGCGCT GGTCCCCCTG GTGCTACTGG

2170      2180      2190      2200      2210      2220
TTTCCCTGGT GCTGCTGGCC GAGTCGGTCC TCCTGSCCCC TCTGGAATG CTGGACCCCC

2230      2240      2250      2260      2270      2280
TGCCCCCTCT GTTCTGCTG GCAAGAAGG CGGCAAGGT CCCCCTGGTG AGACTGGCCC

2290      2300      2310      2320      2330      2340
TGCTGGACGT CCTGGTGAAG TTGGTCCCCC TGGTCCCCCT GGCCCTGCTG GCGAGAAAGG

2350      2360      2370      2380      2390      2400
ATCCCCCTGGT GCTGATGGTC CTGCTGGTGC TCCTGTACT CCCGGGCTC AAGGTATTGC

2410      2420      2430      2440      2450      2460
TGGACAGCGT GGTGTGGTCG GCCTGCCTGG TCAGAGAGGA GAGAGAGGCT TCCCTGGTCT

2470      2480      2490      2500      2510      2520
TCCTGGCCCC TCTGGTGAAC CTGGCAAACA AGGTCCCTCT GGAGCAAGTG GTGAACGTGG

```

FIG. 14B

2530	2540	2550	2560	2570	2580
TCCCCCCGGT	CCCATGGGCC	CCCCTGGATT	GGCTGGACCC	CCTGGTGAAT	CTGCACGTGA
2590	2600	2610	2620	2630	2640
GGGGGCTCCT	GCTGCCGAAG	GTTCCCCTGG	ACGAGACGGT	TCTCCTGGCG	CCAAGGGTGA
2650	2660	2670	2680	2690	2700
CCGTGGTGAG	ACCGGCCCCG	CTGGACCCCC	TGGTGCTCCT	GGTGCTCCTG	GTGCCCCCTGG
2710	2720	2730	2740	2750	2760
CCCCGTTGGC	CCTGCTGGCA	AGAGTGGTGA	TCGTGGTGAG	ACTGGTCCCTG	CTGGTCCCCG
2770	2780	2790	2800	2810	2820
CGGTCCCGTC	GGCCCCGCTG	GCGCCCGTGG	CCCCGCCCGA	CCCCAAGGCC	CCCGTGGTGA
2830	2840	2850	2860	2870	2880
CAAGGGTGAG	ACAGGCGAAC	AGGGCGACAG	AGGCATAAAG	GGTCACCGTG	GCTTCTCTGG
2890	2900	2910	2920	2930	2940
CCTCCAGGGT	CCCCCTGGCC	CTCCTGGCTC	TCCTGGTGAA	CAAGGTCCCT	CTGGAGCCTC
2950	2960	2970	2980	2990	3000
TGGTCTTGCT	GGTCCCCGAG	GTCCCCCTGG	CTCTGCTGGT	GCTCCTGGCA	AAGATGGACT
3010	3020	3030	3040	3050	3060
CAACGGTCTC	CCTGGCCCCA	TTGGGCCCCC	TGGTCTCGC	GGTCGCACTG	GTGATGCTGG
3070	3080	3090	3100	3110	3120
TCCTGTTGGT	CCCCCGGCC	CTCCTGGACC	TCCTGGTCCC	CCTGGTCCCTC	CCAGCGCTGG
3130	3140	3150	3160	3170	3180
TTTCGACTTC	AGCTTCCTCC	CCCAGCCACC	TCAAGAGAAG	GCTCAGGATG	GTGGCCCGCTA
3190	3200	3210	3220	3230	3240
CTACCGGGCT	agatccCAGC	GGGCCAGGAA	GAAGAATAAG	AATGCCGGC	GCCACTCGCT
3250	3260	3270	3280	3290	3300
CTATGTGGAC	TTCAGCGATG	TGGGCTGGAA	TGACTGGATT	GTGGCCCCAC	CAGGCTACCA
3310	3320	3330	3340	3350	3360
GGCCTTCTAC	TGCCATGGGG	ACTGCCCTTT	TCCACTGGCT	GACCACCTCA	ACTCAACCAA
3370	3380	3390	3400	3410	3420
CCATGCCATT	GTGCAGACCC	TGGTCAATTG	TGTCAATTCC	AGTATCCCCA	AAGCCTGTTG
3430	3440	3450	3460	3470	3480
TGTGCCCACT	GAACTGAGTG	CCATCTCCAT	GCTGTACCTG	GATGAGTATG	ATAAGGTGGT
3490	3500	3510	3520	3530	3540
ACTGAAAAAT	TATCAGGAGA	TGGTAGTAGA	GGGATGTGGG	TGCCGCTAAa	agcctt.....

FIG. 14C

10	20	30	40	50	60
QLSYGYDEKS	TGGISVPGPH	GPSGPRGLPG	PPGAPGPOGF	QGPFGPECEP	GASGPMGPRG
70	80	90	100	110	120
PPGPPGKNGD	DGEAGKPGRP	GERGPPGPQG	ARGLPGTAGL	PGMKGHRGFS	GLDGAKGDAG
130	140	150	160	170	180
PAGPKGEPS	PGENGAPQM	GPRGLPGERG	RPGAPGPAGA	RQNDGATGAA	GPPGPTGPAG
190	200	210	220	230	240
PPGFPGAVGA	KGEAGPQGPR	GSEGPQGVRG	EPGPPGPAGA	AGPAGNPGAD	GQPGAKGANG
250	260	270	280	290	300
APGLAGAPGF	PGARGPSGPQ	GPPGPPGPKG	NSGEPGAPGS	KGDTGAKGEP	GPVGVQGPFG
310	320	330	340	350	360
PAGEEGKRG	RGEFGPTGLP	GPPGERGGPG	TRGFPGADGV	AGPKGPAGER	GSPGPAGPKG
370	380	390	400	410	420
SPGEAGRPE	AGLPGAKGLT	GSPGSPGPDG	KTGPPGPAGQ	DGRPGPPGPP	GARGQAGVMG
430	440	450	460	470	480
FPGPKGAAGE	PGKAGERGVP	GPPGAVGPAG	KDGEAGAQQP	PGPAGPAGER	GEQGPAGSPG
490	500	510	520	530	540
FOGLPGPAGP	PGEAGKPGEQ	GVPGLQAPG	PSGARGERG	PGERGVOGPP	GPAGPRGANG
550	560	570	580	590	600
APQNDGAKGD	AGAPGAPGSQ	GAPGLQMPG	ERGAAGLPGP	KGDRGDAGPK	GADGSFGKDG
610	620	630	640	650	660
VRGLTGPIGP	PGPAGAPGDK	GESGSPGPAG	PTGARGAPGD	RGEFGPPGPA	GTAGPPGADG
670	680	690	700	710	720
QPGAKGEPGD	AGAKGDAGFP	GPAGPAGPPG	PIQNVGAPGA	KGARGSAGPP	GATGFFGAAG
730	740	750	760	770	780
RVGPFGPSGN	AGPPGPPGPA	GKEGGKGPGR	ETGPAGRPE	VGPPGPPGPA	GEXGSPGADG
790	800	810	820	830	840
PAGAPGTGPF	QGLAGQRGVV	GLPGQRGERG	FPGLFGPSGE	PGKQGPSGAS	GERGPPGFMG
850	860	870	880	890	900
PPGLAGPPGE	SGREGAPAAE	GSPGRDGSFG	AKGDRGETGP	AGPPGAXGAX	GAPGFVGPAG
910	920	930	940	950	960
KSGDRGETGP	AGPAGFVGPA	GARGPAGPQG	PRGDKGETGE	QGDRGIKGRH	GFSGLQGPPG
970	980	990	1000	1010	1020
PPGSPGEQGP	SGASGPAGPR	GPPGSAGAPG	KDGLNGLPGP	IGPPGPRGRT	GDAGFVGPFG
1030	1040	1050	1060	1070	1080
PPGPPGPPGP	PSAGFDFSFL	PQPPQEKAMD	GGRYYRARS	LDNYCFSST	EKNCCVRQLY
1090	1100	1110	1120	1130	1140
IDFRKDLGWK	WIHEPKGYHA	NFCLGPCPYI	WSLDTQYSKV	LALYNQHNPG	ASAAPCCVPO
1150	1160	1170	1180	1190	1200
ALEPLPIVYY	VGRKPKVEQL	SNMIVRSCXC	S*		

FIG. 15

```

10      20      30      40      50      60
gggaaggatt tccatttccc AGCTGTCTTA TGGCTATGAT GAGAAATCAA CCGGAGGAAT

70      80      90      100     110     120
TTCCGTGCTT GGGCCCATGG GTCCCTCTGG TCCTCGTGGT CTCCCTGGCC CCCCTGGTGC

130     140     150     160     170     180
ACCTGGTCCC CAAGGCTTCC AAGGTCCCCC TGGTGAGCCT GCGGAGCCTG GAGCTTCAGG

190     200     210     220     230     240
TCCCATGGGT CCCCAGGTC CCCCAGGTCC CCCTGMAAAG AATGGAGATG ATGGGGAAGC

250     260     270     280     290     300
TGGAACCTT GGTCTGCTTG GTGAGCGTGG GCCTCTTGGG CCTCAGGGTG CTCGAGGATT

310     320     330     340     350     360
GCCCCGAACA GCTGGCCTCC CTGGAATGAA GGGACACAGA GGTTCAGTG GTTTGGATGG

370     380     390     400     410     420
TGCCAAGGGA GATGCTGGTC CTGCTGGTCC TAAGGATGAG CCTGGCAGCC CTGGTGAAAA

430     440     450     460     470     480
TGGAGCTCCT GGTGAGATGG GCCCCCGTGG CCTGCCTGGT GAGAGAGGTC GCCCTGGAGC

490     500     510     520     530     540
CCCTGGCCCT GCTGGTGCTC GTGGAATGA TGGTGCTACT GGTGCTGCCG GGCECCCTGG

550     560     570     580     590     600
TCCCACCGGC CCCGCTGGTC CTCCTGGCTT CCCTGATGCT GATGGTGCTA AGGGTGAAAGC

610     620     630     640     650     660
TGGTCCCCAA GGGCCCCGAG GCTCTGAAGG TCCCAGGGT GTGCGTGGTG AGCCTGGCCC

670     680     690     700     710     720
CCCTGGCCCT GCTGGTGCTG CTGGCCCTGC TGGAACCCCT GGTGCTGATG GACAGCCTGG

730     740     750     760     770     780
TGCTAAAGGT GCCAATGGTG CTCCTGGTAT TGCTGGTGCT CCTGGCTTCC CTGGTGCCCCG

790     800     810     820     830     840
AGGCCCCCTCT GGACCCCAGG GCCCCGCGG CCTCCTGGT CCCAAGGGTA ACAGCGGTGA

850     860     870     880     890     900
ACCTGGTGCT CCTGGCAGCA AAGGAGACAC TGGTGCTAAG GGAGAGCCTG GCCCTGTGGG

910     920     930     940     950     960
TGTTCAAGGA CCCCCTGGCC CTGCTGGAGA GGAAGGAAAG CGAGGAGCTC GAGGTGAACC

970     980     990     1000    1010    1020
CGGACCCACT GGCCTGCCCC GACCCCTGG CGAGCGTGGT GGACCTGGTA GCCGTGGTTT

1030    1040    1050    1060    1070    1080
CCCTGGCGCA GATGGTGTTG CTGGTCCCAA GGGTCCCGCT GGTGAACGTG GTTCTCTCTGG

1090    1100    1110    1120    1130    1140
CCCCGCTGGC CCCAAGGAT CTCCTGGTGA AGCTGGTCGT CCCGGTGAAG CTGGTCTGCC

1150    1160    1170    1180    1190    1200
TGGTGCCAAG GGTCTGACTG GAAGCCCTGG CAGCCCTGGT CCTGATGGCA AAACCTGGCCC

1210    1220    1230    1240    1250    1260
CCCTGGTCCC GCGGCTCAAG ATGGTCGCCC CGGACCCCA GGCCCACTG GTGCCCCGTGG

```

FIG. 16A

1270	1280	1290	1300	1310	1320
TCAGGCTGGT	GTGATGGGAT	TCCCTGGACC	TAAAGGTGCT	GCTGGAGAGC	CCGGCAAGGC
1330	1340	1350	1360	1370	1380
TGGAGAGCGA	GGTGTTCCTG	GACCCCTGG	CGCTGTCGGT	CCTGCTGGCA	AAGATGGAGA
1390	1400	1410	1420	1430	1440
GGCTGGAGCT	CAGGGACCCC	CTGGCCCTGC	TGGTCCCGCT	GGCGAGAGAG	GTGAACAAGG
1450	1460	1470	1480	1490	1500
CCCTGCTGGC	TCCCCCGGAT	TCCAGGGTCT	CCCTGGTCCT	GCTGGTCCTC	CAGGTGAAGC
1510	1520	1530	1540	1550	1560
AGGCCAAACCT	GGTGAACAGG	GTGTTCTCTG	AGACCTTGGC	GCCCCCTGGC	CCTCTGGAGC
1570	1580	1590	1600	1610	1620
AAGAGGCGAG	AGAGGTTTCC	CTGGCGAGCG	TGGTGTGCAA	GGTCCCCCTG	GTCTGCTGG
1630	1640	1650	1660	1670	1680
ACCCCGAGGG	GCCAACGGTG	CTCCCGGCAA	CGATGGTGCT	AAGGGTGATG	CTGGTGCCCC
1690	1700	1710	1720	1730	1740
TGGAGCTCCC	GGTAGCCAGG	GCGCCCTGG	CCTTCAGGGA	ATGCCTGGTG	AACGTGGTGC
1750	1760	1770	1780	1790	1800
AGCTGGTCTT	CCAGGGCCTA	AGGGTGACAG	AGGTGATGCT	GGTCCCAAAG	GTGCTGATGG
1810	1820	1830	1840	1850	1860
CTCTCCTGGC	AAAGATGGCG	TCCGTGGTCT	GACCGACCCC	ATTGGTCTCT	CTGGCCCTGC
1870	1880	1890	1900	1910	1920
TGGTGCCCCCT	GGTGACAAGG	GTGAAAGTGG	TCCAGACGGC	CCTGCTGGTC	CCACTGGAGC
1930	1940	1950	1960	1970	1980
TGGTGGTGCC	CCCGGAGACC	GTGGTGAGCC	TGGTCCCCC	GGCCCTGCTG	GCTTTGCTGG
1990	2000	2010	2020	2030	2040
CCCCCTGGT	GCTGACGGCC	AACCTGGTGC	TAAAGGCGAA	CCTGGTGATG	CTGGTGCCAA
2050	2060	2070	2080	2090	2100
AGGCGATGCT	GGTCCCCCTG	GGCCTGCCGG	ACCCGCTGGA	CCCCCTGGCC	CCATTGGTAA
2110	2120	2130	2140	2150	2160
TGTTGGTGCT	CCTGGAGCCA	AAGGTGCTCG	CGGCAACGCT	GGTCCCCCTG	GTGCTACTGG
2170	2180	2190	2200	2210	2220
TTTCCCTGGT	GCTGCTGGCC	GAGTCGGTCC	TCCTGSCCCC	TCTGGAATG	CTGGACCCCC
2230	2240	2250	2260	2270	2280
TGGCCCTCCT	GGTCTGCTG	GCAAGAAGG	CGGCAAGGT	CCCCGTGGTG	AGACTGGCCC
2290	2300	2310	2320	2330	2340
TGCTGGACGT	CCTGGTGAAG	TTGGTCCCCC	TGGTCCCCCT	GGCCCTGCTG	GCGAGAAAGG
2350	2360	2370	2380	2390	2400
ATCCCCCTGGT	GCTGATGGTC	CTGCTGGTGC	TCCTGTACT	CCCGGGCTC	AAGGTATTGC
2410	2420	2430	2440	2450	2460
TGGACAGCGT	GGTGTGGTGC	GCCTGCCTGG	TCAGAGAGGA	GAGAGAGGCT	TCCCTGGTCT
2470	2480	2490	2500	2510	2520
TCCTGGCCCC	TCTGGTGAAC	CTGGCAAACA	AGGTCTCTCT	GGAGCAAGTG	GTGAACGTGG

FIG. 16B


```

      2530      2540      2550      2560      2570      2580
TCCCCCGGT CCCATGGGCC CCCCTGGATT GGCTGGACCC CCTGGTGAAT CTGGACGTGA

      2590      2600      2610      2620      2630      2640
GGGGCTCCT GCTGCCGAAG GTCCCCCTGG ACGAGACGGT TCTCCTGGCG CCAAGGGTGA

      2650      2660      2670      2680      2690      2700
CCGTGGTGAG ACCGGCCCCG CTGGACCCCC TGGTGCTCCT GGTGCTCCTG GTGCCCTGG

      2710      2720      2730      2740      2750      2760
CCCCGTGGC CCTGCTGGCA AGAGTGGTGA TCGTGGTGAG ACTGGTCCTG CTGGTCCCGC

      2770      2780      2790      2800      2810      2820
CGGTCCCCGTC GGGCCCGCTG GCGCCCGTGG CCCCCGCGGA CCCCAGGCC CCCGTGGTGA

      2830      2840      2850      2860      2870      2880
CAAGGGTGAG ACAGGCGAAC AGGGCGACAG AGGCA?AAAG GGTACCGTG GCTTCTCTGG

      2890      2900      2910      2920      2930      2940
CCTCCAGGGT CCCCCTGGCC CTCCTGGCTC TCCTGGTGAA CAAGGTCCCT CTGGAGCCTC

      2950      2960      2970      2980      2990      3000
TGGTCCTGCT GGTCCTCGAG GTCCCCCTGG CTCCTGCTGGT GCTCCTGGCA AAGATGGACT

      3010      3020      3030      3040      3050      3060
CAACGGTCTC CCTGGCCCCA TTGGGCCCCC TGGTCCTCGC GGTCCGACTG GTGATGCTGG

      3070      3080      3090      3100      3110      3120
TCCTGTTGCT CCCCCCGGCC CTCCTGGACC TCCTG?TCCC CCTGGTCCTC CCAGCCTGG

      3130      3140      3150      3160      3170      3180
TTTCGACTTC AGCTTCCTCC CCCAGCCACC TCAAG/GAAG GCTCAGCATG GTGGCCGCTA

      3190      3200      3210      3220      3230      3240
CTACCGGGCT agatctGCCC TGGACACCAA CTATTGCTTC AGCTCCACGG AGAAGAACTG

      3250      3260      3270      3280      3290      3300
CTGCGTGCGG CAGCTGTACA TTGACTTCCG CAAGGACCTC GGCTGGAAGT GGATCCACGA

      3310      3320      3330      3340      3350      3360
GCCCAAGGGC TACCATGCCA ACTTCTGCCT CGGGCCCTGC CCCTACATTT GGAGCCTGGA

      3370      3380      3390      3400      3410      3420
CAGCAGTAC AGCAAGGTCC TGGCCCTGTA CAACCAGCAT AACCCGGGCG CCTCGGCGGC

      3430      3440      3450      3460      3470      3480
GCCGTGCTGC GTGCCGCAAG CGCTGGAGCC GCTGCCCATC GTGTACTACG TGGGCCGCAA

      3490      3500      3510      3520      3530      3540
GCCCAAGGTG GAGCA?CTGT CCAACATGAT CGTGCGCTCC TGCAAGTGCA GCTGatctag

      3550      3560      3570      3580      3590      3600
a.....

```

FIG. 16C

10	20	30	40	50	60
QLSYGYDEKS	TGGISVPGPM	GPSGPRGLPG	PPGAPGPQGF	QGPPEGPGEF	GASGPMGPGR
70	80	90	100	110	120
PPGPPGKQGD	DGEAGKPRP	GERGPPGPQG	ARGLPGTAGL	PGMKGHRGFS	GLDGAKGDAG
130	140	150	160	170	180
PAGPKGEFGS	PGENGAPQOM	GPRGLPGERG	RPGAPGPAGA	RGNDGATGAA	GPPGPTGPAG
190	200	210	220	230	240
PPGFPAGVGA	KGEAGPQGP	GSEGPQGVRC	EPGPPGPAGA	AGPAGNPGAD	GQPKGAKGANG
250	260	270	280	290	300
APGLAGAPGF	PGARGPSGPQ	GPGGPPGPKG	NSGEPGAPGS	KGDTGAKGEP	GPVGVQGPFG
310	320	330	340	350	360
PAGEEGKRG	RGEPTGLP	GPPGERGGPG	SRGFFGADGV	AGPKGPAGER	GSPGPAGPKG
370	380	390	400	410	420
SPGEAGRGE	AGLPGAKGLT	GSPGSPGPDG	KTGPPGPAGQ	DGRPGPPGPP	GARGQAGVMG
430	440	450	460	470	480
FPGPKGAAGE	PGKAGERGVP	GPPGAVGPAG	KDGEAGAQQP	PGPAGPAGER	GEQGPAGSPG
490	500	510	520	530	540
FQGLPGPAGP	PGEAGKPGEQ	GVPGDLGAPG	PSGARGERGF	PGERGVQGP	GPAGPRGANG
550	560	570	580	590	600
APGNDGAKGD	AGAPGAPGSQ	GAPGLQGMPP	ERGAAGLPGP	KGDRGDAGPK	GADGSPGKDG
610	620	630	640	650	660
VRGLTGPIGP	PGPAGAPGDK	GESGSPGPAG	PTGARZAPGD	RGEPPGPPGA	GFAGPPGADG
670	680	690	700	710	720
QPGAKGEFGD	AGAKGDAGPP	GPAGPAGPPG	PIQNVZAPGA	KGARGSAGPP	GATGPPGAAG
730	740	750	760	770	780
RVGPPGPGSN	AGPPGPPGPA	GKEGGKGPGR	ETGPAGRGE	VGPPGPPGPA	GEMGSPGADG
790	800	810	820	830	840
PAGAPGTGPP	QGIAGQRGVV	GLPGQRGERG	FPGLPGPSGE	PGKQSPSGAS	GERGPPGPMG
850	860	870	880	890	900
PPGLAGPPGE	SGREGAPAAE	GSPGRDGSFG	AKGDRGETGP	AGPPGAXGAX	GAPGVPGPAG
910	920	930	940	950	960
KSGDRGETGP	AGPAGPVGPA	GARGPAGPQG	PRGDKGETGE	QGDRGIKGR	GFSGLQGPPG
970	980	990	1000	1010	1020
PPGSPGEQGP	SGASGPAGPR	GPPGSAGAPG	KDGLNGLPGP	IGPPGPRGRT	GDAGPVGPPG
1030	1040	1050	1060	1070	1080
PPGPPGPPGP	PSAGFDFSFL	PQPPQEXAHD	GGRYYRARS	EASGIGPEVP	DDRDFEPSLG
1090	1100	1110	1120	1130	1140
PVCPFRQCH	LRVVQCSLG	LDXVPKDLPP	DTLLDLQNN	KITEIKDGD	KNLNLMALI
1150	1160	1170	1180	1190	1200
LWNNKISKVS	PGFTPLVXL	ERLYLSKNQL	KELPEXPKT	LQELRAHENE	ITKVRKVTFN
1210	1220	1230	1240	1250	1260
GLNQMIIVIEL	GTNPLKSSCI	ENGAPQMKK	LSYIRIADTN	ITSIPQGLPP	SLTELHLDGN

FIG. 17A

1270	1280	1290	1300	1310	1320
XISRVDAAASL	KGLNNLAKLG	LSFNSISAVD	NGSLANTPHL	RELHLENNKL	TRVPGGLAEH
1330	1340	1350	1360	1370	1380
KYIQVVYLHN	NNISVVGSSD	FCPPGHNTKK	ASYSGVSLFS	NPVQYWEIQP	STFRCVYVRS
1390	1400	1410	1420	1430	1440
AIQLGNYK*					

FIG. 17B

10	20	30	40	50	60
QLSYGYDEKS	TGGISVPGPM	GPSGPRGLPG	PPGAPGPQGF	QGPPGCEPGE	GASGPMGPRG
70	80	90	100	110	120
PPGPPGKNGD	DGEAGKPRP	GERGPPGPOG	ARGLPGTAGL	PGMKGHRGFS	GLDCAKGDAG
130	140	150	160	170	180
PAGPKGEFGS	PGENGAPGQM	GPRGLPGERG	RPGAPGPAGA	RGNDGATGAA	GPPGPTGPAG
190	200	210	220	230	240
PPGFPGAVGA	KGEAGPQGP	GSEGPQGVRG	EPGPPGPAGA	AGPAGNPGAD	GQPGAKGANG
250	260	270	280	290	300
APGLAGAPGF	PGARGPSGPQ	GPGGPPGPKG	NSGEPGAPGS	KGDTGAKGEP	GPVGVQGPPG
310	320	330	340	350	360
PAGEEGKRG	RGEFGPTGLP	GPPGERGGPG	SRGFPGADGV	AGPKGPAGER	GSPGPAGPKG
370	380	390	400	410	420
SPGEAGRPE	AGLPGAKGLT	GSPGSPGPDG	KTGPPGPAGQ	DGRPGPPGPP	GARGQAGVMG
430	440	450	460	470	480
PPGPKGAAGE	PGKAGERGV	GPPGAVGPAG	KDGEAGAQQP	PGPAGPAGER	GEQGPAGSPG
490	500	510	520	530	540
FQGLPGPAGP	PGEAGKPEQ	GVPGLGAPG	PSGARGERGF	PGERGVQGP	GPAGPRGANG
550	560	570	580	590	600
APGNDGAKGD	AGAPGAPGQ	GAPGLQGMFG	ERGAAGLPGP	KGDRGDAGPK	GADGSPGKDG
610	620	630	640	650	660
VRGLTGPIGP	PGPAGAPGDX	GESGSPGPAG	PTGARGAPGD	RGEPPGPPGA	GFAGPPGADG
670	680	690	700	710	720
QPGAKGEFGD	AGAKGDAGPP	GPAGPAGPPG	PIGNVGAAGA	KGARGSAGPP	GATGFFGAAG
730	740	750	760	770	780
RVGPPGPGSN	AGPPGPPGPA	GKEGCKGPRG	ETGPAGRPE	VGPPGPPGPA	GEKGSFGADG
790	800	810	820	830	840
PAGAPGTGPP	QGLAQGRGVV	GLPGQRGERG	FPGLPGPSGE	PGKQGPGSGS	GERGPPGPMG
850	860	870	880	890	900
PPGLAGPPGE	SGREGAPAAE	GSPGRDGSPG	AKGDRGETGP	AGPPGAXGAX	GAPGPPGPPG
910	920	930	940	950	960
KSGDRGETGP	AGPAGFVGPA	GARGPAGPQG	PRGDKGETGE	QGDRGIKGR	GFSGLQGP
970	980	990	1000	1010	1020
PPGSPGEQGP	SGASGPAGPR	GPPGSAGAPG	KDGLNGLPGP	IGPPGPRGRT	GDAGFVGPPG
1030	1040	1050	1060	1070	1080
PPGPPGPPGP	PSAGFDFSL	PQPPQEKAMD	GGRYYRARS	KDLPPDTLL	DLQNNKITEI
1090	1100	1110	1120	1130	1140
KDGDFFKLN	LHALILVNNK	ISKVSPG*			

FIG. 18

9	18	27	36	45	54
CAG CTG TCT TAT GGC TAT GAT GAG AAA TCA ACC CGA GGA ATT TCC GTG CCT GGC CCC ATG					
69	78	87	96	105	114
GGT CCC TCT GGT CCT CGT GGT CTC CCT GGC CCC CCT GGT GCA CCT GGT CCC CAA GGC TTC					
129	138	147	156	165	174
CAA GGT CCC CCT GGT GAG CCT GGC GAG CCT GGA GCT TCA GGT CCC ATG GGT CCC CGA GGT					
189	198	207	216	225	234
CCC CCA GGT CCC CCT GGA AAG AAT GGA GAT GAT GGG GAA GCT GGA AAA CCT GGT CGT CCT					
249	258	267	276	285	294
GGT GAG CGT GGG CCT CCT GAG CCT CAG GGT GCT CGA GGA TTG CCC GGA ACA GCT GGC CTC					
309	318	327	336	345	354
CCT GGA ATG AAG GGA CAC AGA GGT TTC AGT GGT TTG GAT GGT GCC AAG GGA GAT GCT GGT					
369	378	387	396	405	414
CCT GCT GGT CCT AAG GGT GAG CCT GGC AGC CCT GGT GAA AAT GGA GCT CCT GGT CAG ATG					
429	438	447	456	465	474
GGC CCC CGT GGC CTG CCT GGT GAG AGA GGT GGC CCT GGA GCC CCT GGC CCT GCT GGT GCT					
489	498	507	516	525	534
CGT GGA AAT GAT GGT GCT ACT GGT GCT GCC GGG CCC CCT GGT CCC ACC GGC CCC GCT GGT					
549	558	567	576	585	594
CCT CCT GGC TTC CCT GGT GCT GTT GGT GCT AAG GGT GAA GCT GGT CCC CAA GGG CCC CGA					
609	618	627	636	645	654
GGC TCT GAA GGT CCC CAG GGT GTG CGT GGT GAG CCT GGC CCC CCT GGC CCT GCT GGT GCT					
669	678	687	696	705	714
GCT GGC CCT GCT GGA AAC CCT GGT GCT GAT GGA CAG CCT GGT GCT AAA GGT GCC AAT GGT					
729	738	747	756	765	774
GCT CCT GGT ATT GCT CGT GCT CCT GGC TTC CCT GGT GCC CGA GGC CCC TCT GGA CCC CAG					
789	798	807	816	825	834
GGC CCC GGC GGC CCT CCT GGT CCC AAG GGT AAC AGC GGT GAA CCT GGT GCT CCT GGC AGC					
849	858	867	876	885	894
AAA CGA GAC ACT GGT GCT AAG GGA GAG CCT GGC CCT GTT GGT GTT CAA GGA CCC CCT GGC					
909	918	927	936	945	954
CCT GCT GGA GAG CAA GGA AAG CGA GGA GCT CGA GGT GAA CCC GGA CCC ACT GGC CTG CCC					
969	978	987	996	1005	1014
GGA CCC CCT GGC GAG CGT GGT GGA CCT GGT AGC CGT GGT TTC CCT GGC GCA GAT GGT GTT					
1029	1038	1047	1056	1065	1074
GCT GGT CCC AAG GGT CCC GCT GGT GAA CGT GGT TCT CCT GGC CCC GCT GGC CCC AAA GGA					
1089	1098	1107	1116	1125	1134
TCT GGT CGT GAA GCT GGT CGT CCC GGT GAA GCT GGT CTG CCT GGT GCC AAG GGT CTG ACT					
1149	1158	1167	1176	1185	1194
GGA AGC CCT GGC AGC CCT GGT CCT GAT GGC AAA ACT GGC CCC CCT GGT CCC GCC GGT CAA					
1209	1218	1227	1236	1245	1254
GAT GGT GGC CCC CGA CCC CCA GGC CCA CCT GGT GCC CGT GGT CAG GCT GGT GTG ATG GGA					

FIG. 19A

1269 1278 1287 1296 1305 1314
 TTC CCT GGA CTT AAA GGT GGT GGT GGA GAG CCC CGC AAG GCT GGA GAG CGA GGT GTT CCC
 1329 1338 1347 1356 1365 1374
 GGA CCC CTT GGC GGT GTC GGT CCT GCT GCC AAA GAT GGA GAG GCT GGA GCT CAG GGA CCC
 1389 1398 1407 1416 1425 1434
 CCT GGC CCT GCT GGT CCC GCT GGC GAG AGA GGT GAA CAA GGC CCT GCT GGC TCC CCC GGA
 1449 1458 1467 1476 1485 1494
 TTC CAG GGT CTC CCT GGT CCT GCT GGT CCT CCA GGT GAA GCA GGC AAA CCT GGT GAA CAG
 1509 1518 1527 1536 1545 1554
 GGT GTT CCT GGA GAC CTT GGC GCC CTT GGC CCC TCT GGA GCA AGA GGC GAG AGA GGT TTC
 1569 1578 1587 1596 1605 1614
 CCT GGC GAG CGT GGT GTG CAA GGT CCC CCT GGT CCT GCT GGA CCC CGA GGC GCC AAC GGT
 1629 1638 1647 1656 1665 1674
 GCT CCC GGC AAC GAT GGT GCT AAG GGT GAT GCT GGT GGC CCT GGA GCT CCC GGT AGC CAG
 1689 1698 1707 1716 1725 1734
 GGC GCC CCT GGC CTT CAG GGA ATG CCT GGT GAA CGT GGT GCA GCT GGT CTT CCA GGC CCT
 1749 1758 1767 1776 1785 1794
 AAG GGT GAC AGA GGT GAT GCT GGT CCC AAA GGT CTT GAT GGC TCT CCT GGC AAA GAT GGC
 1809 1818 1827 1836 1845 1854
 GTC CGT GGT CTT ACC GGC CCC ATT GGT CCT CCT GGC CCT GCT GGT GGC CCT GGT GAC AAG
 1869 1878 1887 1896 1905 1914
 GGT GAA AGT GGT CCC AGC GGC CCT GCT GGT CCC ACT GGA GCT CGT GGT GCC CCC GGA GAC
 1929 1938 1947 1956 1965 1974
 CGT GGT GAG CCT GGT CCC CCC GGC CCT GCT GGC TTT GCT GGC CCC CCT GGT GCT GAC GGC
 1989 1998 2007 2016 2025 2034
 CAA CCT GGT GCT AAA GGC GAA CCT GGT GAT CTT GGT GCC AAA GGC GAT GCT GGT CCC CCT
 2049 2058 2067 2076 2085 2094
 GGG CCT GGC GGA CCC GCT GGA CCC CCT GGC CCC ATT GGT AAT GTT GGT GCT CCT GGA GCC
 2109 2118 2127 2136 2145 2154
 AAA GGT GCT CGC CGC AGC GCT GGT CCC CCT GGT GCT ACT GGT TTC CCT GGT GCT GCT GGC
 2169 2178 2187 2196 2205 2214
 CGA GTC GGT CCT CCT GGC CCC TCT GGA AAT GCT GGA CCC CCT GGC CCT CCT GGT CCT GCT
 2229 2238 2247 2256 2265 2274
 GGC AAA GAA GGC GGC AAA GGT CCC CGT GGT GAG ACT GGC CCT GCT GGA CGT CCT GGT GAA
 2289 2298 2307 2316 2325 2334
 GTT GGT CCC CTT GGT CCC CCT GGC CCT GCT GGC GAG AAA GGA TCC CCT GGT GCT GAT GGT
 2349 2358 2367 2376 2385 2394
 CTT CCT GGT GCT CTT GGT ACT CCC GGC CCT CAA GGT ATT GCT GGA CAG CGT GGT GTG GTC
 2409 2418 2427 2436 2445 2454
 GGC CTT CCT GGT CAG AGA GGA GAG AGA GGC TTC CCT GGT CTT CCT GGC CCC TCT GGT GAA
 2469 2478 2487 2496 2505 2514
 CTT CCC AAA CAA GGT CCC TCT GGA GCA AGT GGT GAA CGT GGT CCC CCC GGT CCC ATG GGC

FIG. 19B

2529	2538	2547	2556	2565	2574
CCC CCT GGA TTT CCT GGA CCC CCT GGT GAA TCT GGA CGT GAG GGG GCT CCT GCT GCC GAA					
2589	2598	2607	2616	2625	2634
GGT TCC CCT GGA CGA GAC GGT TCT CCT GGC GCC AAG GGT GAC CGT GGT GAG ACC GGC CCC					
2649	2658	2667	2676	2685	2694
GCT GGA CCC CCT GGT GGT CCT GGT GGT CCT GGT GCC CCT GGC CCC GTT GGC CCT GCT GGC					
2709	2718	2727	2736	2745	2754
AAG AGT CGT GAT CGT GGT GAG ACT GGT CCT GGT GGT CCC GCC GGT CCC GTC GGC CCC GCT					
2769	2778	2787	2796	2805	2814
GGC GCC CGT GGC CCC GGC CGA CCC CAA GGC CCC CGT GGT GAC AAG GGT GAG ACA GGC GAA					
2829	2838	2847	2856	2865	2874
CAG GGC GAC AGA GGC ATA AAG GGT CAC CGT GGC TTC TCT GGC CTC CAG GGT CCC CCT GGC					
2889	2898	2907	2916	2925	2934
CCT CCT GGC TCT CCT GGT GAA CAA GGT CCC TCT GGA GCC TCT GGT CCT GCT GGT CCC CGA					
2949	2958	2967	2976	2985	2994
GGT CCC CCT GGC TCT GGT GGT GCT CCT GGC AAA GAT GGA CTC AAC GGT CTC CCT GGC CCC					
3009	3018	3027	3036	3045	3054
ATT GGC CCC CCT GGT CCT CGC GGT CGC ACT GGT GAT GCT GGT CCT GTT GGT CCC CCC GGC					
3069	3078	3087	3096	3105	3114
CCT CCT GGA CCT CCT GGT CCC CCT GGT CCT CCC AGC GCT GGT TTC GAC TTC AGC TTC CTC					
3129	3138	3147	3156	3165	3174
CCC CAG CCA CCT CAA GAG AAG GCT CAC GAT GGT GGC CGC TAC TAC CGG GCT AGA TCC GAT					
3189	3198	3207	3216	3225	3234
GAG GCT TCT GGG ATA GCC CCA GAA GTT CCT GAT GAC CGC GAC TTC GAG CCC TCC CTA GGC					
3249	3258	3267	3276	3285	3294
CCA GTG TGC CCC TTC CGC TGT CAA TGC CAT CTT CGA GTG GTC CAG TGT TCT GAT TTG GGT					
3309	3318	3327	3336	3345	3354
CTG GAC AAA GTG CCA AAG GAT CTT CCC CCT GAC ACA ACT CTG CTA GAC CTG CAA AAC AAC					
3369	3378	3387	3396	3405	3414
AAA ATA ACC GAA ATC AAA GAT GGA GAC TTT AAG AAC CTG AAG AAC CTT CAC GCA TTG ATT					
3429	3438	3447	3456	3465	3474
CTT GTC AAC AAT AAA ATT AGC AAA GTT AGT CCT GGA GCA TTT ACA CCT TTG GTG AAG TTG					
3489	3498	3507	3516	3525	3534
GAA CGA CTT TAT CTG TCC AAG AAT CAG CTG AAG GAA TTG CCA GAA AAA ATG CCC AAA ACT					
3549	3558	3567	3576	3585	3594
CTT CAG GAG CTG CTT GCC CAT GAG AAT GAG ATC ACC AAA GTG CGA AAA GTT ACT TTC AAT					
3609	3618	3627	3636	3645	3654
GGA CTG AAC CAG ATG ATT GTC ATA GAA CTG GGC ACC AAT CCG CTG AAG AGC TCA GGA ATT					
3669	3678	3687	3696	3705	3714
GAA AAT GAG CCT TTC CAG GGA ATG AAG AAG CTC TCC TAC ATC CGC ATT GCT GAT ACC AAT					
3729	3738	3747	3756	3765	3774
ATC ACC AGC ATT CCT CAA GGT CTT CCT CCT TCC CTT ACG GAA TTA CAT CTT GAT GGC AAC					

FIG. 19C

3785	3798	3807	3816	3825	3834
AAA ATC AGC AGA GTT CAT GCA GCT AGC CTG AAA GGA CTG AAT AAT TTG GCT AAG TTG GGA					
3849	3858	3867	3876	3885	3894
TTG AGT TTC AAC AGC ATC TCT GCT GTT GAC AAT GGC TCT CTG GCC AAC ACG CCT CAT CTG					
3909	3918	3927	3936	3945	3954
AGG GAG CTT CAC TTG GAC AAC AAC AAG CTT ACC AGA GTA CCT GGT GGG CTG GCA GAG CAT					
3969	3978	3987	3996	4005	4014
AAC TAC ATC CAG GTT CTC TAC CTT CAT AAC AAC AAT ATC TCT GTA GTT GGA TCA AGT GAC					
4029	4038	4047	4056	4065	4074
TTC TGC CCA CCT GGA CAC AAC ACC AAA AAG GCT TCT TAT TCG GGT GTG AGT CTT TTC AGC					
4089	4098	4107	4116	4125	4134
AAC CCG GTC CAG TAC TGG GAG ATA CAG CCA TCC ACC TTC AGA TGT GTC TAC GTG CGC TCT					
4149	4158	4167	4176	4185	4194
GCC ATT CAA CTC GGA AAC TAT AAG TAA					

FIG. 19D


```

      10      20      30      40      50      60
gggaaggatt tccatttccc AGCTGTCTTA TGGCTATGAT GAGAAATCAA CCGGAGGAAAT

      70      80      90     100     110     120
TTCCGTGCCT GGGCCCATGG GTCCCTCTGG TCCTCGTGGT CTCCCTGGCC CCCCTGGTGC

      130     140     150     160     170     180
ACCTGGTCCC CAAGGCTTCC AAGGTCCCCC TGGTGAGCCT GCGGAGCCTG GAGCTTCAGG

      190     200     210     220     230     240
TCCCATGGGT CCCCAGGGTC CCCCAGGTCC CCCTGGAAAG AATGGAGATG ATGGGGAAGC

      250     260     270     280     290     300
TGGA AAA CCT GGTCTCTCTG GTGAGCGTGG GCCTCCTGGG CCTCAGGGTG CTCGAGGATT

      310     320     330     340     350     360
GCCCCGAACA GCTGGCCTCC CTGGAATGAA GGGACACAGA GGTTCAGTG GTTTGGATGG

      370     380     390     400     410     420
TGCCAAGGGA GATGCTGGTC CTGCTGGTCC TAAGGCTGAG CCTGGCAGCC CTGGTGAAAA

      430     440     450     460     470     480
TGGAGCTCCT GGTGAGATGG GCCCCCGTGG CCTGCCTGGT GAGAGAGGTC GCCCTGGAGC

      490     500     510     520     530     540
CCCTGGCCCT GCTGGTGCTC GTGGAATGA TGGTGCTACT GGTGCTGCCG GGCCCCCTGG

      550     560     570     580     590     600
TCCCACCGGC CCCGCTGGTC CTCCTGGCTT CCCTGGTGCT GTTGGTGCTA AGGGTGAAGC

      610     620     630     640     650     660
TGGTCCCCAA GGGCCCCGAG GCTCTGAAGG TCCCCAGGGT GTGCGTGGTG AGCCTGGCCC

      670     680     690     700     710     720
CCCTGGCCCT GCTGGTGCTG CTGGCCCTGC TGGAAACCTT GGTGCTGATG GACAGCCTGG

      730     740     750     760     770     780
TGCTAAAGGT GCCAATGGTG CTCCTGGTAT TGCTGGTGCT CCTGGCTTCC CTGGTGCCCC

      790     800     810     820     830     840
AGGCCCCCTCT GGACCCCAAG GCCCCGGCGG CCCTCCTGGT CCCAAGGGTA ACAGCGGTGA

      850     860     870     880     890     900
ACCTGGTGCT CCTGGCAGCA AAGCAGACAC TGGTGCTAAG GGAGAGCCTG GCCCTGTTGG

      910     920     930     940     950     960
TGTTCAAGGA CCCCCTGGCC CTGCTGGAGA GGAAGGAAG CGAGGAGCTC GAGGTGAACC

      970     980     990     1000     1010     1020
CGGACCCACT GGCCTGCCCC GACCCCTGGG CGAGCGTGGT GGACCTGGTA GCCGTGGTTT

      1030     1040     1050     1060     1070     1080
CCCTGGCGCA GATGGTGTG CTGGTCCCAA GGGTCCCGCT GGTGAACGTG GTTCTCCTGG

      1090     1100     1110     1120     1130     1140
CCCCGCTGGC CCCAAAGGAT CTCCTGGTGA AGCTGGTCGT CCCGGTGAAG CTGGTCTGCC

      1150     1160     1170     1180     1190     1200
TGGTGCCAAG GGTCTGACTG GAAGCCCTGG CAGCCCTGGT CCTGATGCCA AAACCTGGCCC

      1210     1220     1230     1240     1250     1260
CCCTGGTCCC GCGGTCAAG ATGGTCCCCC CGGACCCCA GGCCACCTG GTGCCCCGTG

```

FIG. 20A

1270	1280	1290	1300	1310	1320
TCAGGCTGGT	GTGATGGGAT	TCCCTGGACC	TAAAGGTGCT	GCTGGAGAGC	CCGGCAAGGC
1330	1340	1350	1360	1370	1380
TGGAGAGCGA	GGTGTTCCTG	GACCCCTGG	CGCTGTCGGT	CCTGCTGGCA	AAGATGGAGA
1390	1400	1410	1420	1430	1440
GGCTGGAGCT	CAGGGACCCC	CTGGCCCTGC	TGGTCCCCTG	GGCGAGAGAG	GTGAACAAGG
1450	1460	1470	1480	1490	1500
CCCTGCTGGC	TCCCCCGGAT	TCCAGGGTCT	CCCTGGTCTT	GCTGGTCTCT	CAGGTGAAGC
1510	1520	1530	1540	1550	1560
AGGCAACCT	GGTGAACAGG	GTGTTCTCTG	AGACCTTGGC	GCCCCCTGGC	CCTCTGGAGC
1570	1580	1590	1600	1610	1620
AAGAGGCGAG	AGAGGTTTCC	CTGGCGAGCG	TGGTGTGCAA	GGTCCCCCTG	GTCCCTGCTG
1630	1640	1650	1660	1670	1680
ACCCCGAGGG	GCCAACGGTG	CTCCCCGGCA	CGATGGTGCT	AAGGGTGATG	CTGGTGCCCC
1690	1700	1710	1720	1730	1740
TGGAGCTCCC	GGTAGCCAGG	GCGCCCCCTG	CCTTCAGGGA	ATGCCTGGTG	AACGTGGTGC
1750	1760	1770	1780	1790	1800
AGCTGGTCTT	CCAGGGCCTA	AGGGTGACAG	AGGTGATGCT	GGTCCCAAAG	GTGCTGATGG
1810	1820	1830	1840	1850	1860
CTCTCCTGGC	AAAGATGGCG	TCCGTGGTCT	GACCGGCCCC	ATTGGTCTCT	CTGGCCCTGC
1870	1880	1890	1900	1910	1920
TGGTGCCCCCT	GGTGACAAGG	GTGAAAGTGG	TCCCAGCGGC	CCTGCTGGTC	CCACTGGAGC
1930	1940	1950	1960	1970	1980
TCGTGGTCCC	CCCGGAGACC	GTGGTGAGCC	TGGTCCCCCT	GGCCCTGCTG	GCTTTGCTGG
1990	2000	2010	2020	2030	2040
CCCCCTGGT	GCTGACGGCC	AACCTGGTGC	TAAAGGCGAA	CCTGGTGATG	CTGGTGCCAA
2050	2060	2070	2080	2090	2100
AGGCGATGCT	GGTCCCCCTG	GGCCTGCCGG	ACCCGCTGGA	CCCCCTGGCC	CCATGGGTAA
2110	2120	2130	2140	2150	2160
TGTTGGTGCT	CCTGGAGCCA	AAGGTGCTCG	CGGCAGCGCT	GGTCCCCCTG	GTGCTACTGG
2170	2180	2190	2200	2210	2220
TTTCCCTGGT	GCTGCTGGCC	GAGTCGGTCC	TCCTGGCCCC	TCTGGAAATG	CTGGACCCCC
2230	2240	2250	2260	2270	2280
TGGCCCTCCT	GGTCTGCTG	GCAAAGAAGG	CGGCAAAGGT	CCCCGTGGTG	AGACTGGCCC
2290	2300	2310	2320	2330	2340
TGCTGGACGT	CCTGGTGAAG	TGGTCCCCCT	TGGTCCCCCT	GGCCCTGCTG	GCGAGAAAGG
2350	2360	2370	2380	2390	2400
ATCCCCCTGGT	GCTGATGGTC	CTGCTGGTGC	TCCTGGTACT	CCCCGGCCTC	AAGGTATTGC
2410	2420	2430	2440	2450	2460
TGGACAGCGT	GGTGTGCTCG	GCCTGCCTGG	TCAGAGAGGA	GAGAGAGGCT	TCCCTGGTCT
2470	2480	2490	2500	2510	2520
TCCTGGCCCC	TCTGGTGAAC	CTGGCAAACA	AGGTCCCTCT	GGAGCAAAGT	GTGAACGTGG

FIG. 20B

```

      2530      2540      2550      2560      2570      2580
TCCCCCCGGT CCCATGGGCC CCCCTGGATT GGCTGGACCC CCTGGTGAAT CTGGACGTGA

      2590      2600      2610      2620      2630      2640
GGGGGCTCCT GCTGCCGAAG GTTCCCCTGG ACGAGACGGT TCTCCTGGCG CCAAGGGTGA

      2650      2660      2670      2680      2690      2700
CCGTGGTGAG ACCGGCCCCG CTGGACCCCC TGGTGCTCNT GGTGCTCNTG GTGCCCCCTG

      2710      2720      2730      2740      2750      2760
CCCCGTGGC CCTGCTGGCA AGAGTGGTGA TCGTGGTGAG ACTGGTCCTG CTGGTCCCGC

      2770      2780      2790      2800      2810      2820
CGGTCCCGTC GGGCCCGCTG GCGCCCGTGG CCCC GCCGAAGGCC CCCGTGGTGA

      2830      2840      2850      2860      2870      2880
CAAGGGTGAG ACAGGCGAAC AGGCGGACAG AGGCATAAAG GGTCAACCGT GCTTCTCTGG

      2890      2900      2910      2920      2930      2940
CCTCCAGGGT CCCCCTGGCC CTCCTGGCTC TCCTGGTGAA CAAGGTCCCT CTGGAGCCTC

      2950      2960      2970      2980      2990      3000
TGGTCCTGCT GGTCCCCGAG GTCCCCCTGG CTCTGCTGGT GCTCCTGGCA AAGATGGACT

      3010      3020      3030      3040      3050      3060
CAACGGTCTC CCTGGCCCCA TTGGGCCCCC TGGTCCTCGC GGTCCGACTG GTGATGCTGG

      3070      3080      3090      3100      3110      3120
TCCTGTGGT CCCCCCGGCC CTCCTGGACC TCCTGGTCCC CCTGGTCCTC CCAGCCTGG

      3130      3140      3150      3160      3170      3180
TTTCGACTTC AGCTTCCTCC CCCAGCCACC TCAAGAGAAG GCTCACGATG GTGGCCGCTA

      3190      3200      3210      3220      3230      3240
CTACCGGGCT agatctccaa AGGATCTTCC CCCTGACACA ACTCTGCTAG ACCTGCAAAA

      3250      3260      3270      3280      3290      3300
CAACAAAATA ACCGAAATCA AAGATGGAGA CTTTAAGAAC CTGAAGAACC TTCACGCATT

      3310      3320      3330      3340      3350      3360
GATTCTTGTC AACAATAAAA TTAGCAAAGT TAGTCCTGGA TAActgcag. ....

```

FIG. 20C

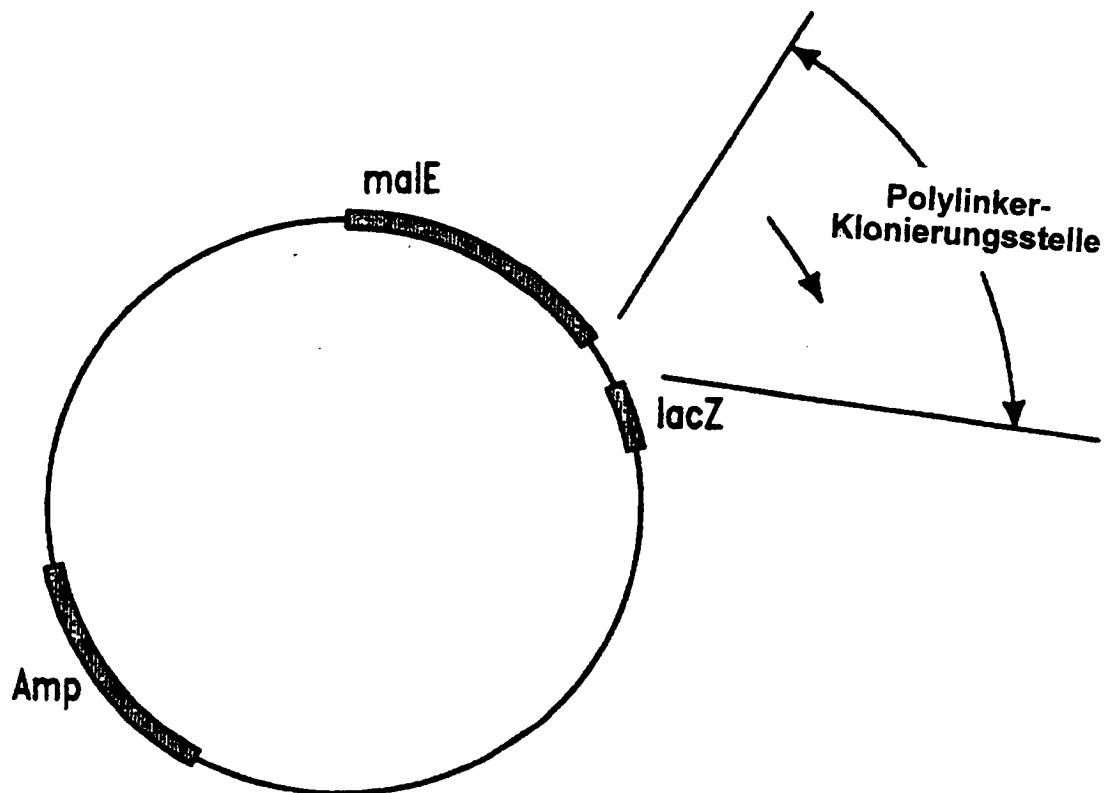


FIG. 2I

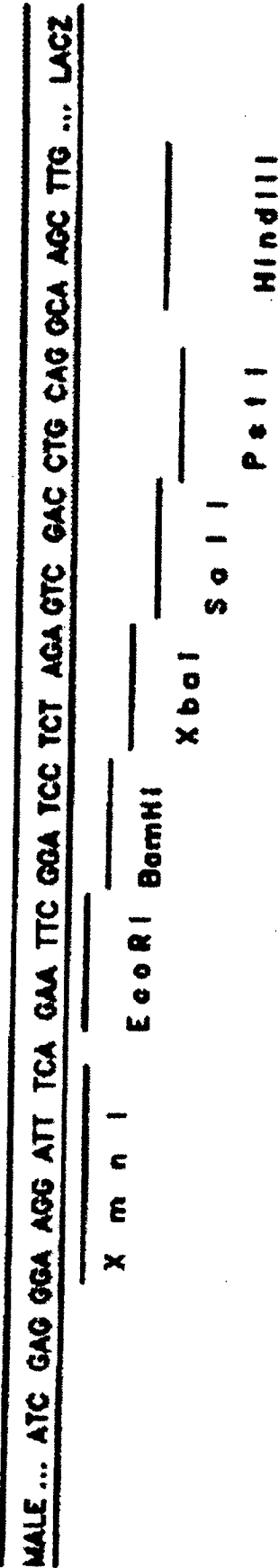


FIG. 22

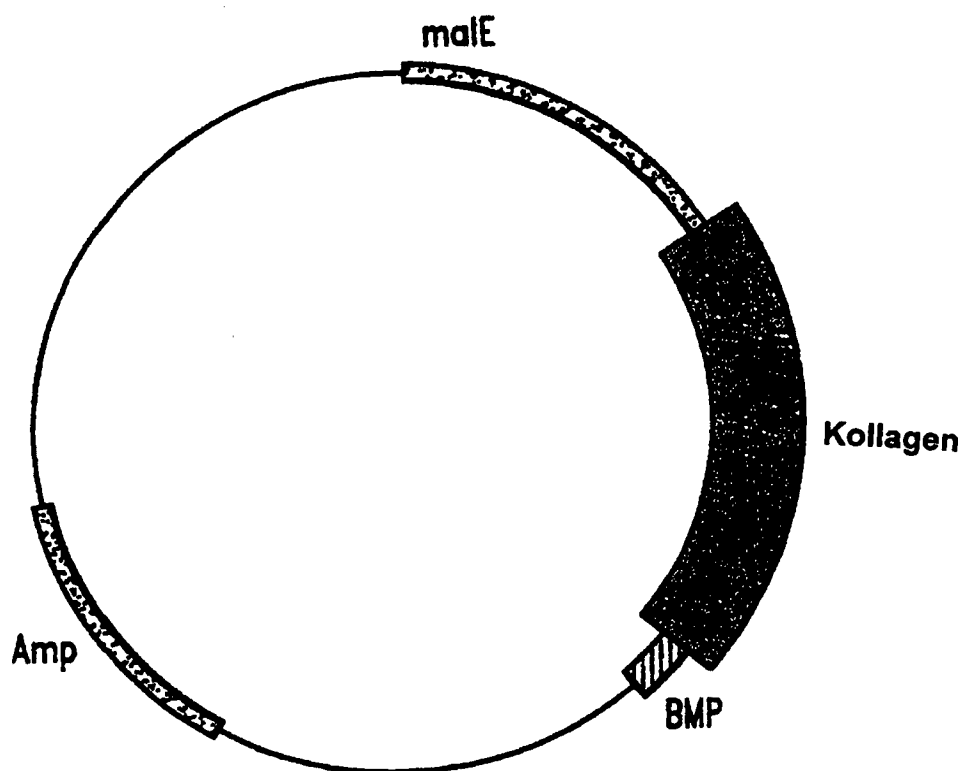


FIG. 23

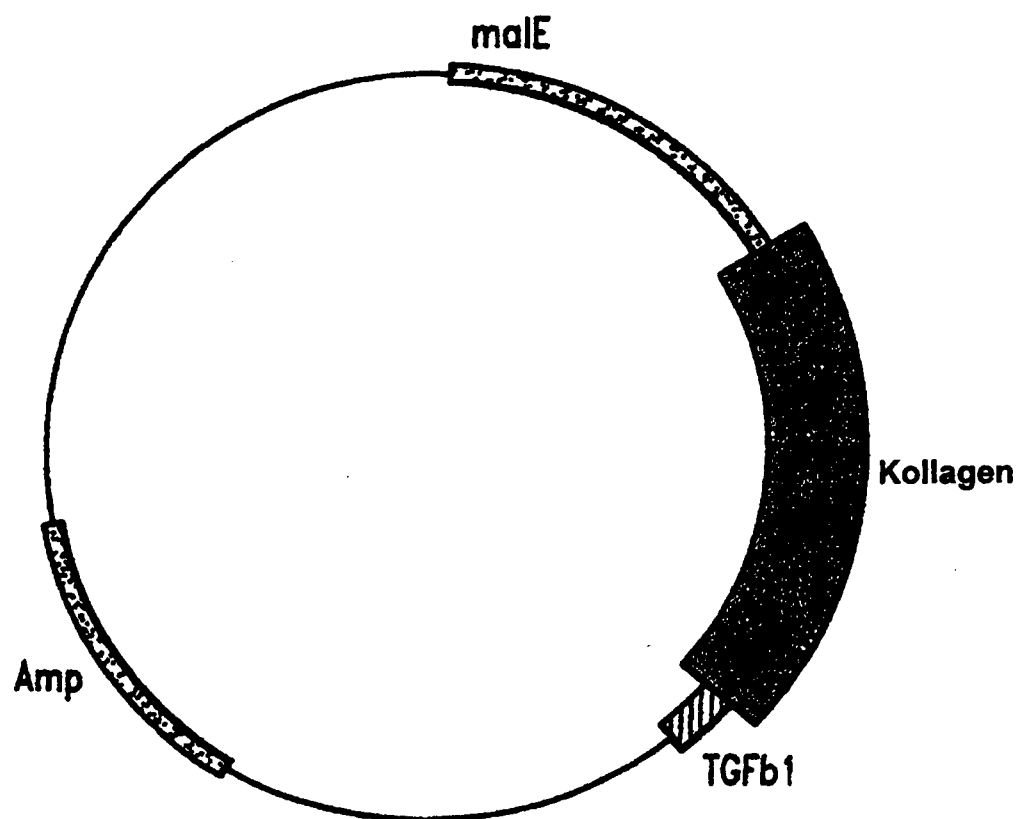


FIG. 24

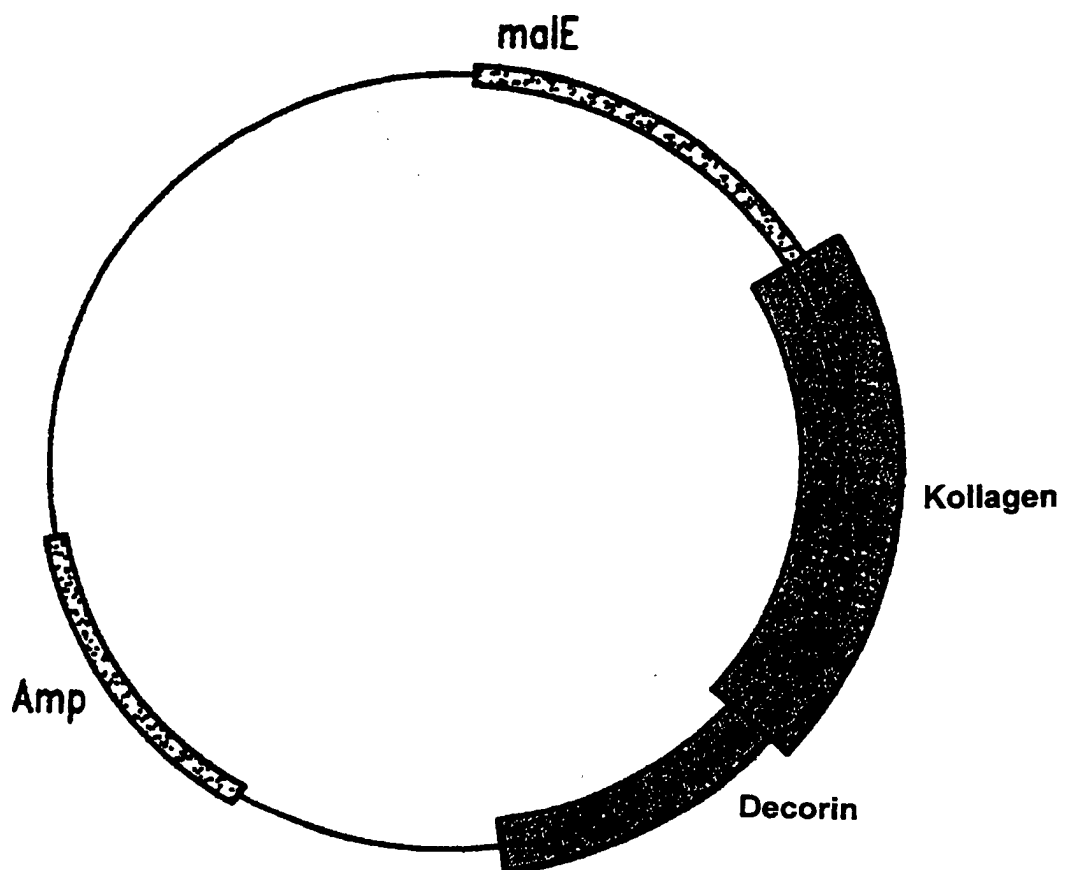


FIG. 25

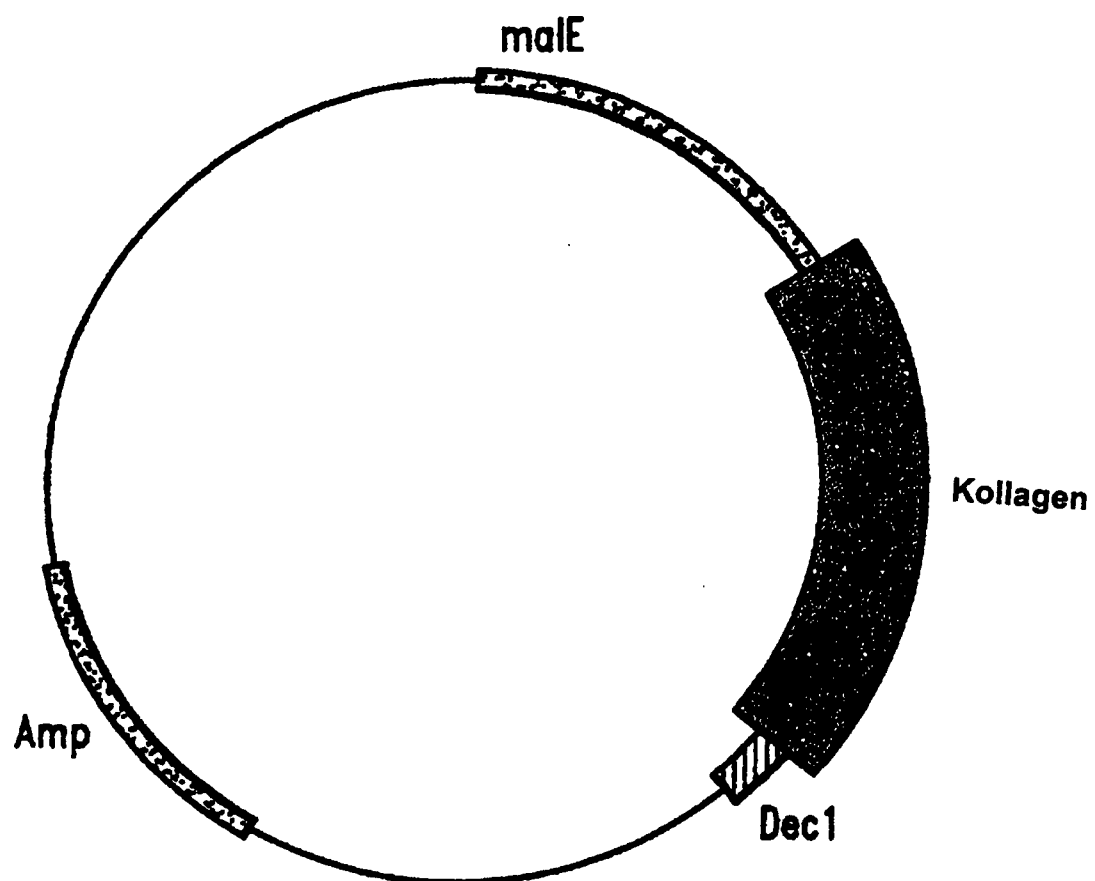


FIG. 26

```

      9      18      27      36      45      54
5'  CAG CTG TCT TAT GGC TAT GAT GAG AAA TCA ACC GGA GGA ATT TCC GTC CCT GGC
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
    Gln Leu Ser Tyr Gly Tyr Asp Glu Lys Ser Thr Gly Gly Ile Ser Val Pro Gly

      63      72      81      90      99      108
    CCC ATG GGT CCC TCT GGT CCT CGT GGT CTC CCT GGC CCC CCT GGT GCA CCT GGT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
    Pro Met Gly Pro Ser Gly Pro Arg Gly Leu Pro Gly Pro Pro Gly Ala Pro Gly

      117      126      135      144      153      162
    CCC CAA GGC TTC CAA GGT CCC CCT GGT GAG CCT GGC GAG CCT GGA GGT TCA GGT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
    Pro Gln Gly Phe Gln Gly Pro Pro Gly Glu Pro Gly Glu Pro Gly Ala Ser Gly

      171      180      189      198      207      216
    CCC ATG GGT CCC CGA GGT CCC CCA GGT CCC CCT GGA AAG AAT GGA GAT GAT GGC
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
    Pro Met Gly Pro Arg Gly Pro Pro Gly Pro Pro Gly Lys Asn Gly Asp Asp Gly

      225      234      243      252      261      270
    GAA GCT GGA AAA CCT GGT CGT CCT GGT GAG CGT GGC CCT CCT GGC CCT CAG GGT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
    Glu Ala Gly Lys Pro Gly Arg Pro Gly Glu Arg Gly Pro Pro Gly Pro Gln Gly

      279      288      297      306      315      324
    GGT CGA GGA TTG CCC GGA ACA GGT GGC CTC CCT GGA ATG AAG GGA CAC AGA GGT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
    Ala Arg Gly Leu Pro Gly Thr Ala Gly Leu Pro Gly Met Lys Gly His Arg Gly

      333      342      351      360      369      378
    TTC AGT GGT TTG GAT GGT GCC AAG GGA GAT GCT GGT CCT GGT GGT CCT AAG GGT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
    Phe Ser Gly Leu Asp Gly Ala Lys Gly Asp Ala Gly Pro Ala Gly Pro Lys Gly

      387      396      405      414      423      432
    GAG CCT GGC AGC CCT GGT GAA AAT GGA GCT CCT GGT CAG ATG GGC CCC CGT GGC
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
    Glu Pro Gly Ser Pro Gly Glu Asn Gly Ala Pro Gly Gln Met Gly Pro Arg Gly

      441      450      459      468      477      486
    CTG CCT GGT GAG AGA GGT CGC CCT GGA GCC CCT GGC CCT GCT GGT GCT CGT GGA
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
    Leu Pro Gly Glu Arg Gly Arg Pro Gly Ala Pro Gly Pro Ala Gly Ala Arg Gly

      495      504      513      522      531      540
    AAT GAT GGT GCT ACT GGT GCT GGC GGG CCC CCT GGT CCC ACC GGC CCC GCT GGT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
    Asn Asp Gly Ala Thr Gly Ala Ala Gly Pro Pro Gly Pro Thr Gly Pro Ala Gly

      549      558      567      576      585      594
    CCT CCT GGC TTC CCT GGT GCT GTT GGT GCT AAG GGT GAA GCT GGT CCC CAA GGC
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
    Pro Pro Gly Phe Pro Gly Ala Val Gly Ala Lys Gly Glu Ala Gly Pro Gln Gly

      603      612      621      630      639      648
    CCC CGA GGC TCT GAA GGT CCC CAG GGT GTG CGT GGT GAG CCT GGC CCC CCT GGC
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
    Pro Arg Gly Ser Glu Gly Pro Gln Gly Val Arg Gly Glu Pro Gly Pro Pro Gly

      657      666      675      684      693      702
    CCT GCT GGT GCT GCT GGC CCT GCT GGA AAC CCT GGT GCT GAT GGA CAG CCT GGT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
    Pro Ala Gly Ala Ala Gly Pro Ala Gly Asn Pro Gly Ala Asp Gly Gln Pro Gly

```

FIG. 27A

```

      711      720      729      738      747      756
GCT AAA GGT GCC AAT GGT GCT CCT GGT ATT GCT GGT GCT CCT GGC TTC CCT GGT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Ala Lys Gly Ala Asn Gly Ala Pro Gly Ile Ala Gly Ala Pro Gly Phe Pro Gly

      765      774      783      792      801      810
GCC CGA GGC CCC TCT GGA CCC CAG GGC CCC GGC GGC CCT CCT GGT CCC AAG GGT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Ala Arg Gly Pro Ser Gly Pro Gln Gly Pro Gly Gly Pro Pro Gly Pro Lys Gly

      819      828      837      846      855      864
AAC AGC GGT GAA CCT GGT GCT CCT GGC AGC AAA GGA GAC ACT GGT GCT AAG GGA
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Asn Ser Gly Glu Pro Gly Ala Pro Gly Ser Lys Gly Asp Thr Gly Ala Lys Gly

      873      882      891      900      909      918
GAG CCT GGC CCT GCT GGT GTT CAA GGA CCC CCT GGC CCT GCT GGA GAG GAA GGA
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Glu Pro Gly Pro Val Gly Val Gln Gly Pro Pro Gly Pro Ala Gly Glu Glu Gly

      927      936      945      954      963      972
AAG CGA GGA GCT CGA GGT GAA CCC GGA CCC ACT GGC CTG CCC GGA CCC CCT GGC
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Lys Arg Gly Ala Arg Gly Glu Pro Gly Pro Thr Gly Leu Pro Gly Pro Pro Gly

      981      990      999      1008      1017      1026
GAG GGT GGT GGA CCT GGT AGC CGT GGT TTC CCT GGC GCA GAT GGT GTT GCT GGT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Glu Arg Gly Gly Pro Gly Ser Arg Gly Phe Pro Gly Ala Asp Gly Val Ala Gly

      1035      1044      1053      1062      1071      1080
CCC AAG GGT CCC GCT GGT GAA CGT GGT TCT CCT GGC CCC GCT GGC CCC AAA GGA
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Pro Lys Gly Pro Ala Gly Glu Arg Gly Ser Pro Gly Pro Ala Gly Pro Lys Gly

      1089      1098      1107      1116      1125      1134
TCT CCT GGT GAA GCT GGT CGT CCC GGT GAA GCT GGT CTG CCT GGT GGC AAG GGT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Ser Pro Gly Glu Ala Gly Arg Pro Gly Glu Ala Gly Leu Pro Gly Ala Lys Gly

      1143      1152      1161      1170      1179      1188
CTG ACT GGA AGC CCT GGC AGC CCT GGT CCT GAT GGC AAA ACT GGC CCC CCT GGT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Leu Thr Gly Ser Pro Gly Ser Pro Gly Pro Asp Gly Lys Thr Gly Pro Pro Gly

      1197      1206      1215      1224      1233      1242
CCC GGC GGT CAA GAT GGT GGC CCC GGA CCC CCA GGC CCA CCT GGT GGC CGT GGT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Pro Ala Gly Gln Asp Gly Arg Pro Gly Pro Pro Gly Pro Pro Gly Ala Arg Gly

      1251      1260      1269      1278      1287      1296
CAG GCT GGT GTG ATG GGA TTC CCT GGA CCT AAA GGT GCT GCT GGA GAG CCC GGC
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Gln Ala Gly Val Met Gly Phe Pro Gly Pro Lys Gly Ala Ala Gly Glu Pro Gly

      1305      1314      1323      1332      1341      1350
AAG GCT GGA GAG CGA GGT GTT CCC GGA CCC CCT GGC GCT GTC GGT CCT GCT GGC
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Lys Ala Gly Glu Arg Gly Val Pro Gly Pro Pro Gly Ala Val Gly Pro Ala Gly

```

FIG. 27B

```

      1359      1368      1377      1386      1395      1404
AAA GAT GGA GAG GCT GGA GCT CAG GGA CCC CCT GGC CCT GCT GGT CCC GCT GGC
---
Lys Asp Gly Glu Ala Gly Ala Gln Gly Pro Pro Gly Pro Ala Gly Pro Ala Gly

      1413      1422      1431      1440      1449      1458
GAG AGA GGT GAA CAA GGC CCT GCT GGC TCC CCC GGA TTC CAG GGT CTC CCT GGT
---
Glu Arg Gly Glu Gln Gly Pro Ala Gly Ser Pro Gly Phe Gln Gly Leu Pro Gly

      1467      1476      1485      1494      1503      1512
CCT GCT GGT CCT CCA GGT GAA GCA GGC AAA CCT GGT GAA CAG GGT GTT CCT GGA
---
Pro Ala Gly Pro Pro Gly Glu Ala Gly Lys Pro Gly Glu Gln Gly Val Pro Gly

      1521      1530      1539      1548      1557      1566
GAC CTT GGC GGC CCT GGC CCC TCT GGA GCA AGA GGC GAG AGA GGT TTC CCT GGC
---
Asp Leu Gly Ala Pro Gly Pro Ser Gly Ala Arg Gly Glu Arg Gly Phe Pro Gly

      1575      1584      1593      1602      1611      1620
GAG CGT GGT GTG CAA GGT CCC CCT GGT CCT GCT GGA CCC CGA GGG GCC AAC GGT
---
Glu Arg Gly Val Gln Gly Pro Pro Gly Pro Ala Gly Pro Arg Gly Ala Asn Gly

      1629      1638      1647      1656      1665      1674
GCT CCC GGC AAC GAT GGT GCT AAG GGT GAT GCT GGT GCC CCT GGA GCT CCC GGT
---
Ala Pro Gly Asn Asp Gly Ala Lys Gly Asp Ala Gly Ala Pro Gly Ala Pro Gly

      1683      1692      1701      1710      1719      1728
AGC CAG GGC GGC CCT GGC CTT CAG GGA ATG CCT GGT GAA CGT GGT GCA GCT GGT
---
Ser Gln Gly Ala Pro Gly Leu Gln Gly Met Pro Gly Glu Arg Gly Ala Ala Gly

      1737      1746      1755      1764      1773      1782
CTT CCA GGG CCT AAG GGT GAC AGA GGT GAT GCT GGT CCC AAA GGT GCT GAT GGC
---
Leu Pro Gly Pro Lys Gly Asp Arg Gly Asp Ala Gly Pro Lys Gly Ala Asp Gly

      1791      1800      1809      1818      1827      1836
TCT CTT GGC AAA GAT GGC GTC CGT GGT CTG ACC GGC CCC ATT GGT CCT CCT GGC
---
Ser Pro Gly Lys Asp Gly Val Arg Gly Leu Thr Gly Pro Ile Gly Pro Pro Gly

      1845      1854      1863      1872      1881      1890
CCT GCT GGT GCC CCT GGT GAC AAG GGT GAA AGT GGT CCC AGC GGC CCT GCT GGT
---
Pro Ala Gly Ala Pro Gly Asp Lys Gly Glu Ser Gly Pro Ser Gly Pro Ala Gly

      1899      1908      1917      1926      1935      1944
CCC ACT GGA GCT CGT GGT GCC CCC GGA GAC CGT GGT GAG CCT GGT CCC CCC GGC
---
Pro Thr Gly Ala Arg Gly Ala Pro Gly Asp Arg Gly Glu Pro Gly Pro Pro Gly

      1953      1962      1971      1980      1989      1998
CCT GCT GGC TTT GCT GGC CCC CCT GGT GCT GAC GGC CAA CCT GGT GCT AAA GGC
---
Pro Ala Gly Phe Ala Gly Pro Pro Gly Ala Asp Gly Gln Pro Gly Ala Lys Gly

      2007      2016      2025      2034      2043      2052
GAA CCT GGT GAT GCT GGT GCC AAA GGC GAT GCT GGT CCC CCT GGG CCT GCC GGA
---
Glu Pro Gly Asp Ala Gly Ala Lys Gly Asp Ala Gly Pro Pro Gly Pro Ala Gly

```

FIG. 27C

2061	2070	2079	2088	2097	2106
CCC GCT GGA CCC	CCT GGC CCC	ATT GGT AAT	GTT GGT GCT	OCT GGA GCC	AAA GGT
Pro Ala Gly	Pro Pro Gly	Pro Ile Gly	Asn Val Gly	Ala Pro Gly	Ala Lys Gly
2115	2124	2133	2142	2151	2160
GCT CGC GGC AGC	GCT GGT CCC	CCT GGT GCT	ACT GGT TTC	OCT GGT GCT	GCT GGC
Ala Arg Gly	Ser Ala Gly	Pro Pro Gly	Ala Thr Gly	Phe Pro Gly	Ala Ala Gly
2169	2178	2187	2196	2205	2214
CGA GTC GGT CCT	CCT GGC CCC	TCT GGA AAT	GCT GGA CCC	OCT GGC CCT	OCT GGT
Arg Val Gly	Pro Pro Gly	Pro Ser Gly	Asn Ala Gly	Pro Pro Gly	Pro Pro Gly
2223	2232	2241	2250	2259	2268
OCT GCT GGC AAA	GAA GGC GGC	AAA GGT CCC	CGT GGT GAG	ACT GGC CCT	GCT GGA
Pro Ala Gly	Lys Glu Gly	Gly Lys Gly	Pro Arg Gly	Glu Thr Gly	Pro Ala Gly
2277	2286	2295	2304	2313	2322
CGT OCT GGT GAA	GTT GGT CCC	CCT GGT CCC	CCT GGC CCT	GCT GGC GAG	AAA GGA
Arg Pro Gly	Glu Val Gly	Pro Pro Gly	Pro Pro Gly	Pro Ala Gly	Glu Lys Gly
2331	2340	2349	2358	2367	2376
TCC OCT GGT GCT	GAT GGT CCT	GCT GGT GCT	CCT GGT ACT	CCC GGC OCT	CAA GGT
Ser Pro Gly	Ala Asp Gly	Pro Ala Gly	Ala Pro Gly	Thr Pro Gly	Pro Gln Gly
2385	2394	2403	2412	2421	2430
ATT GCT GGA CAG	CGT GGT GTG	GTC GGC CTG	CCT GGT CAG	AGA GGA GAG	AGA GGC
Ile Ala Gly	Gln Arg Gly	Val Val Gly	Leu Pro Gly	Gln Arg Gly	Glu Arg Gly
2439	2448	2457	2466	2475	2484
TTC OCT GGT CTT	CCT GGC CCC	TCT GGT GAA	CCT GGC AAA	CAA GGT CCC	TCT GGA
Phe Pro Gly	Leu Pro Gly	Pro Ser Gly	Glu Pro Gly	Lys Gln Gly	Pro Ser Gly
2493	2502	2511	2520	2529	2538
GCA AGT GGT GAA	CGT GGT CCC	CCC GGT CCC	ATG GGC CCC	OCT GGA TTG	GCT GGA
Ala Ser Gly	Glu Arg Gly	Pro Pro Gly	Pro Met Gly	Pro Pro Gly	Leu Ala Gly
2547	2556	2565	2574	2583	2592
CCC OCT GGT GAA	TCT GGA CGT	GAG GGC GCT	CCT GCT GCC	GAA GGT TCC	OCT GGA
Pro Pro Gly	Glu Ser Gly	Arg Glu Gly	Ala Pro Ala	Ala Glu Gly	Ser Pro Gly
2601	2610	2619	2628	2637	2646
CGA GAC GGT TCT	CCT GGC GCC	AAG GGT GAC	CGT GGT GAG	ACC GGC CCC	GCT GGA
Arg Asp Gly	Ser Pro Gly	Ala Lys Gly	Asp Arg Gly	Glu Thr Gly	Pro Ala Gly
2655	2664	2673	2682	2691	2700
CCC OCT GGT GCT	CCT GGT GCT	OCT GGT GGC	CCT GGC CCC	GTT GGC CCT	GCT GGC
Pro Pro Gly	Ala Pro Gly	Ala Pro Gly	Ala Pro Gly	Pro Val Gly	Pro Ala Gly

FIG. 27D

```

      2709      2718      2727      2736      2745      2754
AAG AGT GGT GAT GGT GGT GAG ACT GGT CCT GGT GGT CCC GCC GGT CCC GTC GGC
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Lys Ser Gly Asp Arg Gly Glu Thr Gly Pro Ala Gly Pro Ala Gly Pro Val Gly

      2763      2772      2781      2790      2799      2808
CCC GGT GGC GGC GGT GGC CCC GGC GGA CCC CAA GGC CCC GGT GGT GAC AAG GGT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Pro Ala Gly Ala Arg Gly Pro Ala Gly Pro Gln Gly Pro Arg Gly Asp Lys Gly

      2817      2826      2835      2844      2853      2862
GAG ACA GGC GAA CAG GGC GAC AGA GGC ATA AAG GGT CAC GGT GGC TTC TCT GGC
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Glu Thr Gly Glu Gln Gly Asp Arg Gly Ile Lys Gly His Arg Gly Phe Ser Gly

      2871      2880      2889      2898      2907      2916
CTC CAG GGT CCC CCT GGC CCT CCT GGC TCT CCT GGT GAA CAA GGT CCC TCT GGA
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Leu Gln Gly Pro Pro Gly Pro Pro Gly Ser Pro Gly Glu Gln Gly Pro Ser Gly

      2925      2934      2943      2952      2961      2970
CCC TCT GGT CCT GGT GGT CCC CGA GGT CCC CCT GGC TCT GGT GGT GCT CCT GGC
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Ala Ser Gly Pro Ala Gly Pro Arg Gly Pro Pro Gly Ser Ala Gly Ala Pro Gly

      2979      2988      2997      3006      3015      3024
AAA GAT GGA CTC AAC GGT CTC CCT GGC CCC ATT GGG CCC CCT GGT CCT CGC GGT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Lys Asp Gly Leu Asn Gly Leu Pro Gly Pro Ile Gly Pro Pro Gly Pro Arg Gly

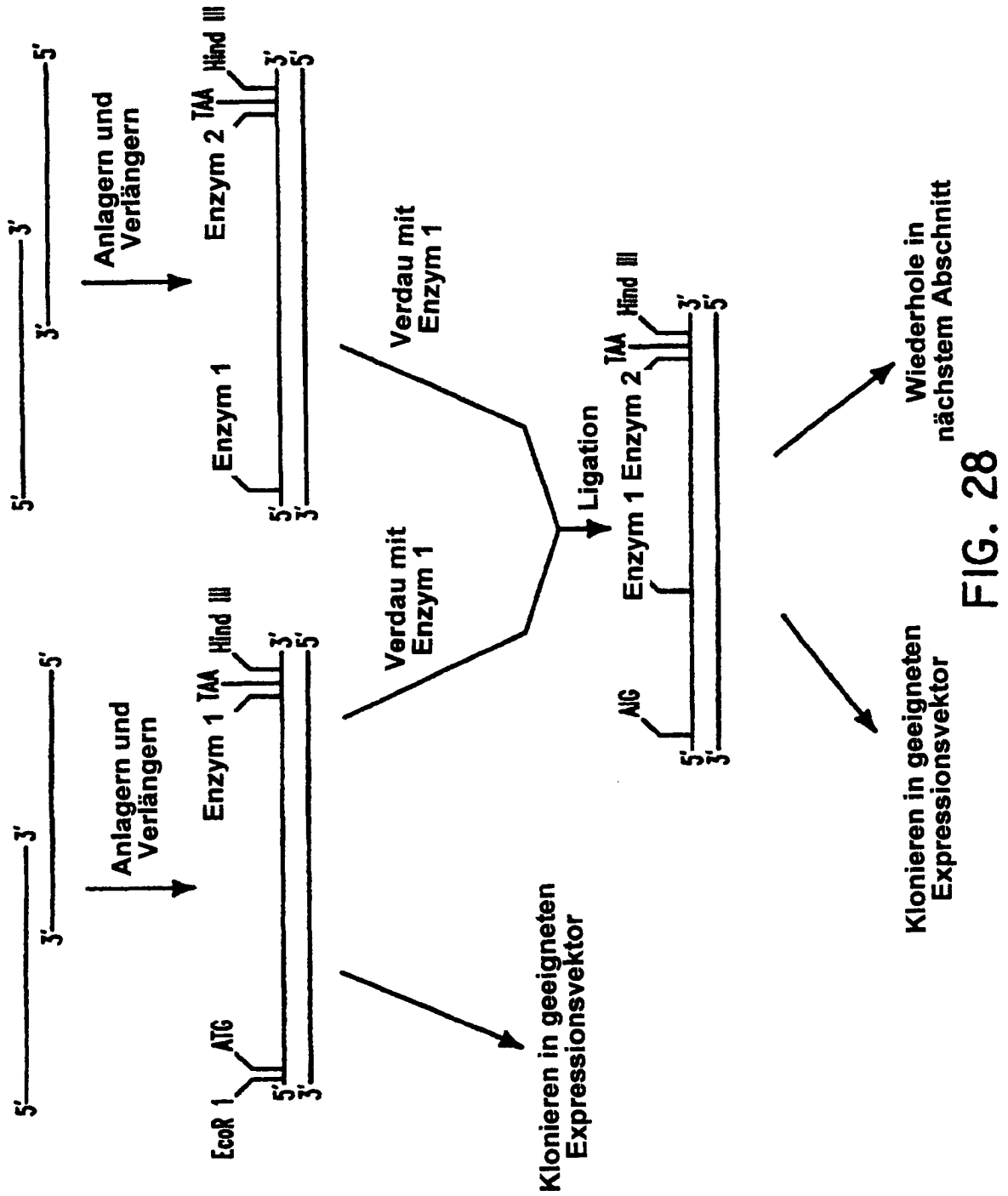
      3033      3042      3051      3060      3069      3078
CGC ACT GGT GAT GGT GGT CCT GTT GGT CCC CCC GGC CCT CCT GGA CCT CCT GGT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Arg Thr Gly Asp Ala Gly Pro Val Gly Pro Pro Gly Pro Pro Gly Pro Pro Gly

      3087      3096      3105      3114      3123      3132
CCC CCT GGT CCT CCC AGC GCT GGT TTC GAC TTC AGC TTC CTC CCC CAG CCA CCT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Pro Pro Gly Pro Pro Ser Ala Gly Phe Asp Phe Ser Phe Leu Pro Gln Pro Pro

      3141      3150      3159      3168
CAA GAG AAG GCT CAC GAT GGT GGC CGC TAC TAC CGG GCT 3'
--- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Gln Glu Lys Ala His Asp Gly Gly Arg Tyr Tyr Arg Ala

```

FIG. 27E



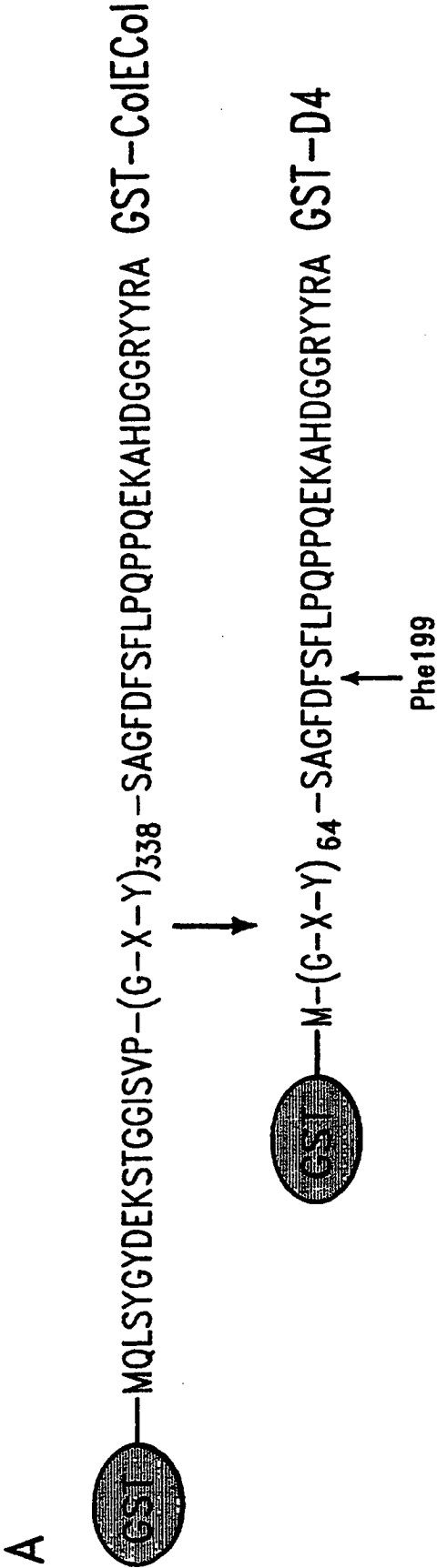


FIG. 29

	HCol	ColECol
Prolin		
CCU	139	11
CCC	93	12
CCA	6	27
CCG	0	189
Glycin		
GGU	174	147
GGC	97	179
GGA	64	8
GGG	11	12

FIG. 30

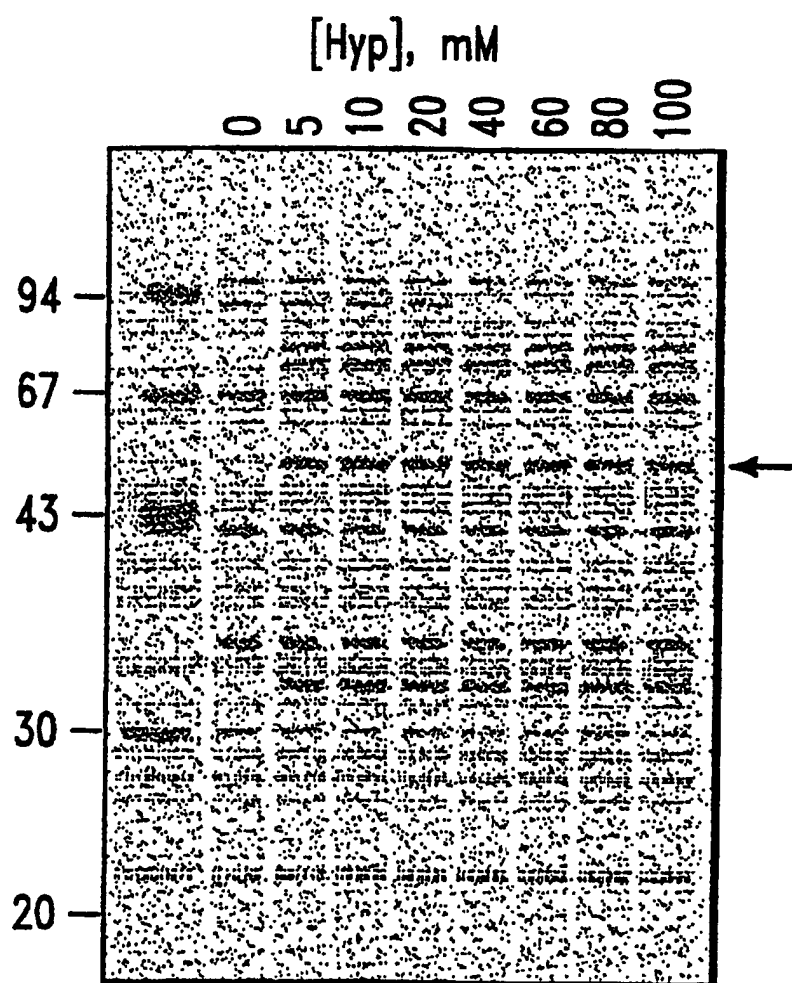


FIG. 31

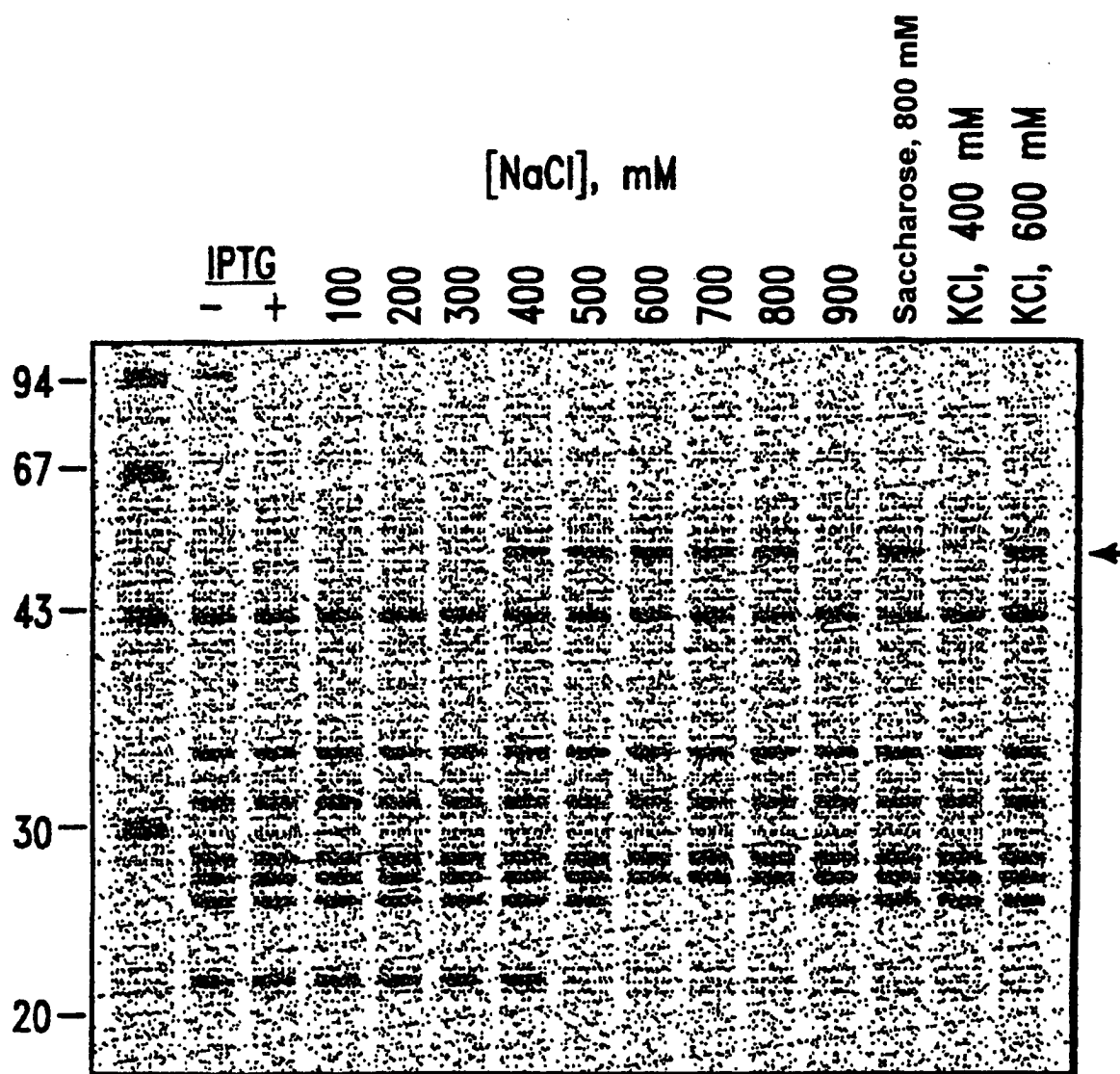


FIG. 32

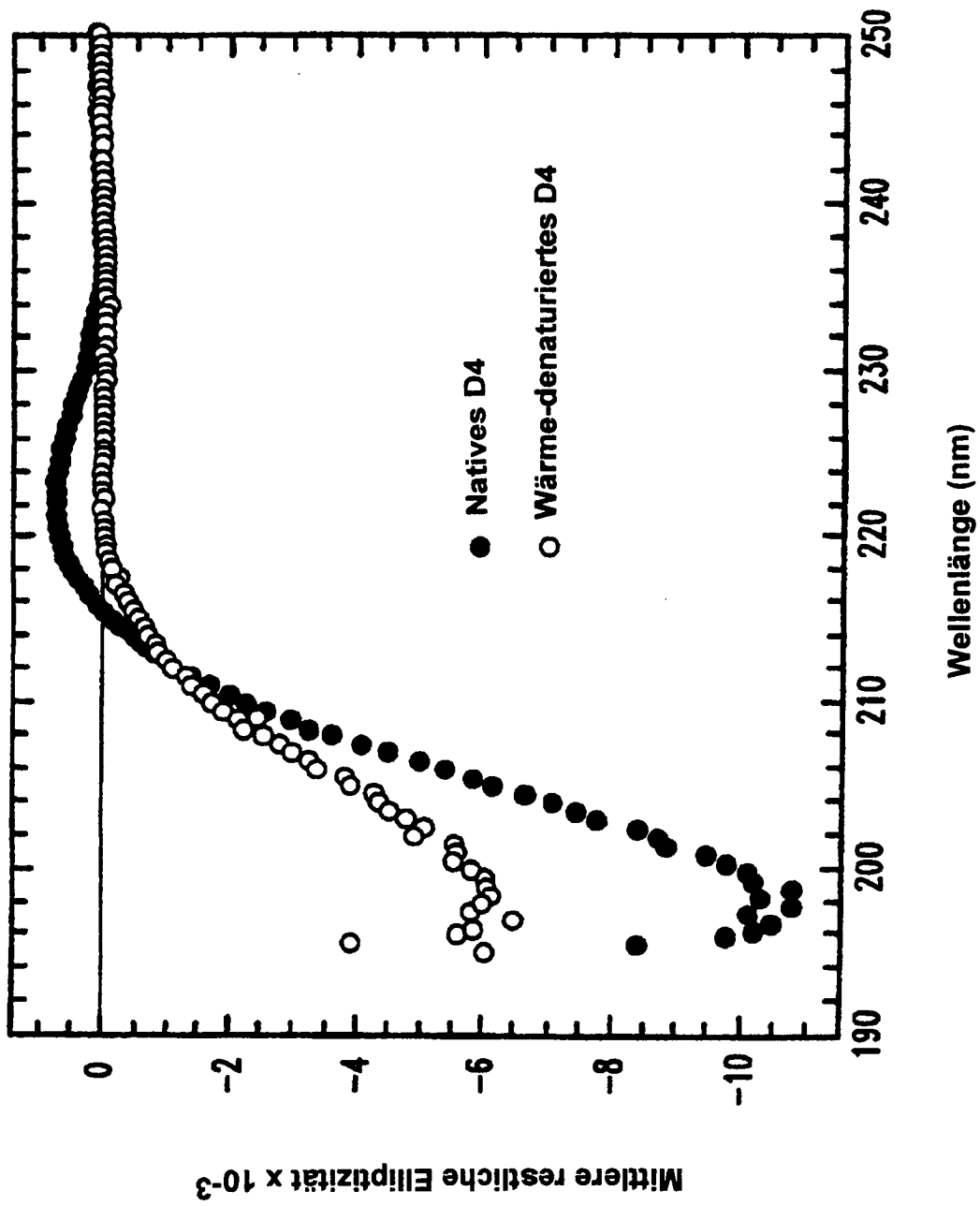


FIG. 33

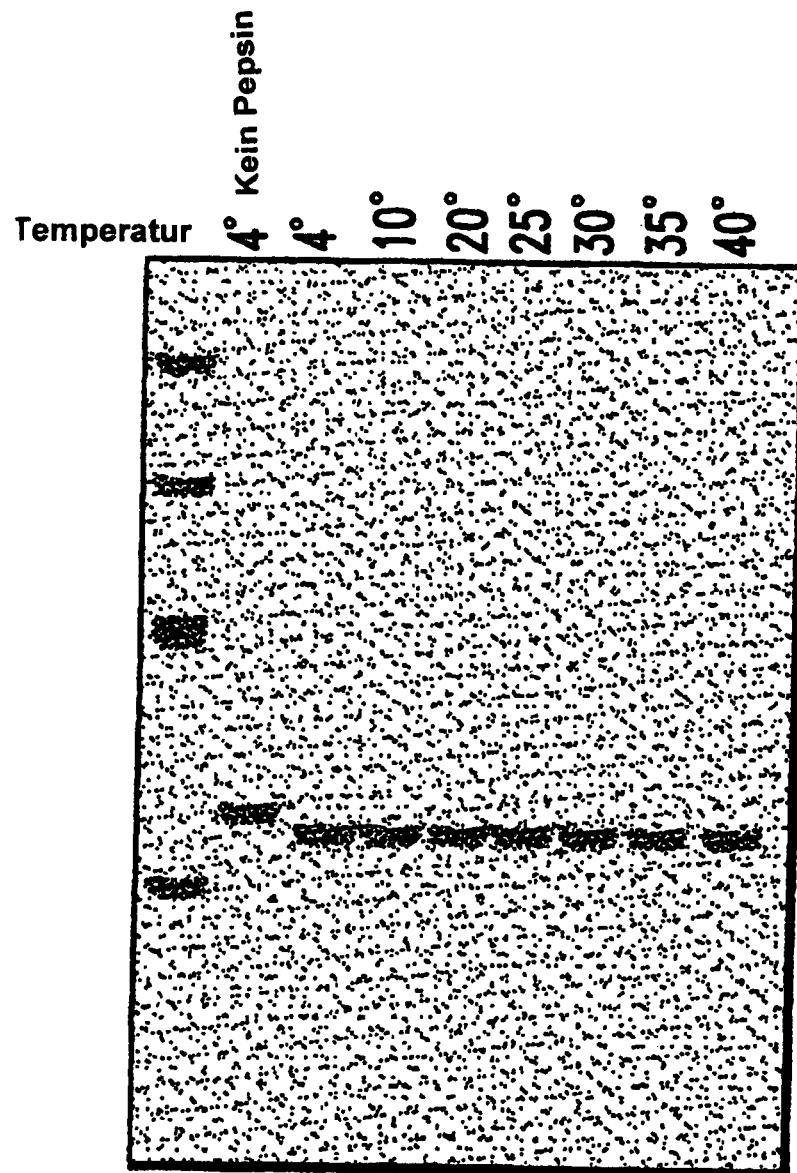


FIG. 34

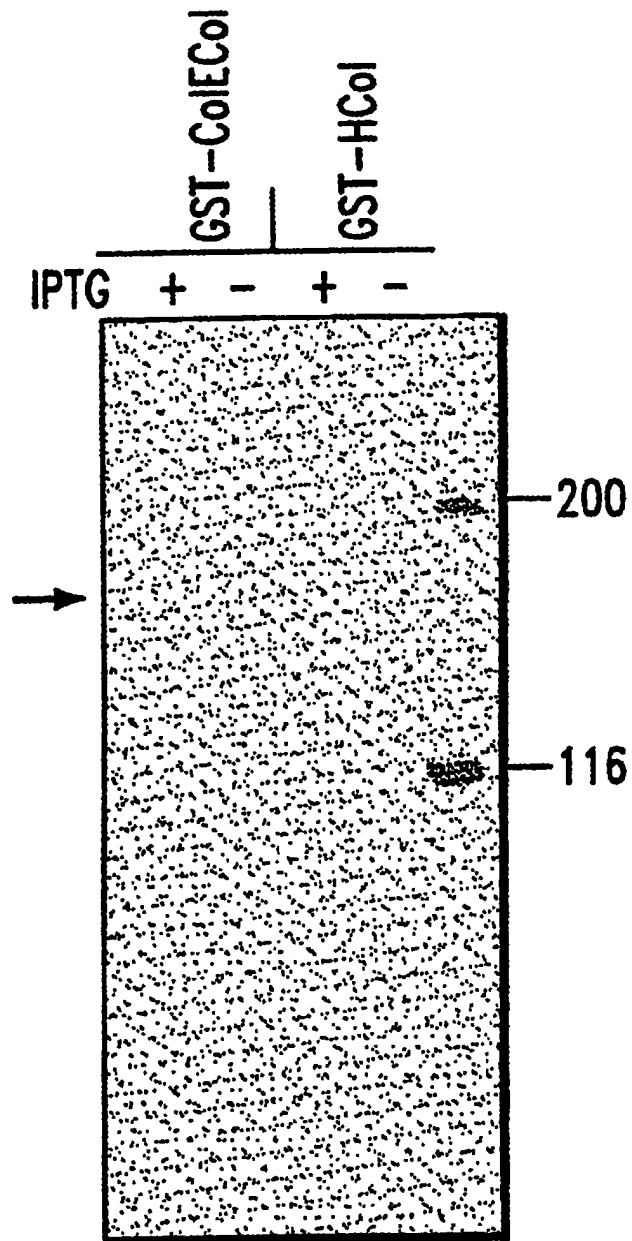


FIG. 35

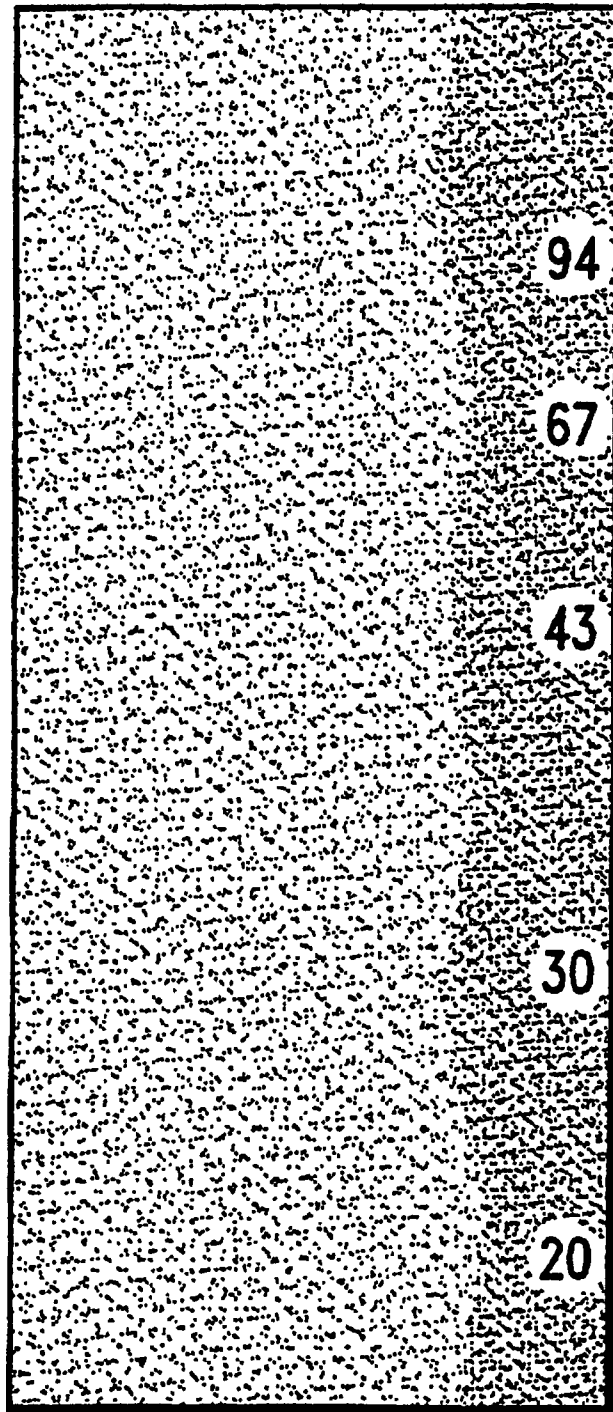


FIG. 36

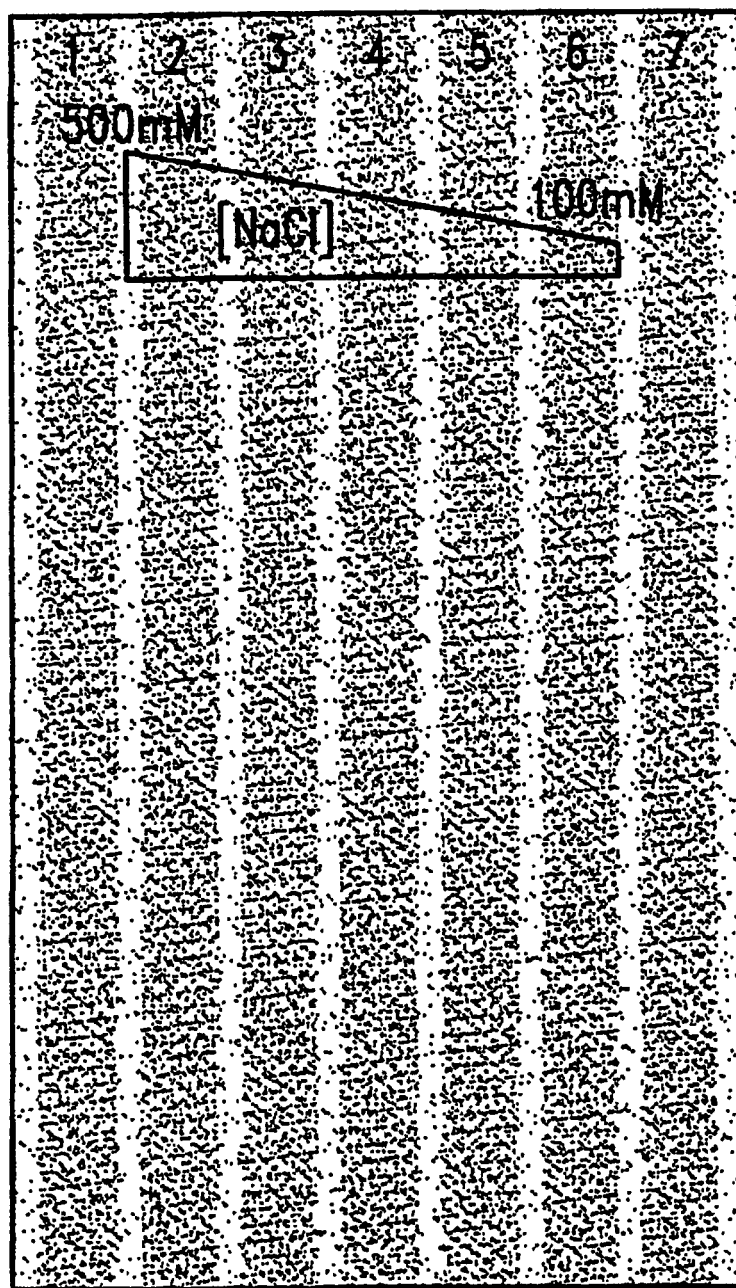


FIG. 37

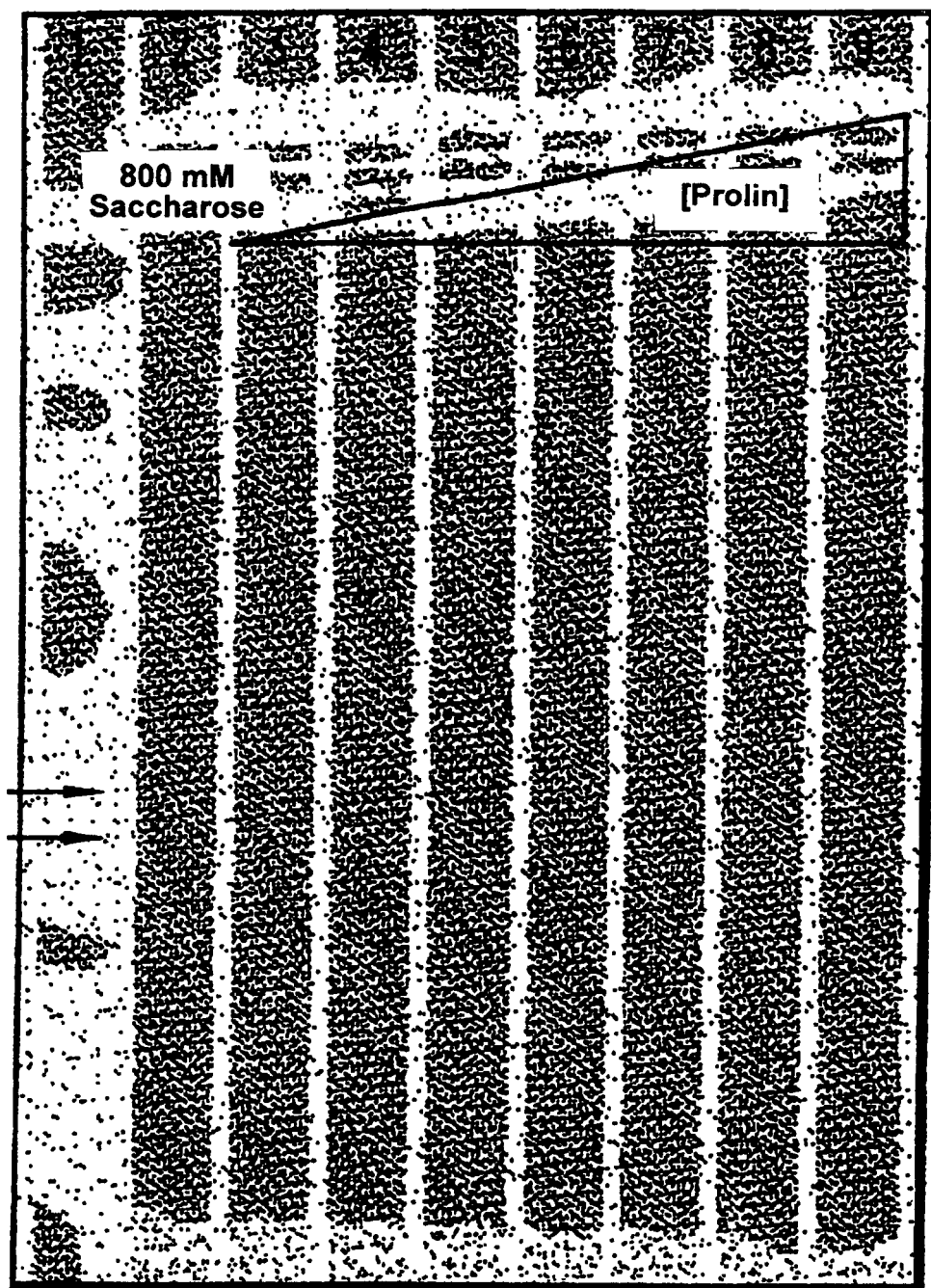


FIG. 38

```

5'   9      18      27      36      45      54
    CAG CTG AGC TAT GGC TAT GAT GAA AAA AGC ACC GGC GGC ATC AGC GTG CCG GGC
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
    Gln Leu Ser Tyr Gly Tyr Asp Glu Lys Ser Thr Gly Gly Ile Ser Val Pro Gly

      63      72      81      90      99      108
    CCG ATG GGT CCG AGC GGC CCG CGT GGC CTG CCG GGC CCG CCA GGT GCG CCC GGT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
    Pro Met Gly Pro Ser Gly Pro Arg Gly Leu Pro Gly Pro Pro Gly Ala Pro Gly

      117      126      135      144      153      162
    CCG CAG GGC TTT CAG GGT CCG CCG GGC GAA CCG GGC GAA OCT GGT GCG AGC GGC
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
    Pro Gln Gly Phe Gln Gly Pro Pro Gly Glu Pro Gly Glu Pro Gly Ala Ser Gly

      171      180      189      198      207      216
    CCG ATG GGC CCG CGC GGC CCG CCG GGT CCG CCA GGC AAA AAC GGC GAT GAT GGC
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
    Pro Met Gly Pro Arg Gly Pro Pro Gly Pro Pro Gly Lys Asn Gly Asp Asp Gly

      225      234      243      252      261      270
    GAA GCG GGC AAA CCG GGA CGT CCG GGT GAA CGT GGC CCC CCG GGC CCG CAG GGC
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
    Glu Ala Gly Lys Pro Gly Arg Pro Gly Glu Arg Gly Pro Pro Gly Pro Gln Gly

      279      288      297      306      315      324
    GCG CGC GGA CTG CCG GGT ACT GCG GGA CTG CCG GGC ATG AAA GGC CAC CCG GGT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
    Ala Arg Gly Leu Pro Gly Thr Ala Gly Leu Pro Gly Met Lys Gly His Arg Gly

      333      342      351      360      369      378
    TTC TCT GGT CTG GAT GGT GCC AAA GGA GAC CCG GGT CCG GCG GGT CCG AAA GGT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
    Phe Ser Gly Leu Asp Gly Ala Lys Gly Asp Ala Gly Pro Ala Gly Pro Lys Gly

      387      396      405      414      423      432
    GAG CCG GGC AGC CCG GGC GAA AAC GGC GCG CCG GGT CAG ATG GGC CCG CGT GGC
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
    Glu Pro Gly Ser Pro Gly Glu Asn Gly Ala Pro Gly Gln Met Gly Pro Arg Gly

      441      450      459      468      477      486
    CTG CCT GGT GAA CGC GGT CGC CCG GGC GGC CCG GGC CCA GCT GGC GCA CGT GGC
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
    Leu Pro Gly Glu Arg Gly Arg Pro Gly Ala Pro Gly Pro Ala Gly Ala Arg Gly

      495      504      513      522      531      540
    AAC GAT GGT GCG ACC GGT GCG GGC GGT CCA CCG GGC CCG ACC GGC CCG GCG GGT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
    Asn Asp Gly Ala Thr Gly Ala Ala Gly Pro Pro Gly Pro Thr Gly Pro Ala Gly

      549      558      567      576      585      594
    CCC CCG GGC TTT CCG GGT GCG GTG GGT GCG AAA GGC GAA GCA GGT CCG CAG GGC
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
    Pro Pro Gly Phe Pro Gly Ala Val Gly Ala Lys Gly Glu Ala Gly Pro Gln Gly

      603      612      621      630      639      648
    CCG CGC GGC AGC GAG GGT CCT CAG GGC GTT CGT GGT GAA CCG GGC CCG CCG GGC
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
    Pro Arg Gly Ser Glu Gly Pro Gln Gly Val Arg Gly Glu Pro Gly Pro Pro Gly

      657      666      675      684      693      702
    CCG GCG GGT GCG GCG GGC CCG GCT GGT AAC CCT GGC GCG GAC GGT CAG CCA GGT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
    Pro Ala Gly Ala Ala Gly Pro Ala Gly Asn Pro Gly Ala Asp Gly Gln Pro Gly

```

FIG. 39A

711	720	729	733	747	756
CGC AAA GGT GGC AAC GGC GCG CCG GGT ATT GCA GGT GCA GCG GGC TTC CCG GGT					
Ala Lys Gly Ala Asn Gly Ala Pro Gly Ile Ala Gly Ala Pro Gly Phe Pro Gly					
765	774	783	792	801	810
GCC CGC GGC CCG TCC GGC CCG CAG GGC CCG GGC GGC CCG CCG GGC CCG AAA GGC					
Ala Arg Gly Pro Ser Gly Pro Gln Gly Pro Gly Gly Pro Pro Gly Pro Lys Gly					
819	828	837	846	855	864
AAC AGC GGT GAA CCG GGT GCG CCG GGC AGC AAA GGC GAC ACC GGT GCG AAA GGT					
Asn Ser Gly Glu Pro Gly Ala Pro Gly Ser Lys Gly Asp Thr Gly Ala Lys Gly					
873	882	891	900	909	918
GAA CCG GGC CCA GTG GGT GAT CAA GGC CCG CCG GGC CCG GCG GGC GAG GAA GGC					
Glu Pro Gly Pro Val Gly Val Gln Gly Pro Pro Gly Pro Ala Gly Glu Glu Gly					
927	936	945	954	963	972
AAA CCG GGT GGT CCG GGT GAA CCG GGC CCG ACC GGC CTG CCG GGC CCG CCG GGA					
Lys Arg Gly Ala Arg Gly Glu Pro Gly Pro Thr Gly Leu Pro Gly Pro Pro Gly					
981	990	999	1008	1017	1026
GAA CCG GGT GGC CCG GGT AGC CCG GGT TTT CCG GGC GCG GAT GGT GTG GCG GGC					
Glu Arg Gly Gly Pro Gly Ser Arg Gly Phe Pro Gly Ala Asp Gly Val Ala Gly					
1035	1044	1053	1062	1071	1080
CCG AAA GGT CCG GCG GGT GAA CCG GGT AGC CCG GGC CCG GCG GGC CCA AAA GGC					
Pro Lys Gly Pro Ala Gly Glu Arg Gly Ser Pro Gly Pro Ala Gly Pro Lys Gly					
1089	1098	1107	1115	1125	1134
AGC CCG GGC GAG GCA GGA CCG CCG GGT GAA GCG GGT CTC CCG GGC GCC AAA GGT					
Ser Pro Gly Glu Ala Gly Arg Pro Gly Glu Ala Gly Leu Pro Gly Ala Lys Gly					
1143	1152	1161	1170	1179	1188
CTG ACC GGC TCT CCG GGC AGC CCG GGT CCG GAT GGC AAA ACC GGC CCG CCG GGT					
Leu Thr Gly Ser Pro Gly Ser Pro Gly Pro Asp Gly Lys Thr Gly Pro Pro Gly					
1197	1206	1215	1224	1233	1242
CCG GCG GGC CAG GAT GGT CCG CCG GGC CCG CCG GGC CCG CCG GGT GCG CCG GGT					
Pro Ala Gly Gln Asp Gly Arg Pro Gly Pro Pro Gly Pro Pro Gly Ala Arg Gly					
1251	1260	1269	1278	1287	1296
CAG GCG GGT GTC ATG GGC TTT CCA GGC CCG AAA GGT GCG GCG GGT GAA CCG GGC					
Gln Ala Gly Val Met Gly Phe Pro Gly Pro Lys Gly Ala Ala Gly Glu Pro Gly					
1305	1314	1323	1332	1341	1350
AAA GCG GGC GAA CCG GGT GTC CCG GGT CCG CCG GGC GCT GTC GGG CCG GCG GGC					
Lys Ala Gly Glu Arg Gly Val Pro Gly Pro Pro Gly Ala Val Gly Pro Ala Gly					
1359	1368	1377	1386	1395	1404
AAA GAT GGC GAA GCG GGC GCG CAA GGC CCG CCG GGA CCA GCG GGT CCG GCG GGC					
Lys Asp Gly Glu Ala Gly Ala Gln Gly Pro Pro Gly Pro Ala Gly Pro Ala Gly					

FIG. 39B

1413	1422	1431	1440	1449	1459
GAG CCG GGT GAA CAG GGC CCG GCA GGC AGC CCG GGT TTC CAG GGT CTG CCG GGC					
---	---	---	---	---	---
Glu Arg Gly Glu Gln Gly Pro Ala Gly Ser Pro Gly Phe Gln Gly Leu Pro Gly					
1467	1476	1485	1494	1503	1512
OCT CCG GGT CCA CCG GGT GAA GCG GGC AAA CCG GCG GAA CAA GGT GTG CCG GGC					
---	---	---	---	---	---
Pro Ala Gly Pro Pro Gly Glu Ala Gly Lys Pro Gly Glu Gln Gly Val Pro Gly					
1521	1530	1539	1548	1557	1566
GAC CTG GGC GCC CCA GGC CCG AGC GGC CCG CCG GGC GAA CCG GGT TTC CCG GGC					
---	---	---	---	---	---
Asp Leu Gly Ala Pro Gly Pro Ser Gly Ala Arg Gly Glu Arg Gly Phe Pro Gly					
1575	1584	1593	1602	1611	1620
GAA CGT GGT GTG CAG GGC CCG CCG GGC CCG GGT CCG CCG GGC GCC AAC GGC					
---	---	---	---	---	---
Glu Arg Gly Val Gln Gly Pro Pro Gly Pro Ala Gly Pro Arg Gly Ala Asn Gly					
1629	1638	1647	1656	1665	1674
GCG CCG GGC AAC GAT GGT GCG AAA GGT GAT GCG GGT GCG CCA GGT GCG CCG GGC					
---	---	---	---	---	---
Ala Pro Gly Asn Asp Gly Ala Lys Gly Asp Ala Gly Ala Pro Gly Ala Pro Gly					
1683	1692	1701	1710	1719	1728
AGC CAG GGC GGC CCG GGC CTG CAA GGC ATG CCG GGT GAA CGT GGT GCC GCG GGT					
---	---	---	---	---	---
Ser Gln Gly Ala Pro Gly Leu Gln Gly Met Pro Gly Glu Arg Gly Ala Ala Gly					
1737	1746	1755	1764	1773	1782
CTA CCG GGT CCG AAA GGC GAC CCG GGT GAT GCG GGT CCA AAA GGT GCG GAT GGC					
---	---	---	---	---	---
Leu Pro Gly Pro Lys Gly Asp Arg Gly Asp Ala Gly Pro Lys Gly Ala Asp Gly					
1791	1800	1809	1818	1827	1836
TCC CCT GGC AAA GAT GGC GTT CGT GGT CTG ACC GGC CCG ATC GGC CCG CCG GGC					
---	---	---	---	---	---
Ser Pro Gly Lys Asp Gly Val Arg Gly Leu Thr Gly Pro Ile Gly Pro Pro Gly					
1845	1854	1863	1872	1881	1890
CCG GCA GGT GGC CCG GGT GAC AAA GGT GAA AGC GGT CCG AGC GGC CCA GCG GGC					
---	---	---	---	---	---
Pro Ala Gly Ala Pro Gly Asp Lys Gly Glu Ser Gly Pro Ser Gly Pro Ala Gly					
1899	1908	1917	1926	1935	1944
CCC ACT GGT GCG CGT GGT GGC CCG GGC GAC CGT GGT GAA CCG GGT CCG CCG GGC					
---	---	---	---	---	---
Pro Thr Gly Ala Arg Gly Ala Pro Gly Asp Arg Gly Glu Pro Gly Pro Pro Gly					
1953	1962	1971	1980	1989	1998
CCG GCG GGC TTT GCG GGC CCG CCA GGC GCT GAC GGC CAG CCG GGT GCG AAA GGC					
---	---	---	---	---	---
Pro Ala Gly Phe Ala Gly Pro Pro Gly Ala Asp Gly Gln Pro Gly Ala Lys Gly					
2007	2016	2025	2034	2043	2052
GAA CCG GGC GAT GCG GGT GCC AAA GGC GAC CCG GGT CCG CCG GGC CCT GCC GGC					
---	---	---	---	---	---
Glu Pro Gly Asp Ala Gly Ala Lys Gly Asp Ala Gly Pro Pro Gly Pro Ala Gly					
2061	2070	2079	2088	2097	2106
CCG GCG GGC CCG CCA GGC CCG ATT GGC AAC GTG GGT GCG CCG GGT GCC AAA GGT					
---	---	---	---	---	---
Pro Ala Gly Pro Pro Gly Pro Ile Gly Asn Val Gly Ala Pro Gly Ala Lys Gly					

FIG. 39C

2115	2124	2133	2142	2151	2160
GGC CGC GGC AGC GGT GGT CCG CCG GGC GCG ACC GGT TTC CCG GGT GCG GCG GCG					
Ala Arg Gly Ser Ala Gly Pro Pro Gly Ala Thr Gly Phe Pro Gly Ala Ala Gly					
2169	2178	2187	2196	2205	2214
CGC GTG GGT CCG CCA GGC CCG AGC GGT AAC GCG GGC CCG CCG GGC CCG CCG GGC					
Arg Val Gly Pro Pro Gly Pro Ser Gly Asn Ala Gly Pro Pro Gly Pro Pro Gly					
2223	2232	2241	2250	2259	2268
CCG GCG GGC AAA GAG GGC GGC AAA GGT CCG CGT GGT GAA ACC GGC CCT GCG GGA					
Pro Ala Gly Lys Glu Gly Gly Lys Gly Pro Arg Gly Glu Thr Gly Pro Ala Gly					
2277	2286	2295	2304	2313	2322
CGT CCA GGT GAA GTG GGT CCG CCG GGC CCG CCG GGC CCG GCG GGC GAA AAA GGT					
Arg Pro Gly Glu Val Gly Pro Pro Gly Pro Pro Gly Pro Ala Gly Glu Lys Gly					
2331	2340	2349	2358	2367	2376
AGC CCG GGT GCG GAT GGT CCG CCG GGT GCG CCA GGC ACG CCG GGT CCG CAA GGT					
Ser Pro Gly Ala Asp Gly Pro Ala Gly Ala Pro Gly Thr Pro Gly Pro Gln Gly					
2385	2394	2403	2412	2421	2430
ATC GGT GGC CAG CGT GGT GTC GTC GGC CTG CCG GGT CAG CCG GGC GAA CCG GGC					
Ile Ala Gly Gln Arg Gly Val Val Gly Leu Pro Gly Gln Arg Gly Glu Arg Gly					
2439	2448	2457	2466	2475	2484
TTT CCG GGT CTG CCG GGC CCG AGC GGT GAG CCG GGC AAA CAG GGT CCA TCT GGC					
Phe Pro Gly Leu Pro Gly Pro Ser Gly Glu Pro Gly Lys Gln Gly Pro Ser Gly					
2493	2502	2511	2520	2529	2538
GCG AGC GGT GAA CGT GGC CCG CCG GGT CCC ATG GGC CCG CCG GGT CTG GCG GGC					
Ala Ser Gly Glu Arg Gly Pro Pro Gly Pro Met Gly Pro Pro Gly Leu Ala Gly					
2547	2556	2565	2574	2583	2592
CCT CCG GGT GAA AGC GGT CGT GAA GGC GCG CCG GGT GCG GAA GGC AGC CCA GGC					
Pro Pro Gly Glu Ser Gly Arg Glu Gly Ala Pro Gly Ala Glu Gly Ser Pro Gly					
2601	2610	2619	2628	2637	2646
GCG GAC GGT AGC CCG GGC GGC AAA GGC GAT CGT GGT GAA ACC GGC CCG GCG GGC					
Arg Asp Gly Ser Pro Gly Ala Lys Gly Asp Arg Gly Glu Thr Gly Pro Ala Gly					
2655	2664	2673	2682	2691	2700
CCC CCG GGT GCA CCG GGC GCG CCG GGT GGC CCA GGC CCG GTG GGC CCG GCG GGC					
Pro Pro Gly Ala Pro Gly Ala Pro Gly Ala Pro Gly Pro Val Gly Pro Ala Gly					
2709	2718	2727	2736	2745	2754
AAA AGC GGT GAT CGT GGT GAG ACC GGT CCG GCG GGC CCG GCG GGT CCG GTG GGC					
Lys Ser Gly Asp Arg Gly Glu Thr Gly Pro Ala Gly Pro Ala Gly Pro Val Gly					
2763	2772	2781	2790	2799	2808
CCA GCG GGC GCG CGT GGC CCG GCG GGT CCG CAG GGC CCG CCG GGT GAC AAA GGT					
Pro Ala Gly Ala Arg Gly Pro Ala Gly Pro Gln Gly Pro Arg Gly Asp Lys Gly					

FIG. 39D

2817	2826	2835	2844	2853	2862
GAA ACG GGC GAA CAG GGC GAC CGT GGC ATT AAA GGC CAC CGT GGC TTC AGC GGC					
---	---	---	---	---	---
Glu Thr Gly Glu Gln Gly Asp Arg Gly Ile Lys Gly His Arg Gly Phe Ser Gly					
2871	2880	2889	2898	2907	2916
CTG CAG GGT CCA CCG GGC CCG CCG GGC AGT CCG GGT GAA CAG GGT CCG TCC GGA					
---	---	---	---	---	---
Leu Gln Gly Pro Pro Gly Pro Pro Gly Ser Pro Gly Glu Gln Gly Pro Ser Gly					
2925	2934	2943	2952	2961	2970
GCC AGC GGG CCG GCG GGC CCA CCG GGT CCG CCG GGC AGC GCG GGT GCG CCG GGC					
---	---	---	---	---	---
Ala Ser Gly Pro Ala Gly Pro Arg Gly Pro Pro Gly Ser Ala Gly Ala Pro Gly					
2979	2988	2997	3006	3015	3024
AAA GAC GGT CTG AAC GGT CTG CCG GGC CCG ATC GGC CCG CCG GGC CCA CCG GGC					
---	---	---	---	---	---
Lys Asp Gly Leu Asn Gly Leu Pro Gly Pro Ile Gly Pro Pro Gly Pro Arg Gly					
3033	3042	3051	3060	3069	3078
CGC ACC GGT GAT GCG GGT CCG GTG GGT CCG CCG GGC CCG CCG GGC CCG CCA GGC					
---	---	---	---	---	---
Arg Thr Gly Asp Ala Gly Pro Val Gly Pro Pro Gly Pro Pro Gly Pro Pro Gly					
3087	3096	3105	3114	3123	3132
CCG CCG GGT CCG CCG AGC GCG GGT TTC GAC TTC AGC TTC CTG CCG CAG CCG CCG					
---	---	---	---	---	---
Pro Pro Gly Pro Pro Ser Ala Gly Phe Asp Phe Ser Phe Leu Pro Gln Pro Pro					
3141	3150	3159	3168		
CAG GAG AAA CCG CAC GAC GGC GGT CCG TAC TAC CCG GCG 3'					
---	---	---	---		
Gln Glu Lys Ala His Asp Gly Gly Arg Tyr Tyr Arg Ala					

FIG. 39E

FIG. 40

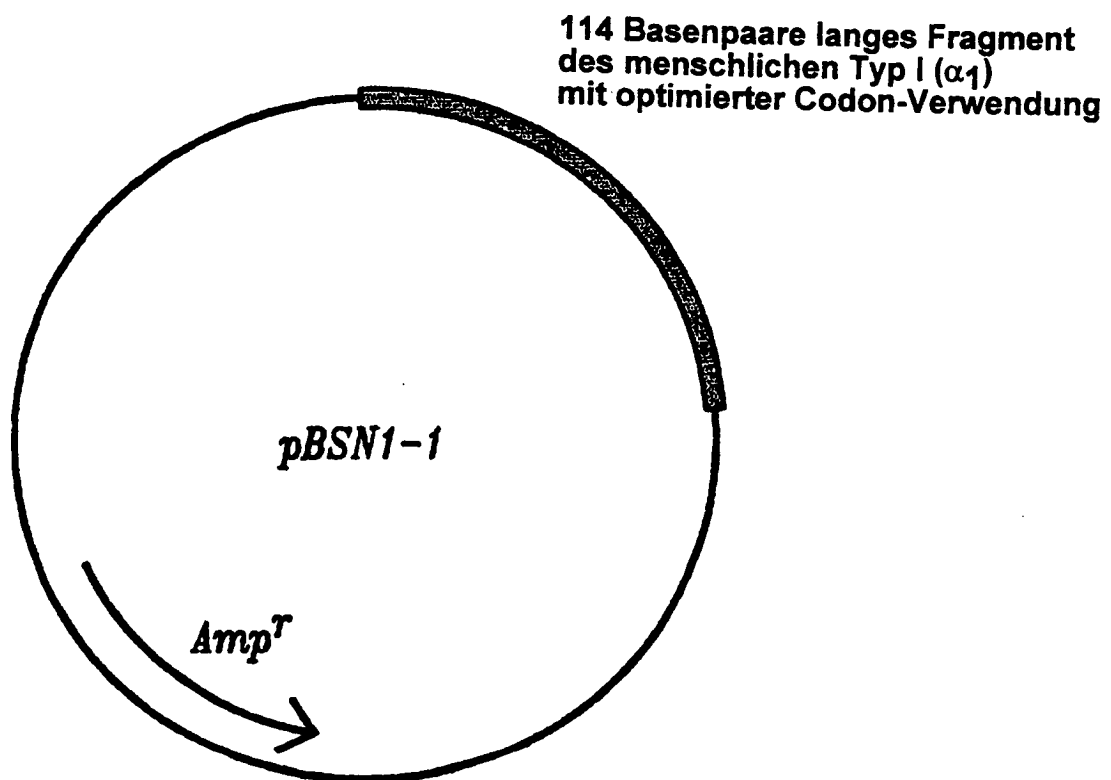


FIG. 4I


```

          9      18      27      36      45      54
5' CAG CTG AGC TAT GGC TAT GAT GAA AAA AGC ACC GGC GGC ATC AGC GTS CCG GGC
   --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
   Gln Leu Ser Tyr Gly Tyr Asp Glu Lys Ser Thr Gly Gly Ile Ser Val Pro Gly

          63      72      81      90      99      108
   CCG ATG GGT CCG AGC GGC CCG CGT GGC CTG CCG GGC CCG CCA GGT GCG CCG GGT
   --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
   Pro Met Gly Pro Ser Gly Pro Arg Gly Leu Pro Gly Pro Pro Gly Ala Pro Gly

CCG
---
Pro

```

FIG. 42

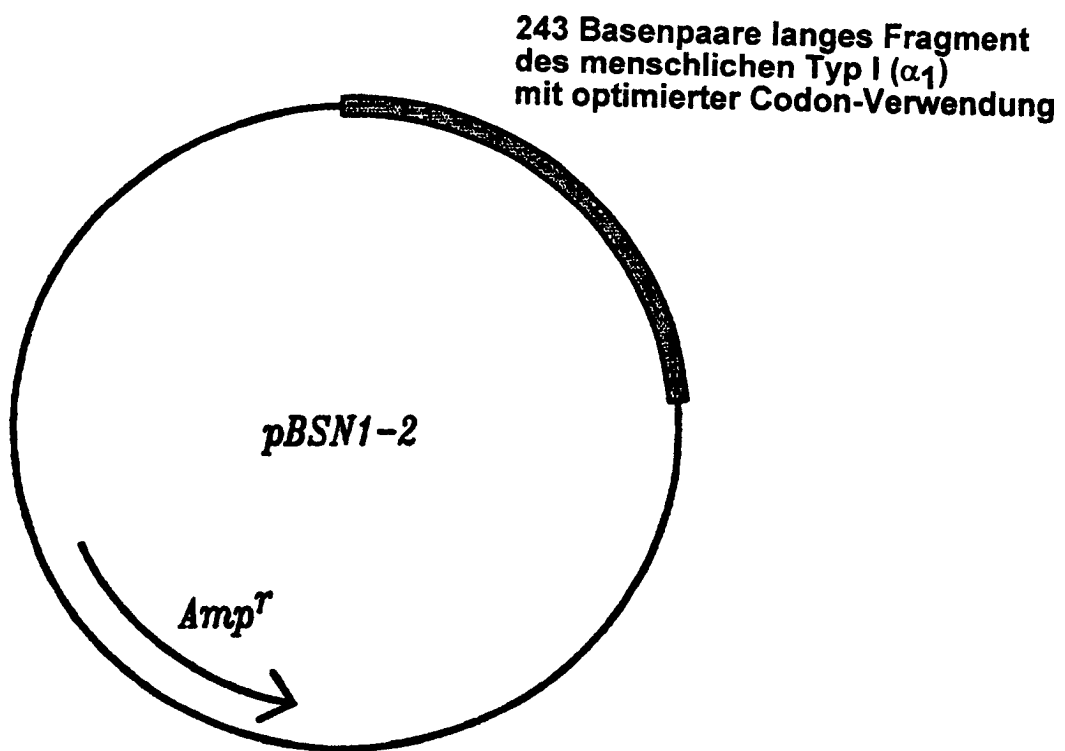


FIG. 43

```

          9      18      27      36      45      54
5' CAG CTG AGC TAT GGC TAT GAT GAA AAA AGC ACC GGC GGC ATC AGC GTG CCG GGC
   --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
   Gln Leu Ser Tyr Gly Tyr Asp Glu Lys Ser Thr Gly Gly Ile Ser Val Pro Gly

          63      72      81      90      99      108
   CCG ATG GGT CCG AGC GGC CCG CGT GGC CTG CCG GGC CCG CCA GGT GCG CCC GGT
   --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
   Pro Met Gly Pro Ser Gly Pro Arg Gly Leu Pro Gly Pro Pro Gly Ala Pro Gly

          117      126      135      144      153      162
   CCG CAG GGC TTT CAG GGT CCG CCG GGC GAA CCG GGC GAA CCT GGT GCG AGC GGC
   --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
   Pro Gln Gly Phe Gln Gly Pro Pro Gly Glu Pro Gly Glu Pro Gly Ala Ser Gly

          171      180      189      198      207      216
   CCG ATG GGC CCG CCG GGC CCG CCG GGT CCG CCA GGC AAA AAC GGC GAT GAT GGC
   --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
   Pro Met Gly Pro Arg Gly Pro Pro Gly Pro Pro Gly Lys Asn Gly Asp Asp Gly

          225      234
   GAA GCG GGC AAA CCG GGA CGT CCG
   --- --- --- --- --- --- --- ---
   Glu Ala Gly Lys Pro Gly Arg Pro

```

FIG. 44

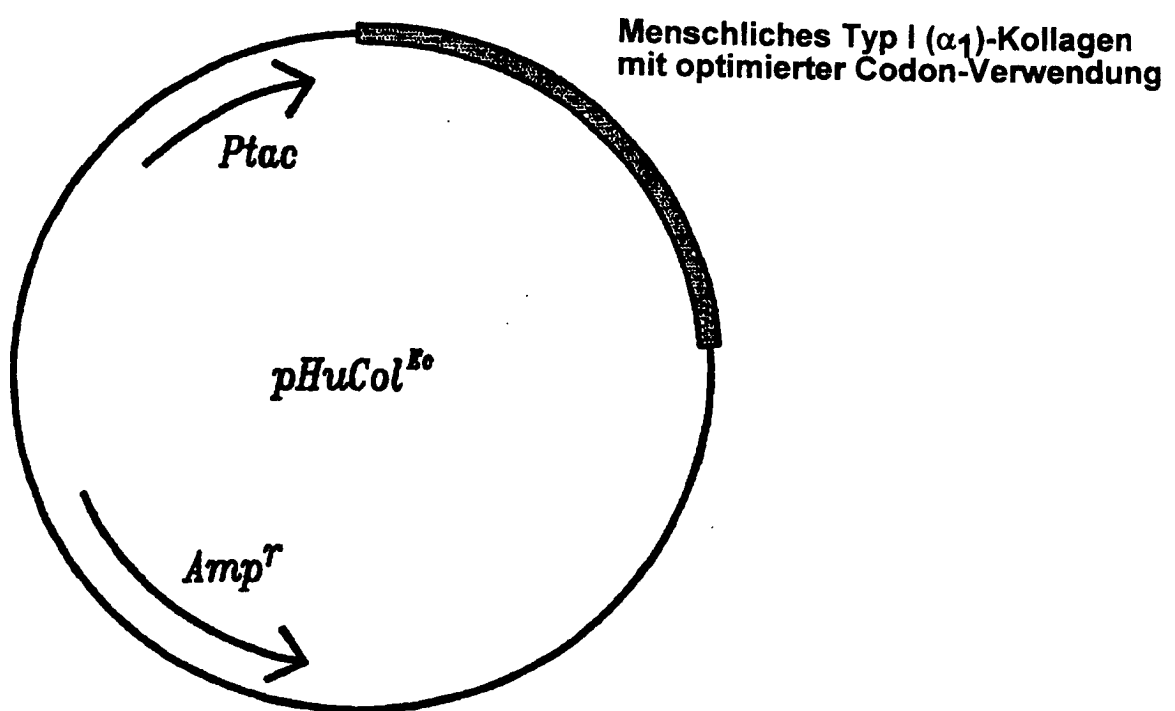


FIG. 45

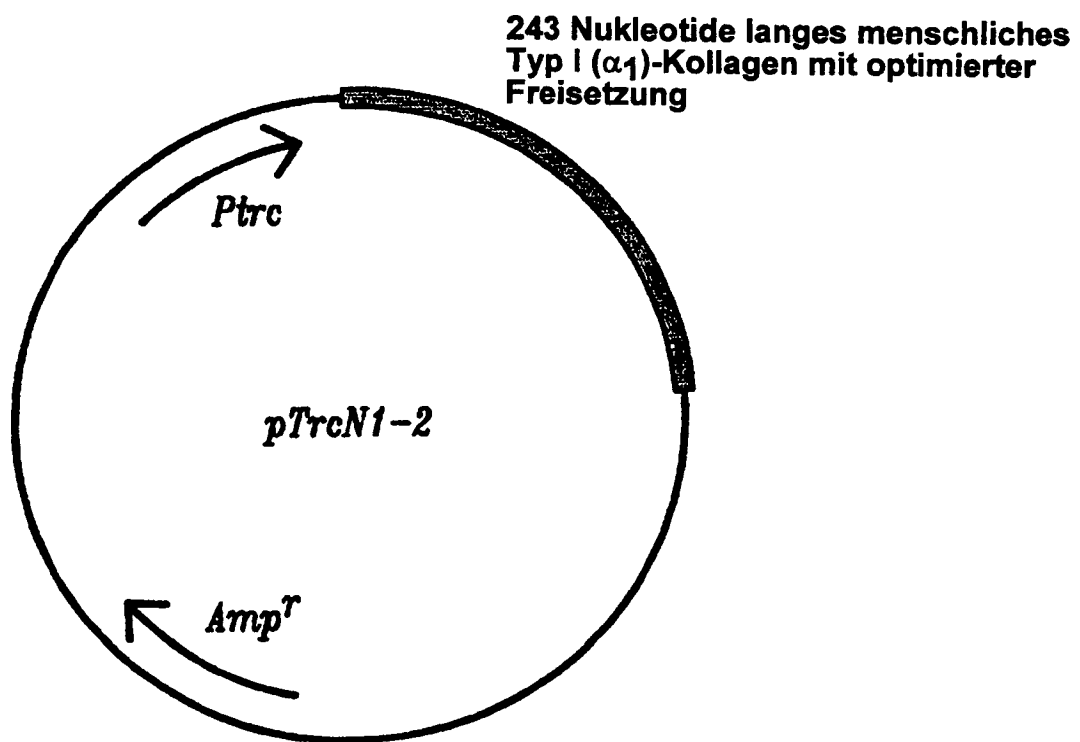


FIG. 46

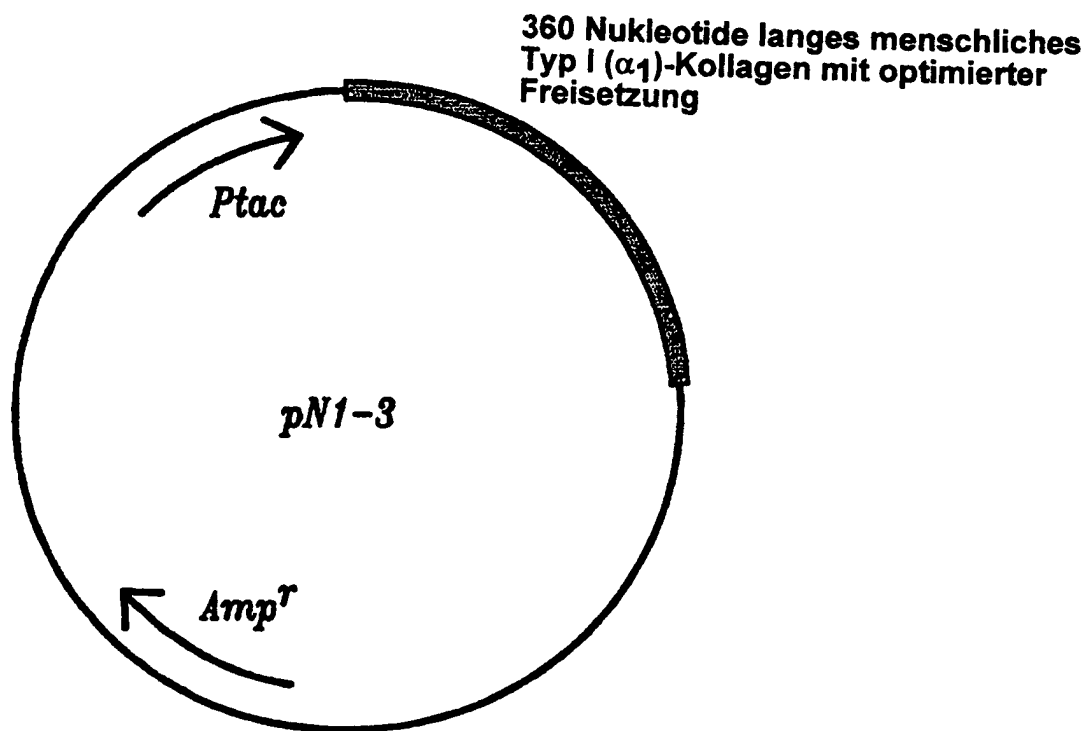


FIG. 47

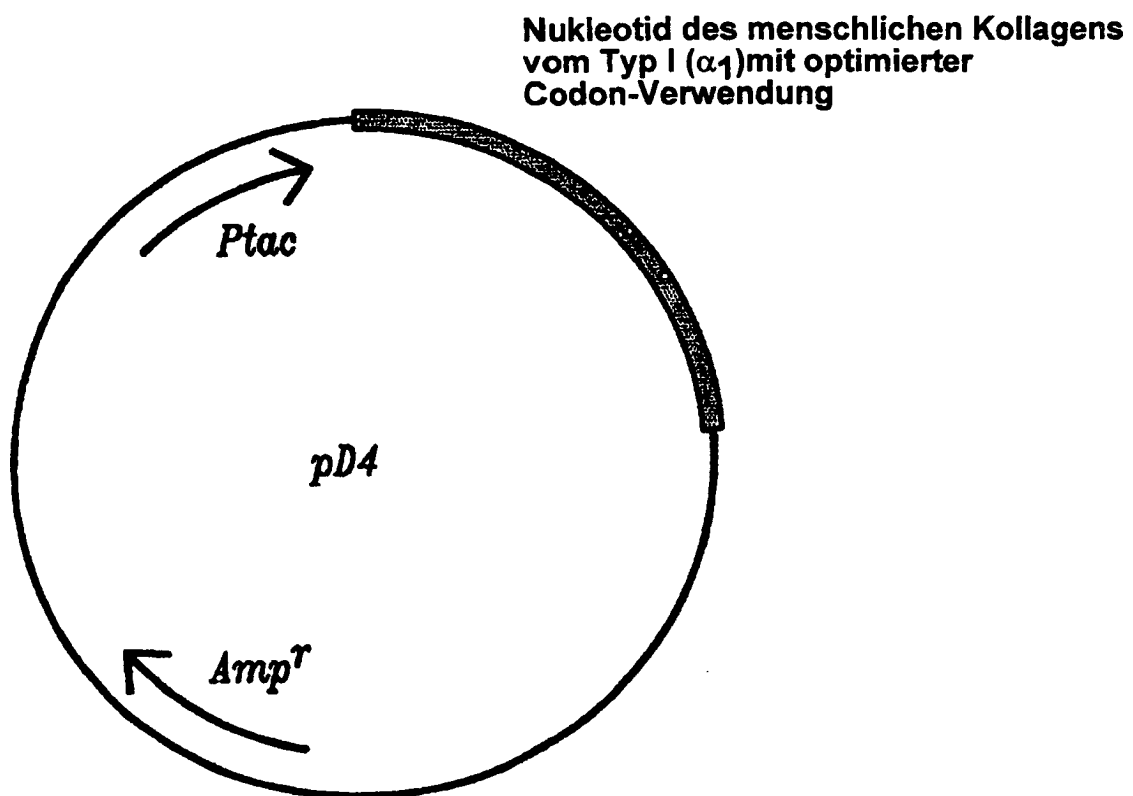


FIG. 48

```

          9      18      27      36      45      54
5'  CAG TAT CAT GGA AAA GGA GTT GGA CTT GGC OCT GGA CCA ATG GGC TTA ATG GGA
    --- --- --- --- --- --- --- --- --- --- --- ---
    Gln Tyr Asp Gly Lys Gly Val Gly Leu Gly Pro Gly Pro Met Gly Leu Met Gly

          63      72      81      90      99      108
    OCT AGA GGC CCA OCT GGT GCA GCT GGA GGC CCA GGC OCT CAA GGT TTC CAA GGA
    --- --- --- --- --- --- --- --- --- --- --- ---
    Pro Arg Gly Pro Pro Gly Ala Ala Gly Ala Pro Gly Pro Gln Gly Phe Gln Gly

          117      126      135      144      153      162
    OCT GCT GGT GAG OCT GGT GAA OCT GGT CAA ACT GGT OCT GCA GGT GCT OCT GGT
    --- --- --- --- --- --- --- --- --- --- --- ---
    Pro Ala Gly Glu Pro Gly Glu Pro Gly Gln Thr Gly Pro Ala Gly Ala Arg Gly

          171      180      189      198      207      216
    CCA GCT GGC OCT OCT GGC AAG GCT GGT GAA GAT GGT CAC OCT GGA AAA CCC GGA
    --- --- --- --- --- --- --- --- --- --- --- ---
    Pro Ala Gly Pro Pro Gly Lys Ala Gly Glu Asp Gly His Pro Gly Lys Pro Gly

          225      234      243      252      261      270
    CGA OCT GGT GAG AGA GGA GTT GTT GGA CCA CAG GGT GCT OCT GGT TTC OCT GGA
    --- --- --- --- --- --- --- --- --- --- --- ---
    Arg Pro Gly Glu Arg Gly Val Val Gly Pro Gln Gly Ala Arg Gly Phe Pro Gly

          279      288      297      306      315      324
    ACT OCT GGA CTT OCT GGC TTC AAA GGC ATT AGG GGA CAC AAT GGT CTG GAT GGA
    --- --- --- --- --- --- --- --- --- --- --- ---
    Thr Pro Gly Leu Pro Gly Phe Lys Gly Ile Arg Gly His Asn Gly Leu Asp Gly

          333      342      351      360      369      378
    TTG AAG GGA CAG CCC GGT GCT OCT GGT GTG AAG GGT GAA OCT GGT GGC OCT GGT
    --- --- --- --- --- --- --- --- --- --- --- ---
    Leu Lys Gly Gln Pro Gly Ala Pro Gly Val Lys Gly Glu Pro Gly Ala Pro Gly

          387      396      405      414      423      432
    GAA AAT GGA ACT CCA GGT CAA ACA GGA GGC OCT GGG CTT OCT GGT GAG AGA GGA
    --- --- --- --- --- --- --- --- --- --- --- ---
    Glu Asn Gly Thr Pro Gly Gln Thr Gly Ala Arg Gly Leu Pro Gly Glu Arg Gly

          441      450      459      468      477      486
    CGT GTT GGT GGC OCT GGC CCA GCT GGT GGC OCT GGC AGT GAT GGA AGT GTG GGT
    --- --- --- --- --- --- --- --- --- --- --- ---
    Arg Val Gly Ala Pro Gly Pro Ala Gly Ala Arg Gly Ser Asp Gly Ser Val Gly

          495      504      513      522      531      540
    CCC GTG GGT OCT GCT GGT CCC ATT GGG TCT GCT GGC OCT CCA GGC TTC CCA GGT
    --- --- --- --- --- --- --- --- --- --- --- ---
    Pro Val Gly Pro Ala Gly Pro Ile Gly Ser Ala Gly Pro Pro Gly Phe Pro Gly

          549      558      567      576      585      594
    GGC OCT GGC CCC AAG GGT GAA ATT GGA GCT GTT GGT AAC GCT GGT OCT GCT GGT
    --- --- --- --- --- --- --- --- --- --- --- ---
    Ala Pro Gly Pro Lys Gly Glu Ile Gly Ala Val Gly Asn Ala Gly Pro Ala Gly

          603      612      621      630      639      648
    CCC GGC GGT CCC CGT GGT GAA GTG GGT CTT CCA GGC CTC TCC GGC CCC GTT GGA
    --- --- --- --- --- --- --- --- --- --- --- ---
    Pro Ala Gly Pro Arg Gly Glu Val Gly Leu Pro Gly Leu Ser Gly Pro Val Gly

          657      666      675      684      693      702
    CCT CCT GGT AAT OCT GGA GCA AAC GGC CTT ACT GGT GGC AAG GGT GCT GCT GGC
    --- --- --- --- --- --- --- --- --- --- --- ---
    Pro Pro Gly Asn Pro Gly Ala Asn Gly Leu Thr Gly Ala Lys Gly Ala Ala Gly

```

FIG. 49A

711	720	729	738	747	756
CTT CCC GGC GTT GCT GGG GCT CCC GGC CTC OCT GGA CCC GGC GGT ATT CCT GGC					
Leu Pro Gly Val Ala Gly Ala Pro Gly Leu Pro Gly Pro Arg Gly Ile Pro Gly					
765	774	783	792	801	810
OCT GTT GGT GCT GGC GGT GCT ACT GGT GGC AGA GGA CTT GTT GGT GAG OCT GGT					
Pro Val Gly Ala Ala Gly Ala Thr Gly Ala Arg Gly Leu Val Gly Glu Pro Gly					
819	828	837	846	855	864
CCA GCT GGC TOC AAA GGA GAG AGC GGT AAC AAG GGT GAG CCC GGC TCT GCT GGC					
Pro Ala Gly Ser Lys Gly Glu Ser Gly Asn Lys Gly Glu Pro Gly Ser Ala Gly					
873	882	891	900	909	918
CCC CAA GGT CCT OCT GGT CCC AGT GGT GAA GAA GGA AAG AGA GGC OCT AAT GGC					
Pro Gln Gly Pro Pro Gly Pro Ser Gly Glu Glu Gly Lys Arg Gly Pro Asn Gly					
927	936	945	954	963	972
GAA GCT GGA TCT GGC GGC OCT CCA GGA OCT OCT GGG CTG AGA GGT AGT OCT GGT					
Glu Ala Gly Ser Ala Gly Pro Pro Gly Pro Pro Gly Leu Arg Gly Ser Pro Gly					
981	990	999	1008	1017	1026
TCT OCT GGT CTT OCT GGA GCT GAT GGC AGA GCT GGC GTC ATG GGC OCT OCT GGT					
Ser Arg Gly Leu Pro Gly Ala Asp Gly Arg Ala Gly Val Met Gly Pro Pro Gly					
1035	1044	1053	1062	1071	1080
AGT OCT GGT CCA AGT GGC OCT GCT GGA GTC CCA GGA OCT AAT GGA GAT GCT GGT					
Ser Arg Gly Ala Ser Gly Pro Ala Gly Val Arg Gly Pro Asn Gly Asp Ala Gly					
1089	1098	1107	1116	1125	1134
CGC OCT GGC GAG OCT GGT CTC ATG GGA CCC AGA GGT CTT OCT GGT TOC OCT GGA					
Arg Pro Gly Glu Pro Gly Leu Met Gly Pro Arg Gly Leu Pro Gly Ser Pro Gly					
1143	1152	1161	1170	1179	1188
AAT ATC GGC CCC GCT GGA AAA GAA GGT OCT GTC GGC CTC OCT GGC ATC GAC GGC					
Asn Ile Gly Pro Ala Gly Lys Glu Gly Pro Val Gly Leu Pro Gly Ile Asp Gly					
1197	1206	1215	1224	1233	1242
AGG OCT GGC CCA ATT GGC CCA GCT GGA CCA AGA GGA GAG OCT GGC AAC ATT GGA					
Arg Pro Gly Pro Ile Gly Pro Ala Gly Ala Arg Gly Glu Pro Gly Asn Ile Gly					
1251	1260	1269	1278	1287	1296
TTC OCT GGA CCC AAA GGC CCC ACT GGT GAT OCT GGC AAA AAC GGT GAT AAA GGT					
Phe Pro Gly Pro Lys Gly Pro Thr Gly Asp Pro Gly Lys Asn Gly Asp Lys Gly					
1305	1314	1323	1332	1341	1350
CAT GCT GGT CTT GCT GGT GCT CGG GGT GCT CCA GGT CCT GAT GGA AAC AAT GGT					
His Ala Gly Leu Ala Gly Ala Arg Gly Ala Pro Gly Pro Asp Gly Asn Asn Gly					
1359	1368	1377	1386	1395	1404
GCT CAG GGA OCT OCT GGA CCA CAG GGT GTT CAA GGT GGA AAA GGT GAA CAG GGT					

FIG. 49B

Ala Gln Gly Pro Pro Gly Pro Gln Gly Val Gln Gly Gly Lys Gly Glu Gln Gly

1413 1422 1431 1440 1449 1458
 --- --- --- --- --- ---
 OCC GAT GGT OCT CCA GGC TTC CAG GGT CTG OCT GGC CCC TCA GGT CCC GCT GGT
 --- --- --- --- --- ---
 Pro Asp Gly Pro Pro Gly Phe Gln Gly Leu Pro Gly Pro Ser Gly Pro Ala Gly

1467 1476 1485 1494 1503 1512
 --- --- --- --- --- ---
 GAA GTT GGC AAA CCA GGA GAA AGG GGT CTC CAT GGT GAG TTT GGT CTC OCT GGT
 --- --- --- --- --- ---
 Glu Val Gly Lys Pro Gly Glu Arg Gly Leu His Gly Glu Phe Gly Leu Pro Gly

1521 1530 1539 1548 1557 1566
 --- --- --- --- --- ---
 CCT GCT GGT CCA AGA GGG GAA CGC GGT CCC CCA GGT GAG AGT GGT GCT GOC GGT
 --- --- --- --- --- ---
 Pro Ala Gly Pro Arg Gly Glu Arg Gly Pro Pro Gly Glu Ser Gly Ala Ala Gly

1575 1584 1593 1602 1611 1620
 --- --- --- --- --- ---
 OCT ACT GGT OCT ATT GGA AGC CGA GGT OCT TCT GGA CCC CCA GGG OCT GAT GGA
 --- --- --- --- --- ---
 Pro Thr Gly Pro Ile Gly Ser Arg Gly Pro Ser Gly Pro Pro Gly Pro Asp Gly

1629 1638 1647 1656 1665 1674
 --- --- --- --- --- ---
 AAC AAG GGT GAA OCT GGT GTG GTT GGT GCT GTG GGC ACT GCT GGT CCA TCT GGT
 --- --- --- --- --- ---
 Asn Lys Gly Glu Pro Gly Val Val Gly Ala Val Gly Thr Ala Gly Pro Ser Gly

1683 1692 1701 1710 1719 1728
 --- --- --- --- --- ---
 OCT AGT GGA CTC CCA GGA GAG AGG GGT GCT GCT GGC ATA OCT GGA GGC AAG GGA
 --- --- --- --- --- ---
 Pro Ser Gly Leu Pro Gly Glu Arg Gly Ala Ala Gly Ile Pro Gly Gly Lys Gly

1737 1746 1755 1764 1773 1782
 --- --- --- --- --- ---
 GAA AAG GGT GAA OCT GGT CTC AGA GGT GAA ATT GGT AAC OCT GGC AGA GAT GGT
 --- --- --- --- --- ---
 Glu Lys Gly Glu Pro Gly Leu Arg Gly Glu Ile Gly Asn Pro Gly Arg Asp Gly

1791 1800 1809 1818 1827 1836
 --- --- --- --- --- ---
 GCT CGT GGT GCT CAT GGT GCT GTA GGT GOC OCT GGT OCT GCT GGA GOC ACA GGT
 --- --- --- --- --- ---
 Ala Arg Gly Ala His Gly Ala Val Gly Ala Pro Gly Pro Ala Gly Ala Thr Gly

1845 1854 1863 1872 1881 1890
 --- --- --- --- --- ---
 GAC CGG GGC GAA GCT GGG GCT GCT GGT OCT GCT GGT OCT GCT GGT OCT CGG GGA
 --- --- --- --- --- ---
 Asp Arg Gly Glu Ala Gly Ala Ala Gly Pro Ala Gly Pro Ala Gly Pro Arg Gly

1899 1908 1917 1926 1935 1944
 --- --- --- --- --- ---
 AGC OCT GGT GAA CGT GGC GAG GTC GGT OCT GCT GGC CCC AAC GGA TTT GCT GGT
 --- --- --- --- --- ---
 Ser Pro Gly Glu Arg Gly Glu Val Gly Pro Ala Gly Pro Asn Gly Phe Ala Gly

1953 1962 1971 1980 1989 1998
 --- --- --- --- --- ---
 COG GCT GGT GCT GCT GGT CAA COG GGT GCT AAA GGA GAA AGA GGA GGC AAA GGG
 --- --- --- --- --- ---
 Pro Ala Gly Ala Ala Gly Gln Pro Gly Ala Lys Gly Glu Arg Gly Ala Lys Gly

2007 2016 2025 2034 2043 2052
 --- --- --- --- --- ---
 OCT AAG GGT GAA AAC GGT GTT GTT GGT CCC ACA GGC CCC GTT GGA GCT GCT GGC
 --- --- --- --- --- ---
 Pro Lys Gly Glu Asn Gly Val Val Gly Pro Thr Gly Pro Val Gly Ala Ala Gly

2061 2070 2079 2088 2097 2106
 --- --- --- --- --- ---
 CCN NNN GGT CCA AAT GGT CCC CCC GGT OCT GCT GGA AGT CGT GGT GAT GGA GGC

FIG. 49C

```

--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Pro Xxx Gly Pro Asn Gly Pro Pro Gly Pro Ala Gly Ser Arg Gly Asp Gly Gly
      2115      2124      2133      2142      2151      2160
COC OCT GGT ATG ACT GGT TTC OCT GGT GCT GCT GGA CCG ACT GGT COC CCA GGA
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Pro Pro Gly Met Thr Gly Phe Pro Gly Ala Ala Gly Arg Thr Gly Pro Pro Gly

      2169      2178      2187      2196      2205      2214
COC TCT GGT ATT TCT GGC OCT OCT GGT COC OCT GGT CCT GCT GGG AAA GAA GGG
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Pro Ser Gly Ile Ser Gly Pro Pro Gly Pro Pro Gly Pro Ala Gly Lys Glu Gly

      2223      2232      2241      2250      2259      2268
CTT CGT GGA CCN CGA GGN GAC CAA GGA CCA GCA GGC CGA CCT GGA GAA GTA GGA
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Leu Arg Gly Pro Arg Gly Asp Gln Gly Pro Ala Gly Arg Pro Gly Glu Val Gly

      2277      2286      2295      2304      2313      2322
GCA CCG GGT CCC OCT GGC TTC GCT GGT GAG AAG GGT COC TCT GGA GAG GCT GGT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Ala Pro Gly Pro Pro Gly Phe Ala Gly Glu Lys Gly Pro Ser Gly Glu Ala Gly

      2331      2340      2349      2358      2367      2376
ACT GCT GGA OCT OCT GGC ACT CCA GGT OCT CAG GGT CTT CTT GGT GCT OCT GGT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Thr Ala Gly Pro Pro Gly Thr Pro Gly Pro Gln Gly Leu Leu Gly Ala Pro Gly

      2385      2394      2403      2412      2421      2430
ATT CTG GGT CTC OCT GGC TCG AGA GGT GAA CGT GGT CTA OCT GGT GTT GCT GGT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Ile Leu Gly Leu Pro Gly Ser Arg Gly Glu Arg Gly Leu Pro Gly Val Ala Gly

      2439      2448      2457      2466      2475      2484
GCT GTG GGT GAA OCT GGT CCT CTT GGC ATT GCG GGC CCT CCT GGG GCG CGT GGT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Ala Val Gly Glu Pro Gly Pro Leu Gly Ile Ala Gly Pro Pro Gly Ala Arg Gly

      2493      2502      2511      2520      2529      2538
CCT OCT GGT GCT GTG GGT AGT OCT GGA GTC AAC GGT GCT OCT GGT GAA GCT GGT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Pro Pro Gly Ala Val Gly Ser Pro Gly Val Asn Gly Ala Pro Gly Glu Ala Gly

      2547      2556      2565      2574      2583      2592
CGT GAT GGC AAC OCT GGG AAC GAT GGT COC CCA GGT CCG GAT GGT CAA COC GGA
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Arg Asp Gly Asn Pro Gly Asn Asp Gly Pro Pro Gly Arg Asp Gly Gln Pro Gly

      2601      2610      2619      2628      2637      2646
CAC AAG GGA GAG CGC GGT TAC OCT GGC AAT ATT GGT COC GTT GGT GCT GCA GGT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
His Lys Gly Glu Arg Gly Tyr Pro Gly Asn Ile Gly Pro Val Gly Ala Ala Gly

      2655      2664      2673      2682      2691      2700
GCA OCT GGT CCT CAT GGC CCC GTG GGT OCT GCT GGC AAA CAT GGA AAC CGT GGT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Ala Pro Gly Pro His Gly Pro Val Gly Pro Ala Gly Lys His Gly Asn Arg Gly

      2709      2718      2727      2736      2745      2754
GAA ACT GGT OCT TCT GGT CCT GTT GGT OCT GCT GGT GCT GTT GGC CCA AGA GGT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Glu Thr Gly Pro Ser Gly Pro Val Gly Pro Ala Gly Ala Val Gly Pro Arg Gly

      2763      2772      2781      2790      2799      2808

```

FIG. 49D

```

OCT AGT GGC CCA CAA GGC ATT CGT GGC GAT AAG GGA GAG CCC GGT GAA AAG GGG
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Pro Ser Gly Pro Gln Gly Ile Arg Gly Asp Lys Gly Glu Pro Gly Glu Lys Gly

      2817      2826      2835      2844      2853      2862
CCC AGA GGT CTT OCT GGC TTC AAG GGA CAC AAT GGA TTG CAA GGT CTG OCT GGT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Pro Arg Gly Leu Pro Gly Phe Lys Gly His Asn Gly Leu Gln Gly Leu Pro Gly

      2871      2880      2889      2898      2907      2916
ATC GCT GGT CAC CAT GGT GAT CAA GGT GCT OCT GGC TCC GTG GGT OCT GCT GGT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Ile Ala Gly His His Gly Asp Gln Gly Ala Pro Gly Ser Val Gly Pro Ala Gly

      2925      2934      2943      2952      2961      2970
OCT AGG GGC CCT GCT GGT OCT TCT GGC OCT GCT GGA AAA GAT GGT CGC ACT GGA
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Pro Arg Gly Pro Ala Gly Pro Ser Gly Pro Ala Gly Lys Asp Gly Arg Thr Gly

      2979      2988      2997      3006      3015      3024
CAT OCT GGT ACG GTT GGA OCT GCT GGC ATT GGA GGC OCT CAG GGT CAC CAA GGC
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
His Pro Gly Thr Val Gly Pro Ala Gly Ile Arg Gly Pro Gln Gly His Gln Gly

      3033      3042      3051      3060      3069      3078
OCT GCT GGC CCC OCT GGT CCC OCT GGC OCT CTT GGA OCT CTA GGT GTA AGC GGT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Pro Ala Gly Pro Pro Gly Pro Pro Gly Pro Leu Gly Pro Leu Gly Val Ser Gly

      3087      3096      3105      3114
GGT GGT TAT GAC TTT GGT TAC GAT GGA GAC TTC TAC AGG GCT 3'
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Gly Gly Tyr Asp Phe Gly Tyr Asp Gly Asp Phe Tyr Arg Ala

```

FIG. 49E

		9		18		27		36		45		54					
CAG	TAC	GAC	GGT	AAA	GGC	GTA	GGC	CTG	GGT	CCG	GGT	CCG	ATG	GGC	CTG	ATG	GGT
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
Gln	Tyr	Asp	Gly	Lys	Gly	Val	Gly	Leu	Gly	Pro	Gly	Pro	Met	Gly	Leu	Met	Gly
		63		72		81		90		99		108					
CCA	CGT	GGC	CCA	CCG	GGT	GCA	GCA	GGT	GCG	CCG	GGT	CCG	CAG	GGC	TTC	CAA	GGT
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
Pro	Arg	Gly	Pro	Pro	Gly	Ala	Ala	Gly	Ala	Pro	Gly	Pro	Gln	Gly	Phe	Gln	Gly
		117		126		135		144		153		162					
CCG	GCG	GGT	GAA	CCG	GGC	GAA	CCG	GGT	CAG	ACG	GGT	CCG	GCG	GGT	GCT	CGC	GGT
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
Pro	Ala	Gly	Glu	Pro	Gly	Glu	Pro	Gly	Gln	Thr	Gly	Pro	Ala	Gly	Ala	Arg	Gly
		171		180		189		198		207		216					
CCG	GCT	GGC	CCA	CCG	GGC	AAA	GCT	GGC	GAA	GAC	GGT	CAC	CCG	GGT	AAG	CCA	GGC
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
Pro	Ala	Gly	Pro	Pro	Gly	Lys	Ala	Gly	Glu	Asp	Gly	His	Pro	Gly	Lys	Pro	Gly
		225		234		243		252		261		270					
CGC	CCG	GGC	GAA	CGT	GGC	GTC	GTG	GGT	CCG	CAA	GGT	GCG	CGT	GGT	TTC	CCG	GGC
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
Arg	Pro	Gly	Glu	Arg	Gly	Val	Val	Gly	Pro	Gln	Gly	Ala	Arg	Gly	Phe	Pro	Gly
		279		288		297		306		315		324					
ACG	CCG	GGT	CTG	CCG	GGT	TTC	AAA	GGC	ATT	CGT	GGT	CAC	AAC	GGT	CTG	GAC	GGT
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
Thr	Pro	Gly	Leu	Pro	Gly	Phe	Lys	Gly	Ile	Arg	Gly	His	Asn	Gly	Leu	Asp	Gly
		333		342		351		360		369		378					
CTG	AAA	GGC	CAA	CCG	GGT	GCT	CCG	GGC	GTC	AAA	GGC	GAA	CCG	GGT	GCC	CCA	GGC
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
Leu	Lys	Gly	Gln	Pro	Gly	Ala	Pro	Gly	Val	Lys	Gly	Glu	Pro	Gly	Ala	Pro	Gly
		387		396		405		414		423		432					
GAA	AAC	GGT	ACG	CCG	GGC	CAG	ACT	GGT	GCG	CGT	GGT	CTG	CCG	GGT	GAA	CGC	GGC
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
Glu	Asn	Gly	Thr	Pro	Gly	Gln	Thr	Gly	Ala	Arg	Gly	Leu	Pro	Gly	Glu	Arg	Gly
		441		450		459		468		477		486					
CGT	GTT	GGC	GCT	CCG	GGT	CCG	GCT	GGC	GCG	CGT	GGC	AGC	GAT	GGC	TCC	GTC	GGT
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
Arg	Val	Gly	Ala	Pro	Gly	Pro	Ala	Gly	Ala	Arg	Gly	Ser	Asp	Gly	Ser	Val	Gly
		495		504		513		522		531		540					
CCG	GTT	GGC	CCT	GCG	GGT	CCG	ATT	GGT	TCC	GCT	GGC	OCT	CCG	GGT	TTC	CCG	GGT
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
Pro	Val	Gly	Pro	Ala	Gly	Pro	Ile	Gly									

FIG. 50A

657	666	675	684	693	702
CCA CCG GGT AAC CCG GGC GCA AAC GGC CTG ACC GGT GCA AAA GGT GCG GCT GGC					
Pro Pro Gly Asn Pro Gly Ala Asn Gly Leu Thr Gly Ala Lys Gly Ala Ala Gly					
711	720	729	738	747	756
CTG CCG GGC GTT GCC GGT GCC CCG GGC CTG CCG GGT CCG CCG GGT ATT CCG GGT					
Leu Pro Gly Val Ala Gly Ala Pro Gly Leu Pro Gly Pro Arg Gly Ile Pro Gly					
765	774	783	792	801	810
CCG GTA GGC GCA GCC GGT GCA ACT GGT GCC CGT GGC CTG GTT GGC GAA CCG GGT					
Pro Val Gly Ala Ala Gly Ala Thr Gly Ala Arg Gly Leu Val Gly Glu Pro Gly					
819	828	837	846	855	864
CCG GCG GGT TCT AAA GGC GAA AGC GGT AAC AAA GGT GAG CCG GGT TCC CCG GGC					
Pro Ala Gly Ser Lys Gly Glu Ser Gly Asn Lys Gly Glu Pro Gly Ser Ala Gly					
873	882	891	900	909	918
CCG CAG GGT CCG CCG GGT CCG AGC GGC GAA GAA GGT AAA CGT GGT CCG AAC GGC					
Pro Gln Gly Pro Pro Gly Pro Ser Gly Glu Glu Gly Lys Arg Gly Pro Asn Gly					
927	936	945	954	963	972
GAG GCT GGT TCC GCA GGC CCT CCG GGT CCG CCG GGT CTG CGT GGC AGC CCG GGT					
Glu Ala Gly Ser Ala Gly Pro Pro Gly Pro Pro Gly Leu Arg Gly Ser Pro Gly					
981	990	999	1008	1017	1026
AGC CGT GGC CTG CCG GGC GCG GAC GGC CGT GCG GGC GTG ATG GGT CCG CCG GGT					
Ser Arg Gly Leu Pro Gly Ala Asp Gly Arg Ala Gly Val Met Gly Pro Pro Gly					
1035	1044	1053	1062	1071	1080
TCC CGT GGT GCC TCT GGT CCG GCT GGT GTC CGT GGT CCG AAT GGC GAC GCG GGC					
Ser Arg Gly Ala Ser Gly Pro Ala Gly Val Arg Gly Pro Asn Gly Asp Ala Gly					
1089	1098	1107	1116	1125	1134
CGT CCG GGT GAA CCG GGC CTG ATG GGT CCG CGT GGC CTG CCG GGT AGC CCG GGT					
Arg Pro Gly Glu Pro Gly Leu Met Gly Pro Arg Gly Leu Pro Gly Ser Pro Gly					
1143	1152	1161	1170	1179	1188
AAC ATT GGT CCG CCG GGT AAG GAG GGT CCG GTA GGT CTG CCG GGT ATT GAT GGT					
Asn Ile Gly Pro Ala Gly Lys Glu Gly Pro Val Gly Leu Pro Gly Ile Asp Gly					
1197	1206	1215	1224	1233	1242
CGT CCG GGT CCG ATC GGC CCT GCG GGC GCT CGT GGC GAG CCG GGT AAC ATC GGT					
Arg Pro Gly Pro Ile Gly Pro Ala Gly Ala Arg Gly Glu Pro Gly Asn Ile Gly					
1251	1260	1269	1278	1287	1296
TTT CCG GGT CCG AAG GGT CCG ACG GGC GAC CCG GGC AAG AAC GGT GAT AAA GGC					
Phe Pro Gly Pro Lys Gly Pro Thr Gly Asp Pro Gly Lys Asn Gly Asp Lys Gly					
1305	1314	1323	1332	1341	1350
CAT GCA GGT CTG GCA GGT GCC CGT GGT GCA CCG GGT CCG GAT GGT AAC AAT GGT					
His Ala Gly Leu Ala Gly Ala Arg Gly Ala Pro Gly Pro Asp Gly Asn Asn Gly					

FIG. 50B

1359	1368	1377	1386	1395	1404
CCG CAG GGT CCG CCG GGT CCG CAG GGC GTA CAG GGT GGC AAA GGT GAA CAG GGT					
Ala Gln Gly Pro Pro Gly Pro Gln Gly Val Gln Gly Gly Lys Gly Glu Gln Gly					
1413	1422	1431	1440	1449	1458
CCG GCA GGC CCA CCG GGC TTC CAG GGT CTG CCG GGT CCG AGC GGC CCG GCT GGT					
Pro Ala Gly Pro Pro Gly Phe Gln Gly Leu Pro Gly Pro Ser Gly Pro Ala Gly					
1467	1476	1485	1494	1503	1512
GAA GTG GGC AAA CCG GGC GAA CGT GGC CTC CAT GGC GAG TTT GGC CTG CCG GGT					
Glu Val Gly Lys Pro Gly Glu Arg Gly Leu His Gly Glu Phe Gly Leu Pro Gly					
1521	1530	1539	1548	1557	1566
CCG GCC GGT CCG CGT GGT GAG CGC GGC CCT CCG GGC GAA TCC GGC GCG GCA GGT					
Pro Ala Gly Pro Arg Gly Glu Arg Gly Pro Pro Gly Glu Ser Gly Ala Ala Gly					
1575	1584	1593	1602	1611	1620
CCG ACC GGC CCG ATT GGT TCC CGT GGT CCG AGC GGC CCA CCG GGT CCG GAC GGC					
Pro Thr Gly Pro Ile Gly Ser Arg Gly Pro Ser Gly Pro Pro Gly Pro Asp Gly					
1629	1638	1647	1656	1665	1674
AAC AAA GGC GAG CCG GGT GTT GTT GGT GCT GTT GGT ACC GCC GGC CCG TCT GGT					
Asn Lys Gly Glu Pro Gly Val Val Gly Ala Val Gly Thr Ala Gly Pro Ser Gly					
1683	1692	1701	1710	1719	1728
CCG AGC GGT CTG CCG GGC GAA CGC GGT GCC GCT GGT ATT CCG GGC GGC AAA GGT					
Pro Ser Gly Leu Pro Gly Glu Arg Gly Ala Ala Gly Ile Pro Gly Gly Lys Gly					
1737	1746	1755	1764	1773	1782
GAA AAA GGT GAA CCG GGT CTG CGC GGT GAG ATT GGC AAC CCG GGC CGT GAC GGT					
Glu Lys Gly Glu Pro Gly Leu Arg Gly Glu Ile Gly Asn Pro Gly Arg Asp Gly					
1791	1800	1809	1818	1827	1836
GCT CGC GGT GCA CAC GGC GCG GTT GGC GCA CCG GGT CCG GCA GGC GCG ACT GGT					
Ala Arg Gly Ala His Gly Ala Val Gly Ala Pro Gly Pro Ala Gly Ala Thr Gly					
1845	1854	1863	1872	1881	1890
GAT CGT GGC GAA GCT GGT GCA CGC GGT CCG CGC GGT CCG GCC GGC CCT CGC GGT					
Asp Arg Gly Glu Ala Gly Ala Ala Gly Pro Ala Gly Pro Ala Gly Pro Arg Gly					
1899	1908	1917	1926	1935	1944
TCC CCG GGC GAA CGC GGC GAA GTC GGC CCG GCT GGC CCG AAT GGC TTT GCT GGC					
Ser Pro Gly Glu Arg Gly Glu Val Gly Pro Ala Gly Pro Asn Gly Phe Ala Gly					
1953	1962	1971	1980	1989	1998
CCA CCG GGC GCT GCG GGC CAA CCG GGT CCG AAA GGT GAG CGC GGT GCC AAA GGC					
Pro Ala Gly Ala Ala Gly Gln Pro Gly Ala Lys Gly Glu Arg Gly Ala Lys Gly					
2007	2016	2025	2034	2043	2052
CCG AAA GGT GAA AAT GGT GTA GTT GGT CCG ACG GGT CCG GTT GGT CCG GCT GGT					
Pro Lys Gly Glu Asn Gly Val Val Gly Pro Thr Gly Pro Val Gly Ala Ala Gly					

FIG. 50C

2061	2070	2079	2088	2097	2106
CCG GCT GGC CCG AAT GGT CCG CCG GGT CCG GCA GGC AGC CGT GGC GAT GGT GGC					
Pro Ala Gly	Pro Asn Gly	Pro Pro Gly	Pro Ala Gly	Ser Arg Gly	Asp Gly Gly
2115	2124	2133	2142	2151	2160
CCA CCG GGC ATG ACC GGT TTC CCT GGC GCG GCC GGT CCG ACC GGC CCG CCG GGT					
Pro Pro Gly	Met Thr Gly	Phe Pro Gly	Ala Ala Gly	Arg Thr Gly	Pro Pro Gly
2169	2178	2187	2196	2205	2214
CCG TCT GGC ATT TCT GGC CCA CCG GGT CCG CCG GGT CCG GCG GGC AAA GAA GGT					
Pro Ser Gly	Ile Ser Gly	Pro Pro Gly	Pro Pro Gly	Pro Ala Gly	Lys Glu Gly
2223	2232	2241	2250	2259	2268
CTG CGT GGC CCA CGC GGC GAC CAG GGT CCG GTG GGC CGT ACC GGC GAA GTC GGT					
Leu Arg Gly	Pro Arg Gly	Asp Gln Gly	Pro Val Gly	Arg Thr Gly	Glu Val Gly
2277	2286	2295	2304	2313	2322
GCT GTT GGC CCT CCG GGC TTT GCG GGT GAG AAA GGT CCG AGC GGT GAA GCT GGC					
Ala Val Gly	Pro Pro Gly	Phe Ala Gly	Glu Lys Gly	Pro Ser Gly	Glu Ala Gly
2331	2340	2349	2358	2367	2376
ACC GCA GGC CCG CCG GGT ACG CCG GGT CCG CAA GGT CTG CTG GGT GCT CCG GGT					
Thr Ala Gly	Pro Pro Gly	Thr Pro Gly	Pro Gln Gly	Leu Leu Gly	Ala Pro Gly
2385	2394	2403	2412	2421	2430
ATC CTG GGC CTG CCG GGC TCC CGT GGC GAA CGC GGT CTG CCG GGC GTT GCA GGC					
Ile Leu Gly	Leu Pro Gly	Ser Arg Gly	Glu Arg Gly	Leu Pro Gly	Val Ala Gly
2439	2448	2457	2466	2475	2484
GCT GTA GGC GAA CCG GGT CCG CTG GGT ATC GCG GGT CCG CCG GGT GCG CGT GGT					
Ala Val Gly	Glu Pro Gly	Pro Leu Gly	Ile Ala Gly	Pro Pro Gly	Ala Arg Gly
2493	2502	2511	2520	2529	2538
CCG CCG GGT GCC GTG GGC TCT CCG GGT GTT AAC GGC GCC CCT GGT GAA GCG GGC					
Pro Pro Gly	Ala Val Gly	Ser Pro Gly	Val Asn Gly	Ala Pro Gly	Glu Ala Gly
2547	2556	2565	2574	2583	2592
CGC GAC GGC AAT CCG GGC AAC GAT GGT CCG CCG GGT CGT GAT GGT CAG CCG GGT					
Arg Asp Gly	Asn Pro Gly	Asn Asp Gly	Pro Pro Gly	Arg Asp Gly	Gln Pro Gly
2601	2610	2619	2628	2637	2646
CAC AAA GGT GAG CGT GGC TAC CCG GGT AAC ATC GGT CCG GTT GGT GCG GCC GGC					
His Lys Gly	Glu Arg Gly	Tyr Pro Gly	Asn Ile Gly	Pro Val Gly	Ala Ala Gly
2655	2664	2673	2682	2691	2700
GCT CCG GGT CCG CAC GGT CCG GTA GGC CCA GCC GGC AAA CAC GGT AAC CGT GGT					
Ala Pro Gly	Pro His Gly	Pro Val Gly	Pro Ala Gly	Lys His Gly	Asn Arg Gly
2709	2718	2727	2736	2745	2754
GAA ACG GGT CCG TCC GGT CCG GTA GGT CCG GCG GGT GCT GTT GGT CCA CCG GGC					
Glu Thr Gly	Pro Ser Gly	Pro Val Gly	Pro Ala Gly	Ala Val Gly	Pro Arg Gly

FIG. 50D

2763	2772	2781	2790	2799	2808
CCG TCC GGC CCG CAG GGT ATT CGC GGT GAC AAA GGC GAA CCG GGC GAA AAA GGT					
Pro Ser Gly Pro Gln Gly Ile Arg Gly Asp Lys Gly Glu Pro Gly Glu Lys Gly					
2817	2826	2835	2844	2853	2862
CCG CGT GGT CTG CCG GGC CTT AAG GGC CAC AAC GGT CTG CAA GGT CTG CCG GGT					
Pro Arg Gly Leu Pro Gly Leu Lys Gly His Asn Gly Leu Gln Gly Leu Pro Gly					
2871	2880	2889	2898	2907	2916
ATC GCG GGT CAC CAC GGT GAT CAG GGT GCT CCG GGT TCC GTT GGT CCG GCC GGT					
Ile Ala Gly His His Gly Asp Gln Gly Ala Pro Gly Ser Val Gly Pro Ala Gly					
2925	2934	2943	2952	2961	2970
CCG CGT GGC CCG GCT GGT CCG TCT GGT CCG GCC GGT AAA GAC GGC CGT ACG GGC					
Pro Arg Gly Pro Ala Gly Pro Ser Gly Pro Ala Gly Lys Asp Gly Arg Thr Gly					
2979	2988	2997	3006	3015	3024
CAC CCG GGT ACG GTG GGT CCG GCC GGC ATT CGC GGT CCG CAA GGT CAC CAG GGT					
His Pro Gly Thr Val Gly Pro Ala Gly Ile Arg Gly Pro Gln Gly His Gln Gly					
3033	3042	3051	3060	3069	3078
CCG CCG GGT CCG CCG GGT CCG CCG GGT CCG CCG GGT CCG CCG GGT GTT AGC GGT					
Pro Ala Gly Pro Pro Gly Pro Pro Gly Pro Pro Gly Pro Pro Gly Val Ser Gly					
3087	3096	3105	3114		
GGC GGT TAT GAT TTT GGT TAT GAC GGT GAT TTC TAT CGT GCG 3'					
Gly Gly Tyr Asp Phe Gly Tyr Asp Gly Asp Phe Tyr Arg Ala					

FIG. 50E

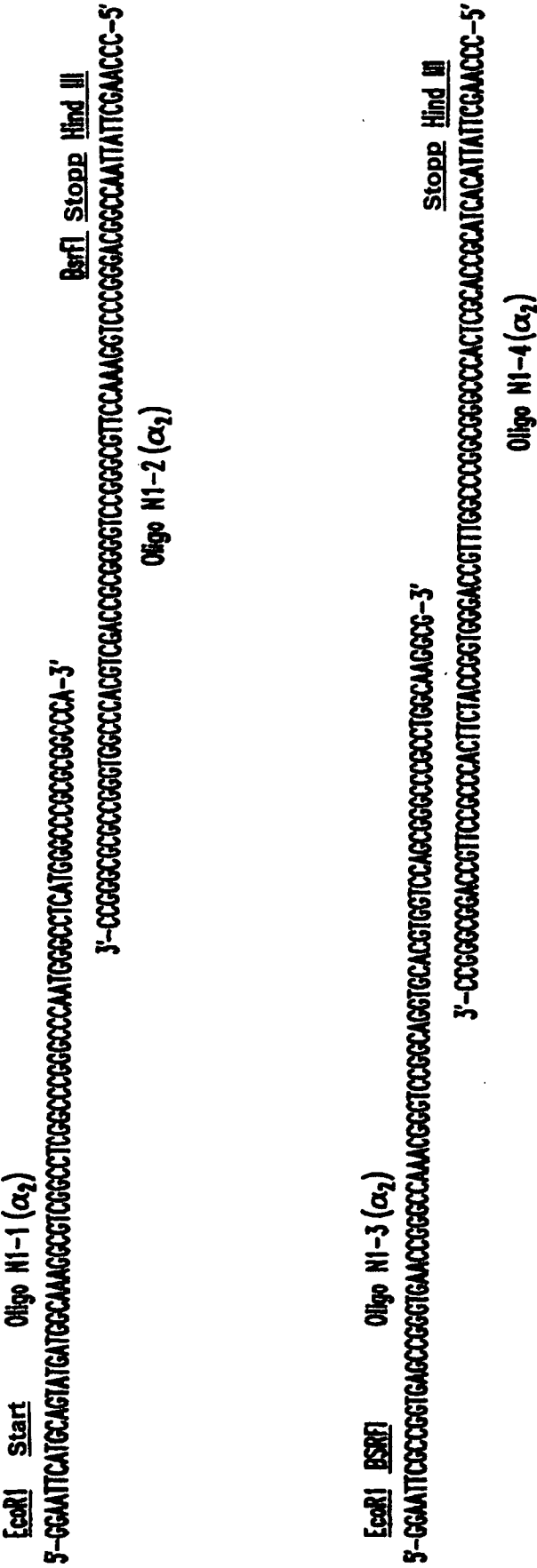


FIG. 5I

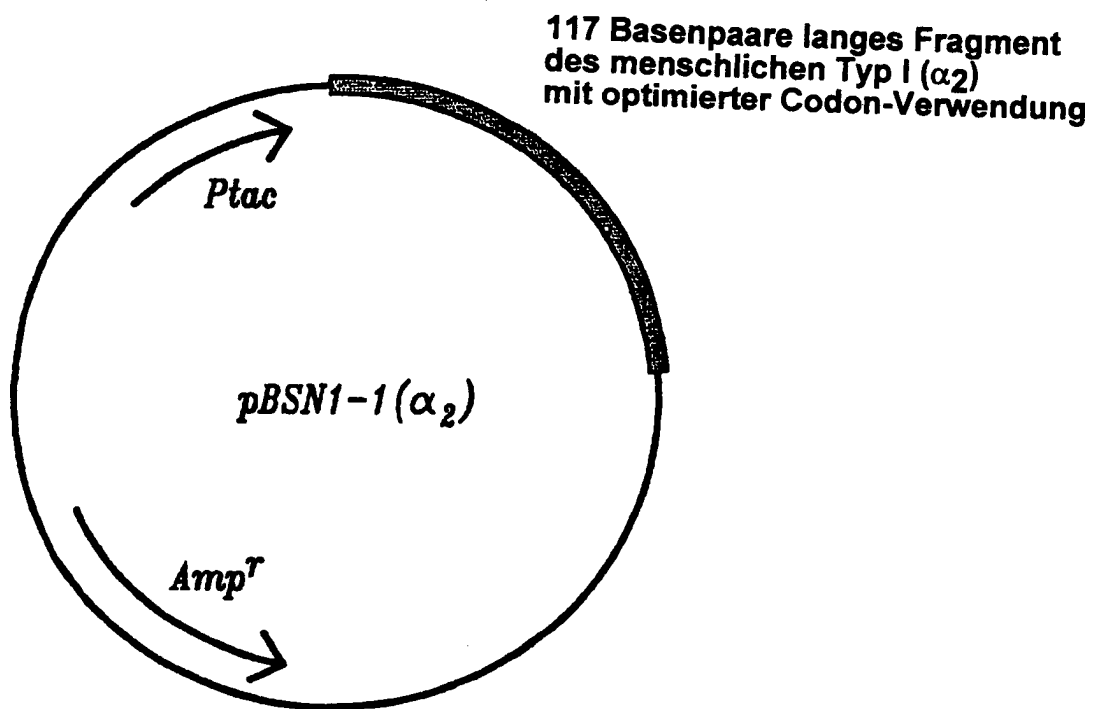


FIG. 52

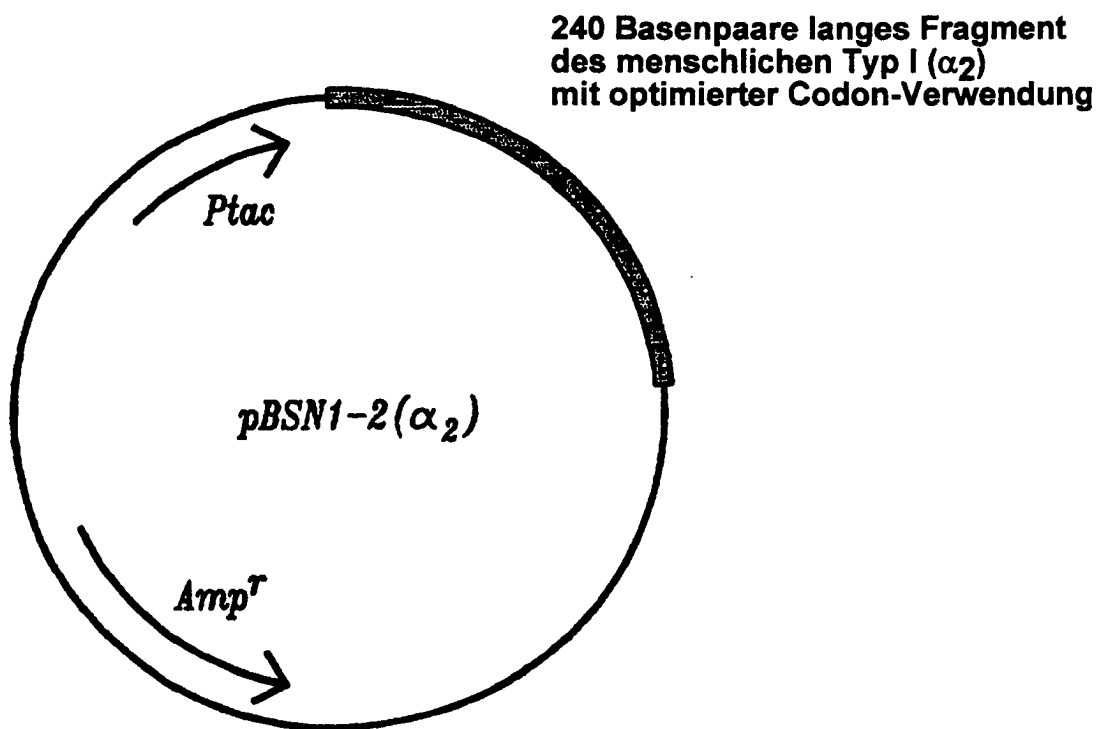


FIG. 53

```

      9      18      27      36      45      54
5'  CAG TAT GAT GGC AAA GGC GTC GGC CTC GGC CCG GGC OCA ATG GGC CTC ATG GGC
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
    Gln Tyr Asp Gly Lys Gly Val Gly Leu Gly Pro Gly Pro Met Gly Leu Met Gly

      63      72      81      90      99      108
    CCG CGC GGC OCA CCG GGT GCA GCT GGC GGC OCA GGC CCG CAA GGT TTC CAG GGC
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
    Pro Arg Gly Pro Pro Gly Ala Ala Gly Ala Pro Gly Pro Gln Gly Phe Gln Gly

      117      126      135      144      153      162
    OCT GGC GGT GAG CCG GGT GAA CCG GGC CAA ACG GGT CCG GCA GGT GCA CGT GGT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
    Pro Ala Gly Glu Pro Gly Glu Pro Gly Gln Thr Gly Pro Ala Gly Ala Arg Gly

      171      180      189      198      207      216
    CCA GCG GGC CCG OCT GGC AAG GCG GGT GAA GAT GGC CAC OCT GGC AAA CCG GGC
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
    Pro Ala Gly Pro Pro Gly Lys Ala Gly Glu Asp Gly His Pro Gly Lys Pro Gly

      225      234
    OGC CCG GGT GAG CGT GGC GTA GTG
    --- --- --- --- --- --- ---
    Arg Pro Gly Glu Arg Gly Val Val

```

FIG. 54

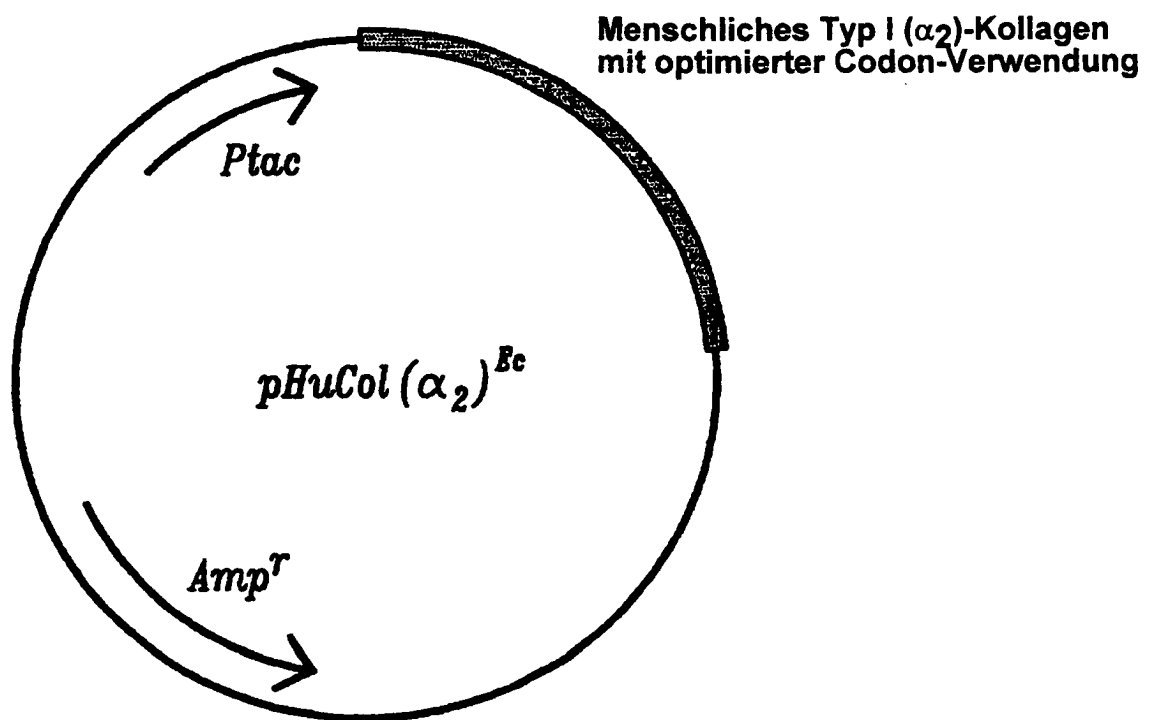


FIG. 55

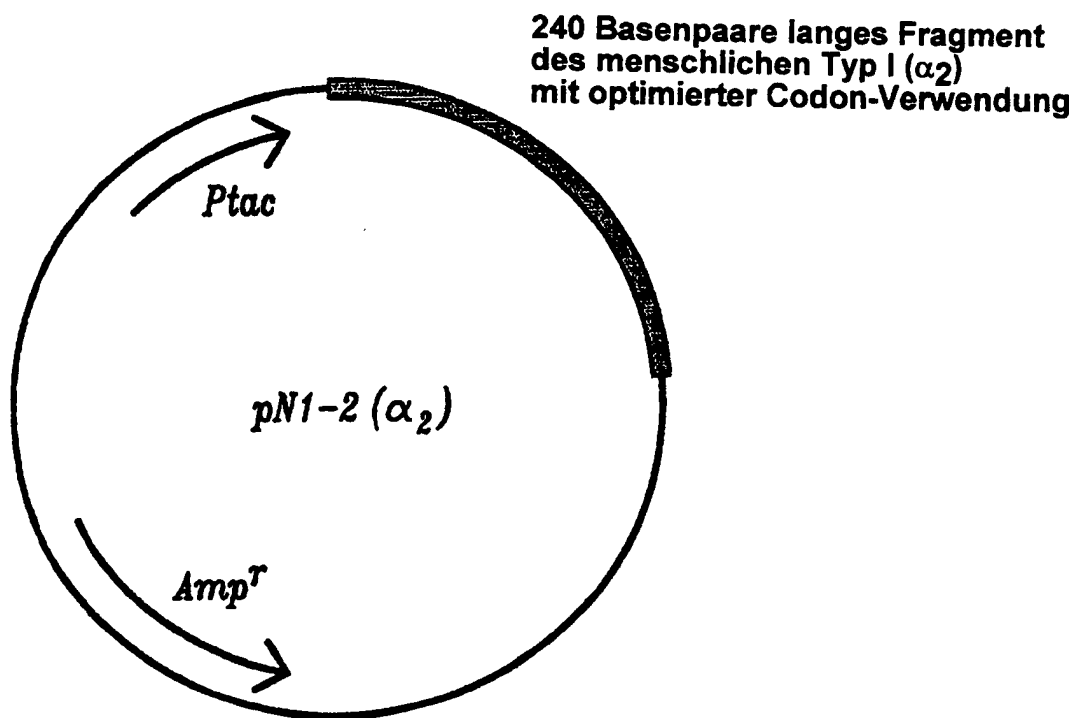


FIG. 56

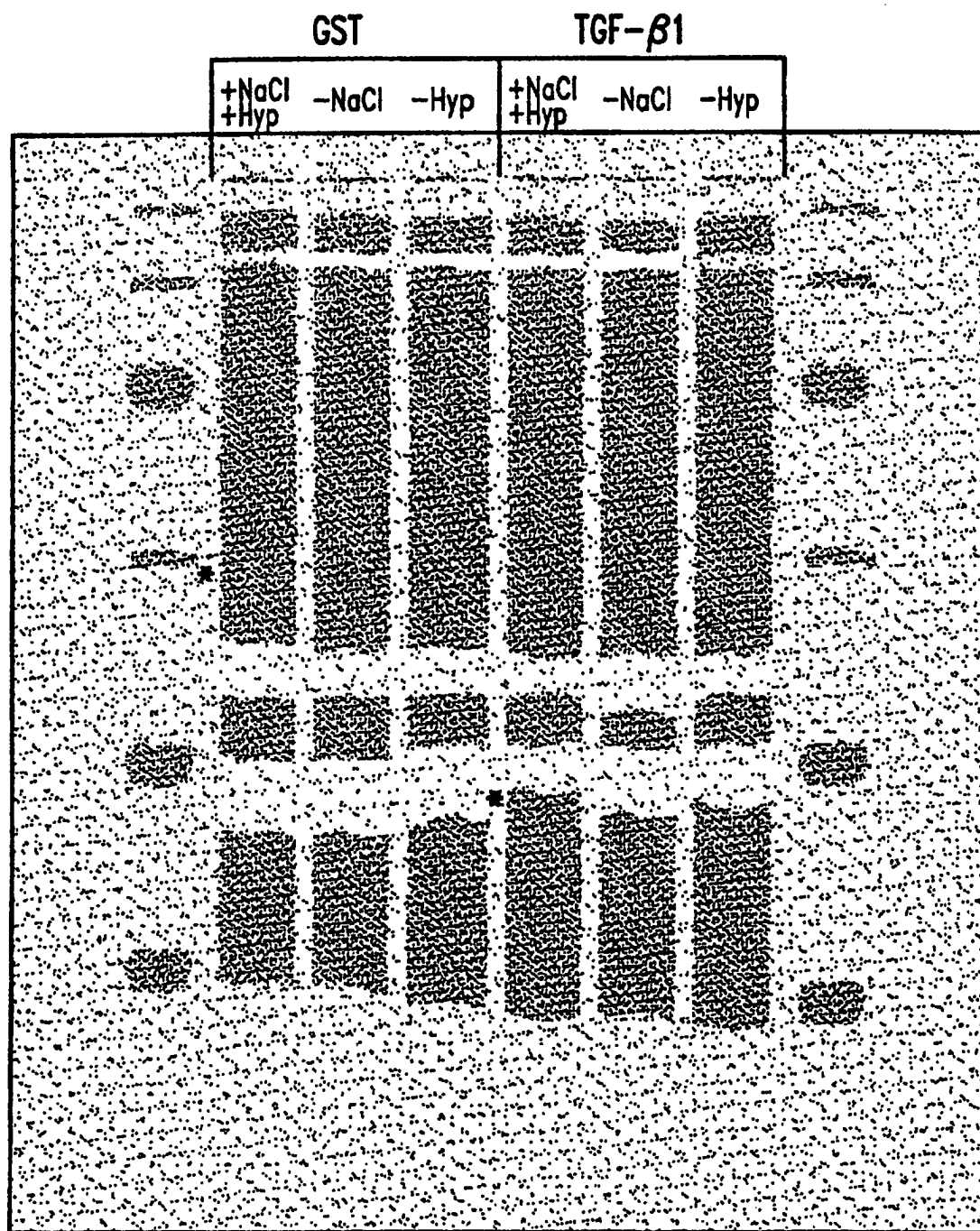


FIG. 57

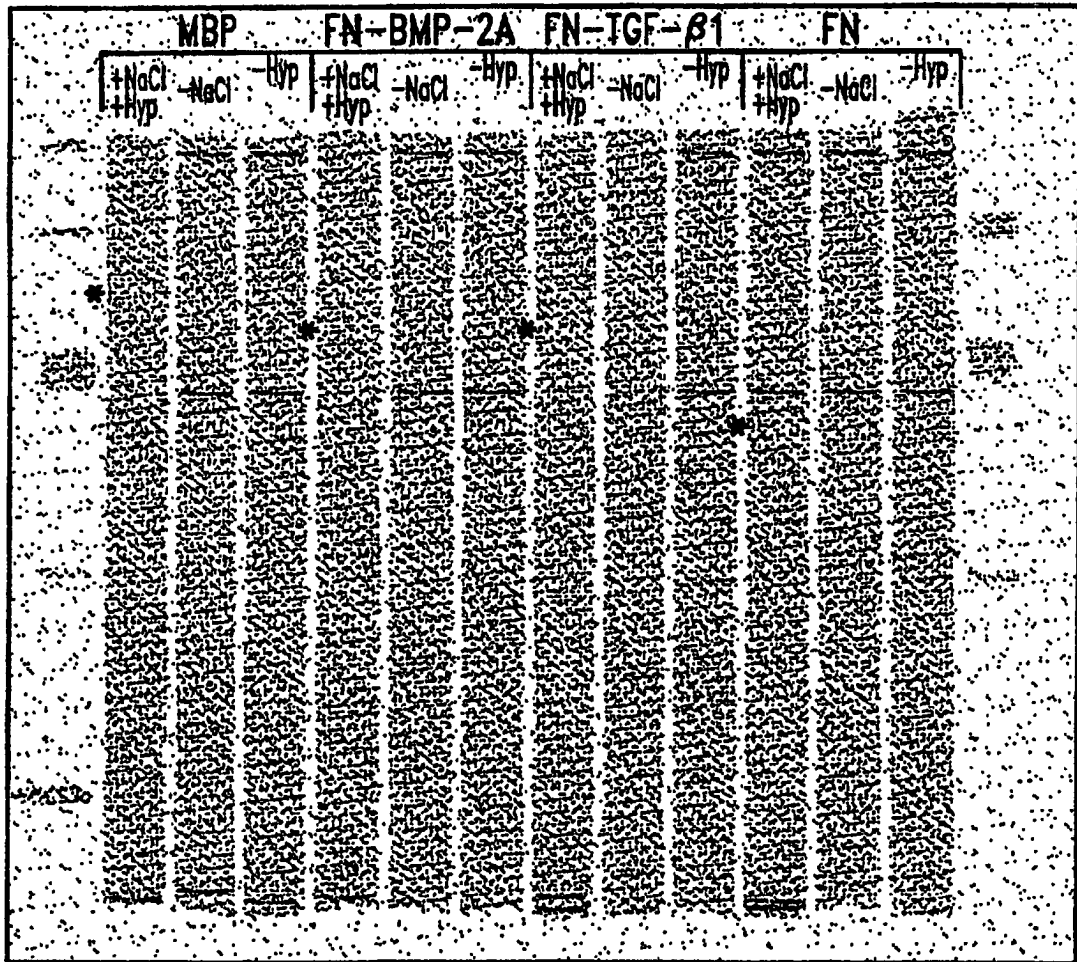


FIG. 58

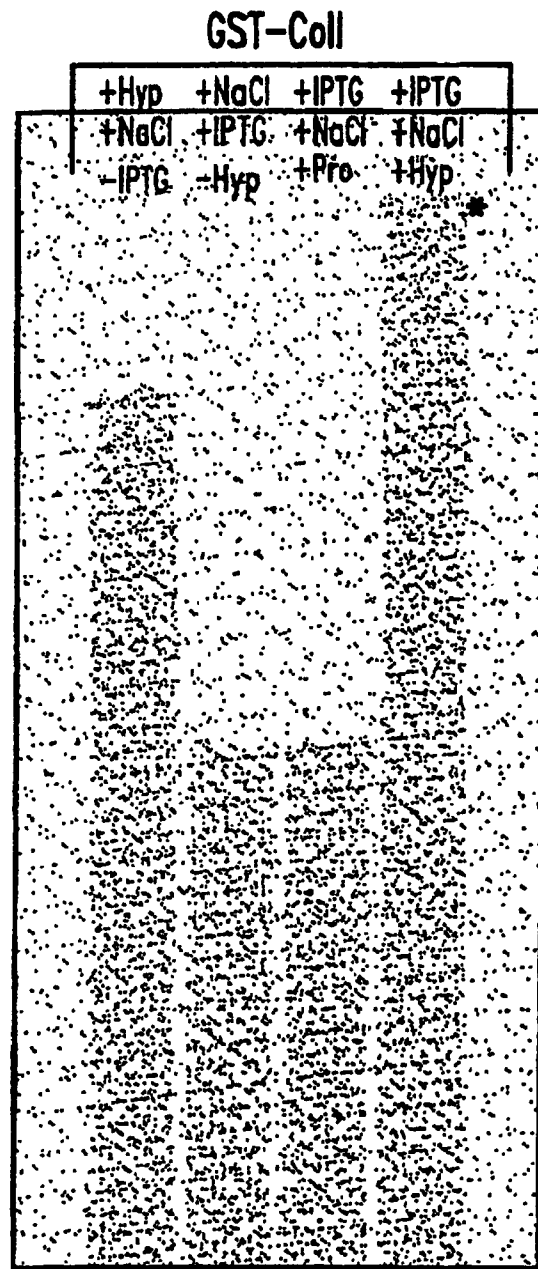


FIG. 59

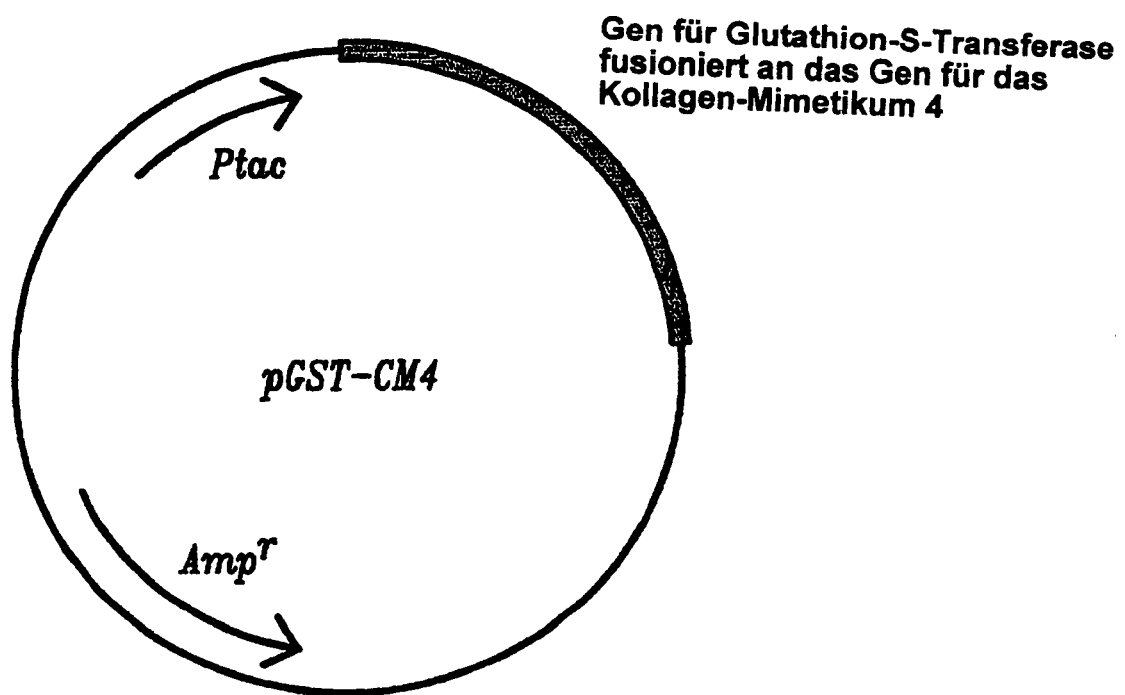


FIG. 60

5'	ATG	GGG	CTC	GCT	GGC	CCA	CCG	GGC	GAA	CCG	GGT	CCG	CCA	GGC	CCG	AAA	GGT	CCG	54
	M	G	L	A	G	P	P	G	E	P	G	P	P	G	P	K	G	P	
	CGT	GGC	GAT	AGC	GGG	CTC	GCT	GGC	CCA	CCG	GGC	GAA	CCG	GGT	CCG	CCA	GGC	CCG	108
	R	G	D	S	G	L	A	G	P	P	G	E	P	G	P	P	G	P	
	AAA	GGT	CCG	CGT	GGC	GAT	AGC	GGG	CTC	GCT	GGC	CCA	CCG	GGC	GAA	CCG	GGT	CCG	162
	K	G	P	R	G	D	S	G	L	A	G	P	P	G	E	P	G	P	
	CCA	GGC	CCG	AAA	GGT	CCG	CGT	GGC	GAT	AGC	GGG	CTC	GCT	GGC	CCA	CCG	GGC	GAA	216
	P	G	P	K	G	P	R	G	D	S	G	L	A	G	P	P	G	E	
	CCG	GGT	CCG	CCA	GGC	CCG	AAA	GGT	CCG	CGT	GGC	GAT	AGC	GGG	CTC	CCG	GGC	GAT	270
	P	G	P	P	G	P	K	G	P	R	G	D	S	G	L	P	G	D	
	TCC	TAA	3'																
	S	*																	

FIG. 61

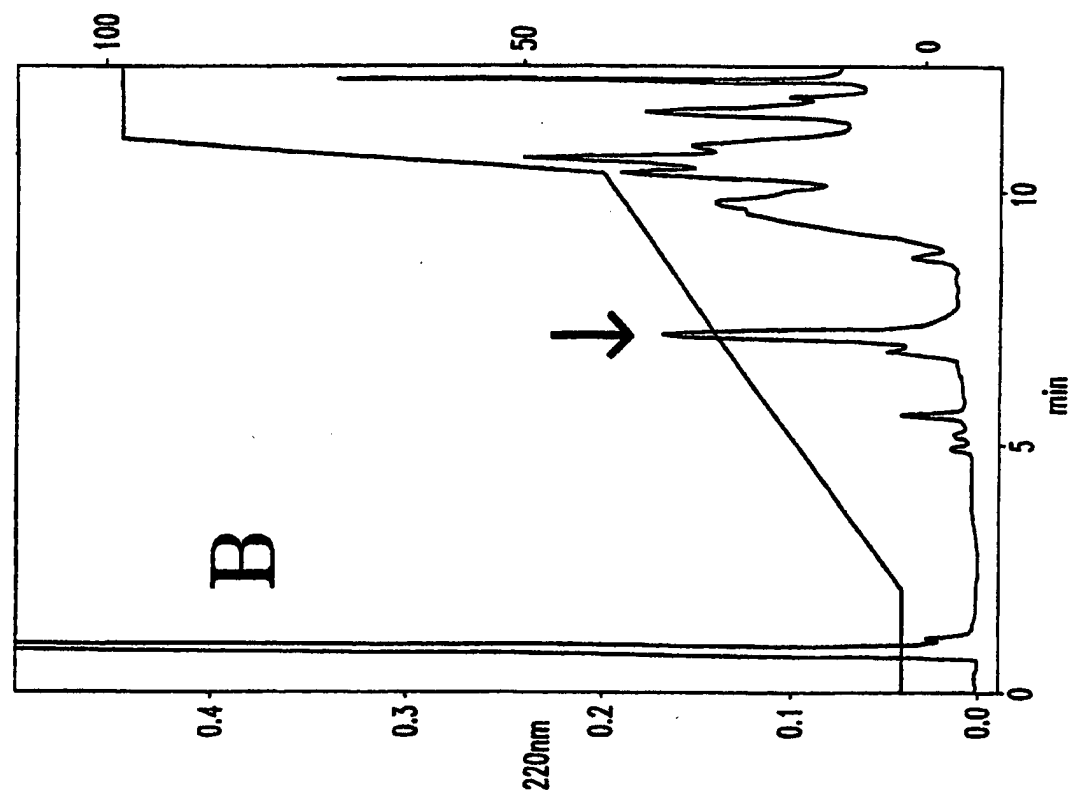


FIG. 62B

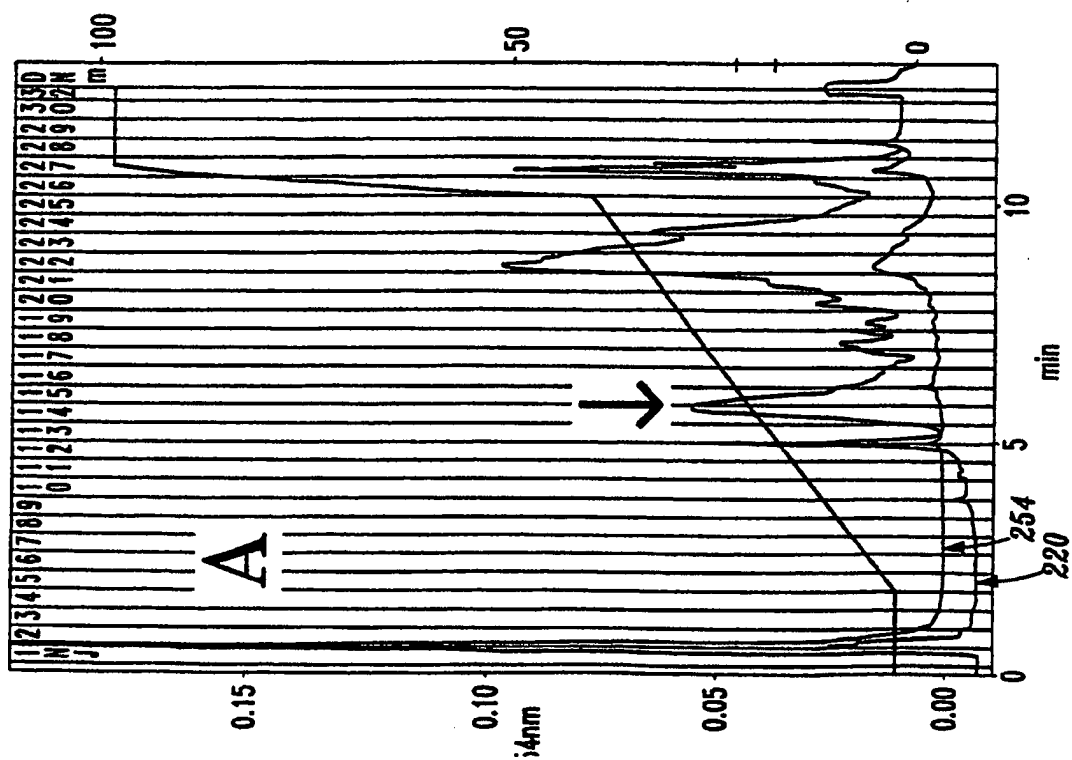


FIG. 62A

FIG. 63A

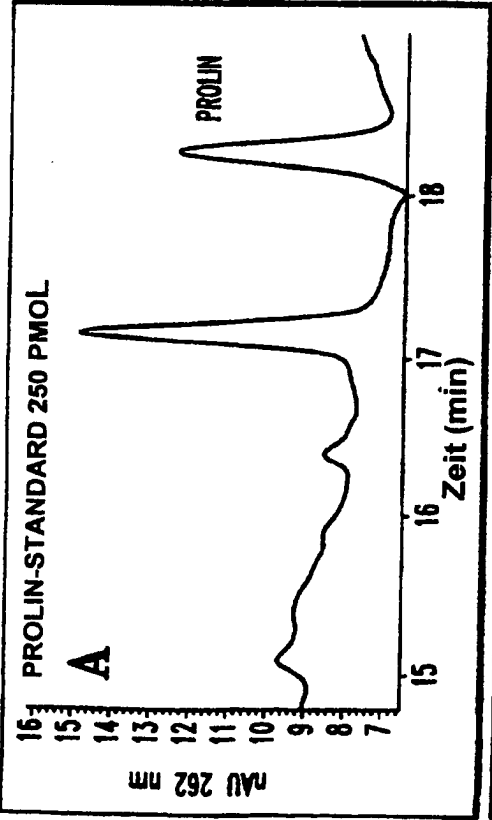


FIG. 63C

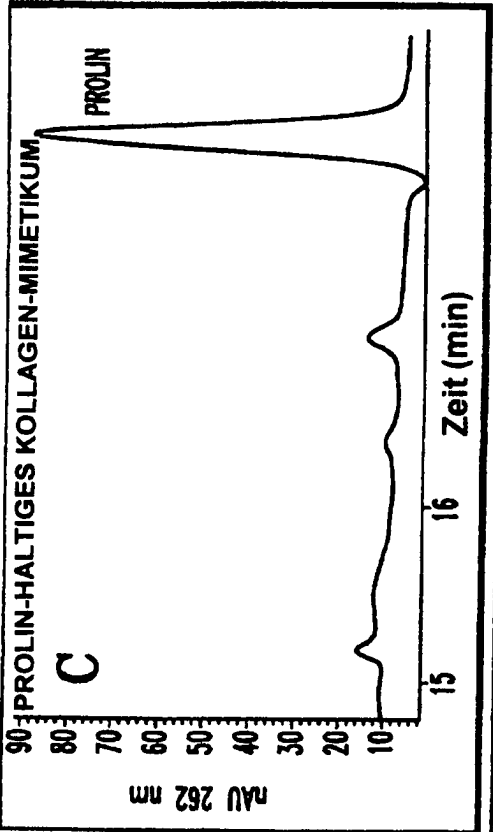


FIG. 63B

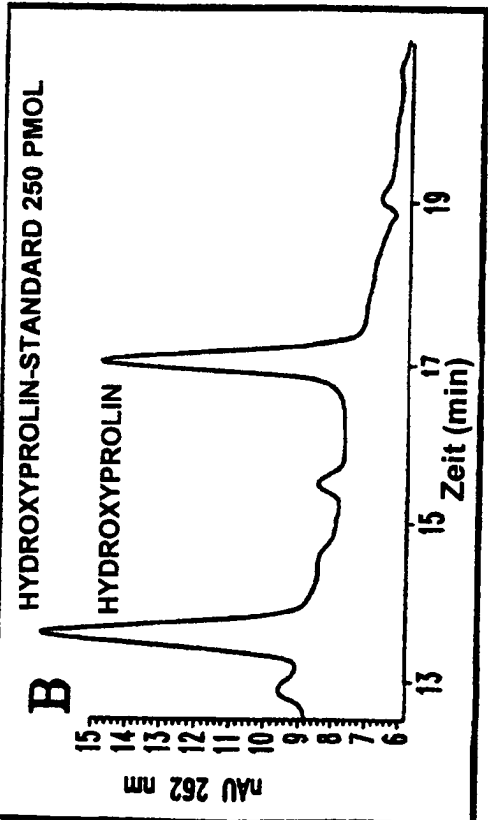
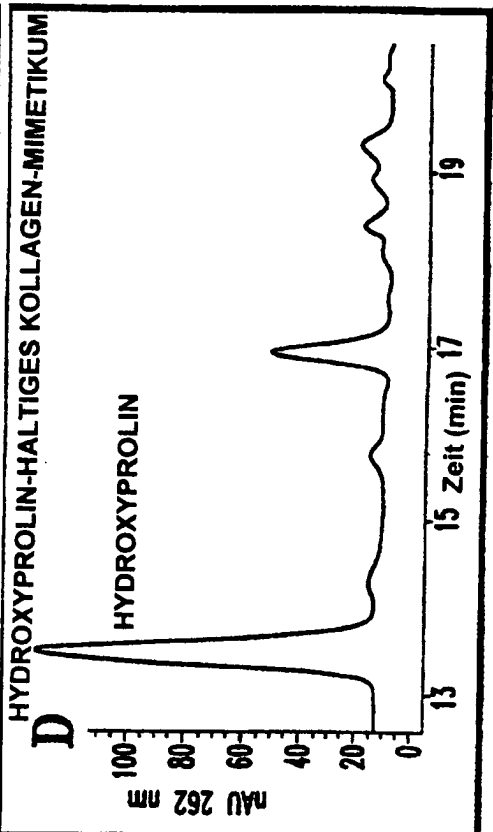


FIG. 63D



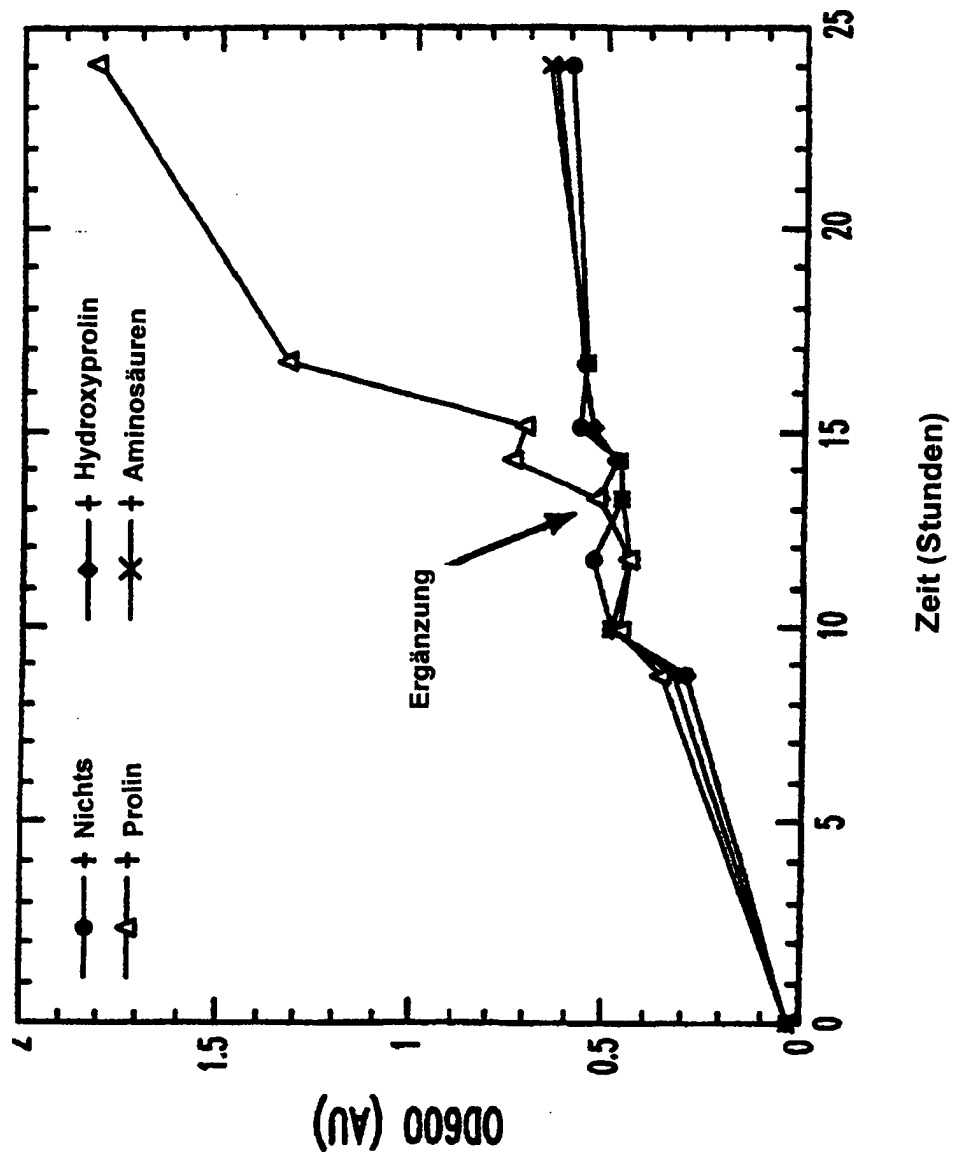


FIG. 64

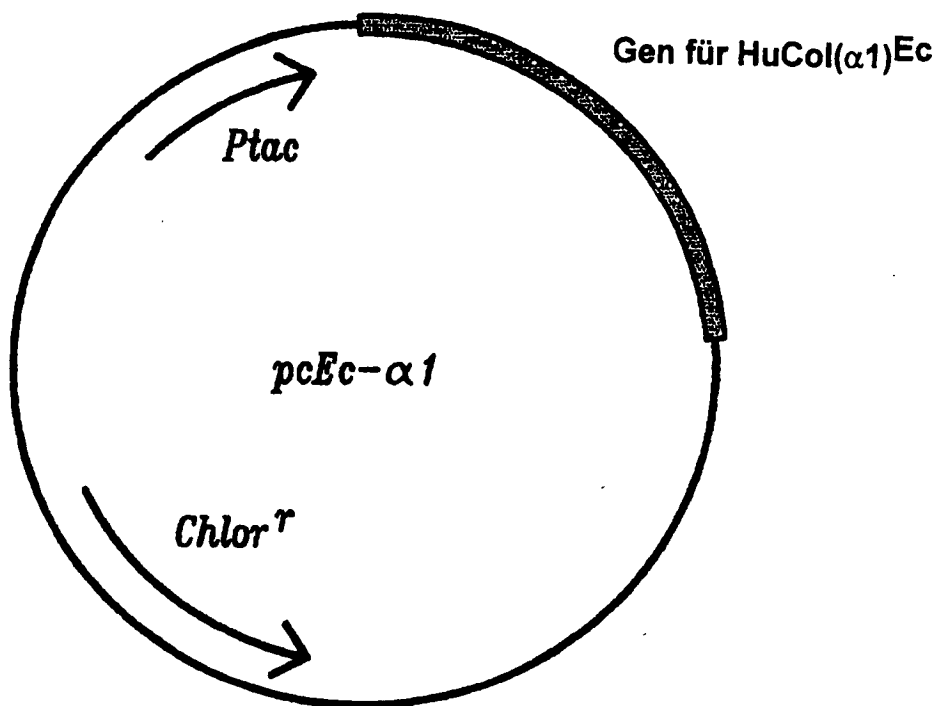


FIG. 65

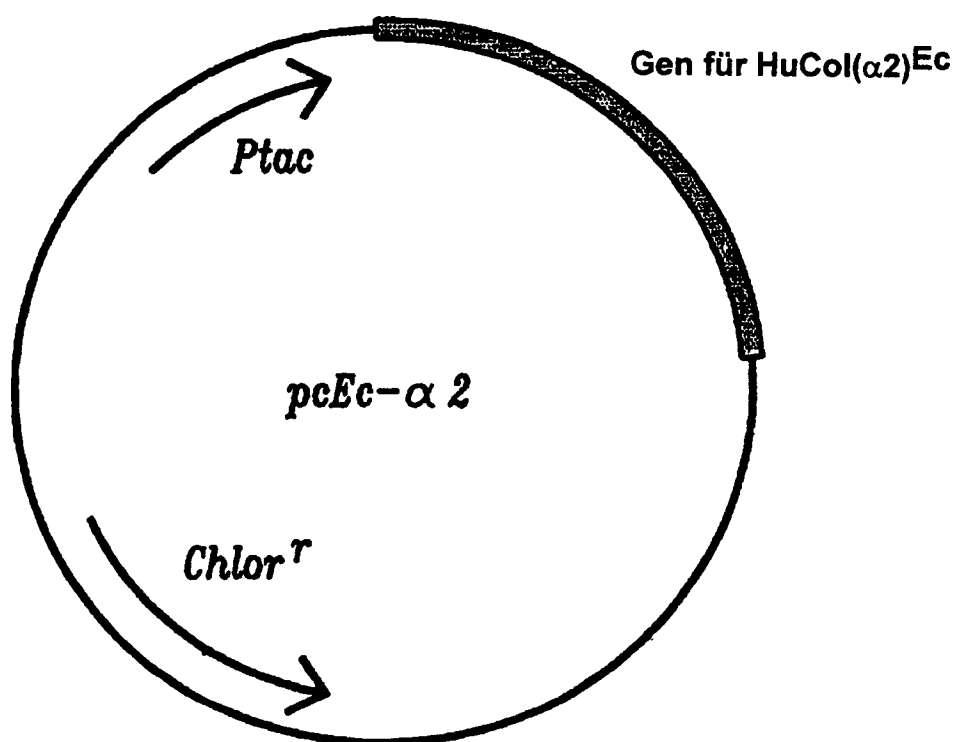


FIG. 66

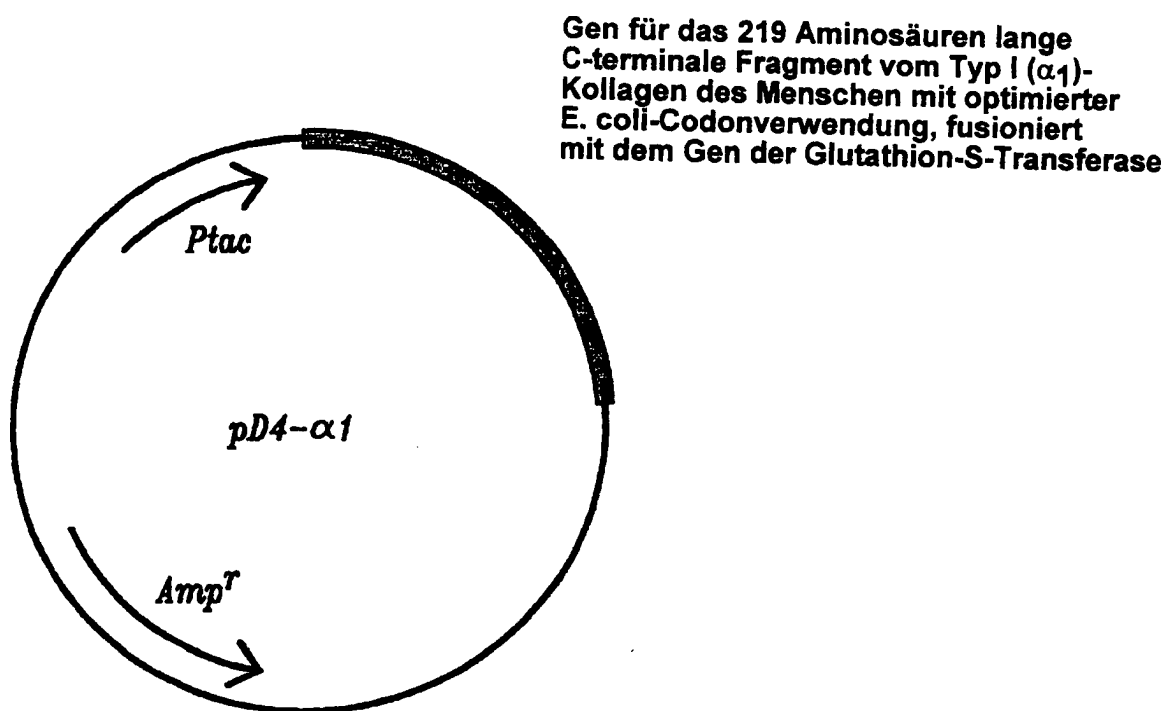


FIG. 67

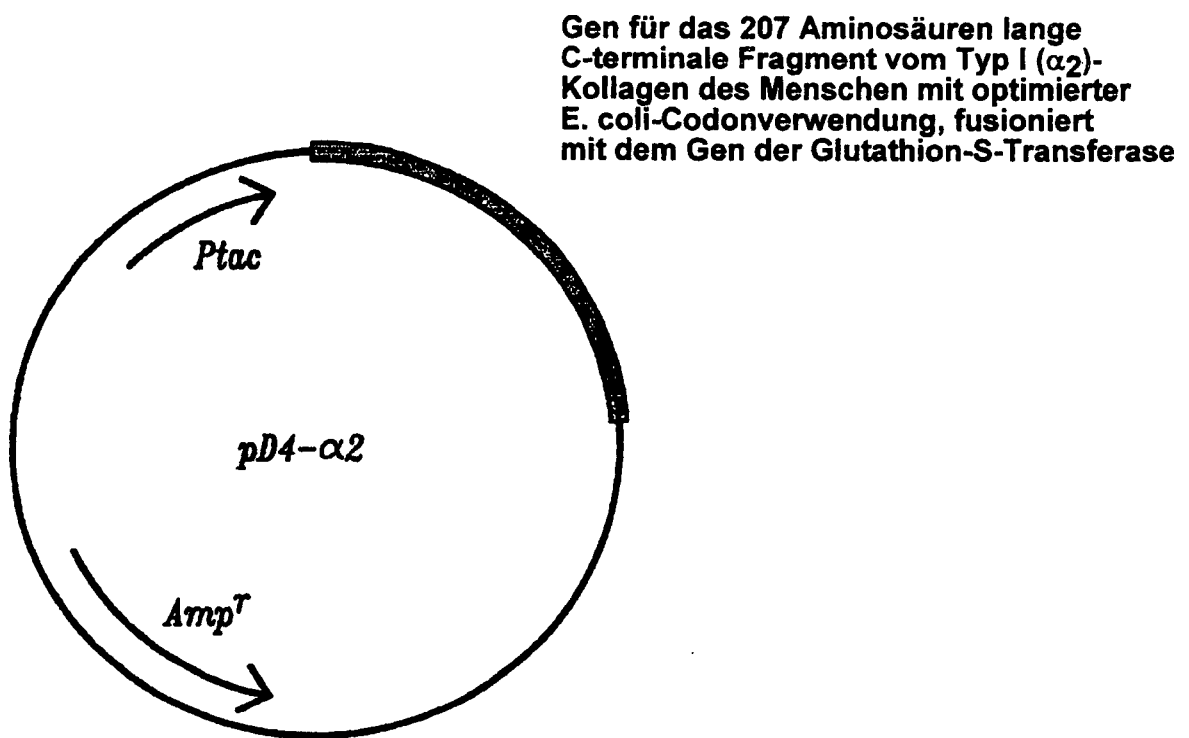


FIG. 68

**Proteinsequenz der ersten 13 Aminosäuren von D4- α 1,
vorhergesagt aufgrund der DNA-Sequenz:**

H₂N-Gly-Pro-Pro-Gly-Leu-Ala-Gly-Pro-Pro-Gly-Glu-Ser-Gly

**Experimentell bestimmte Proteinsequenz der ersten
13 Aminosäuren von D4- α 1:**

H₂N-Gly-Hyp-Hyp-Gly-Leu-Ala-Gly-Hyp-Hyp-Gly-Glu-Ser-Gly

FIG. 69

6050 1 (3.963) Sm (SG, 4x2.00); Tr (895:1593,0.13,Mid); Sm (SG, 3x1.50)

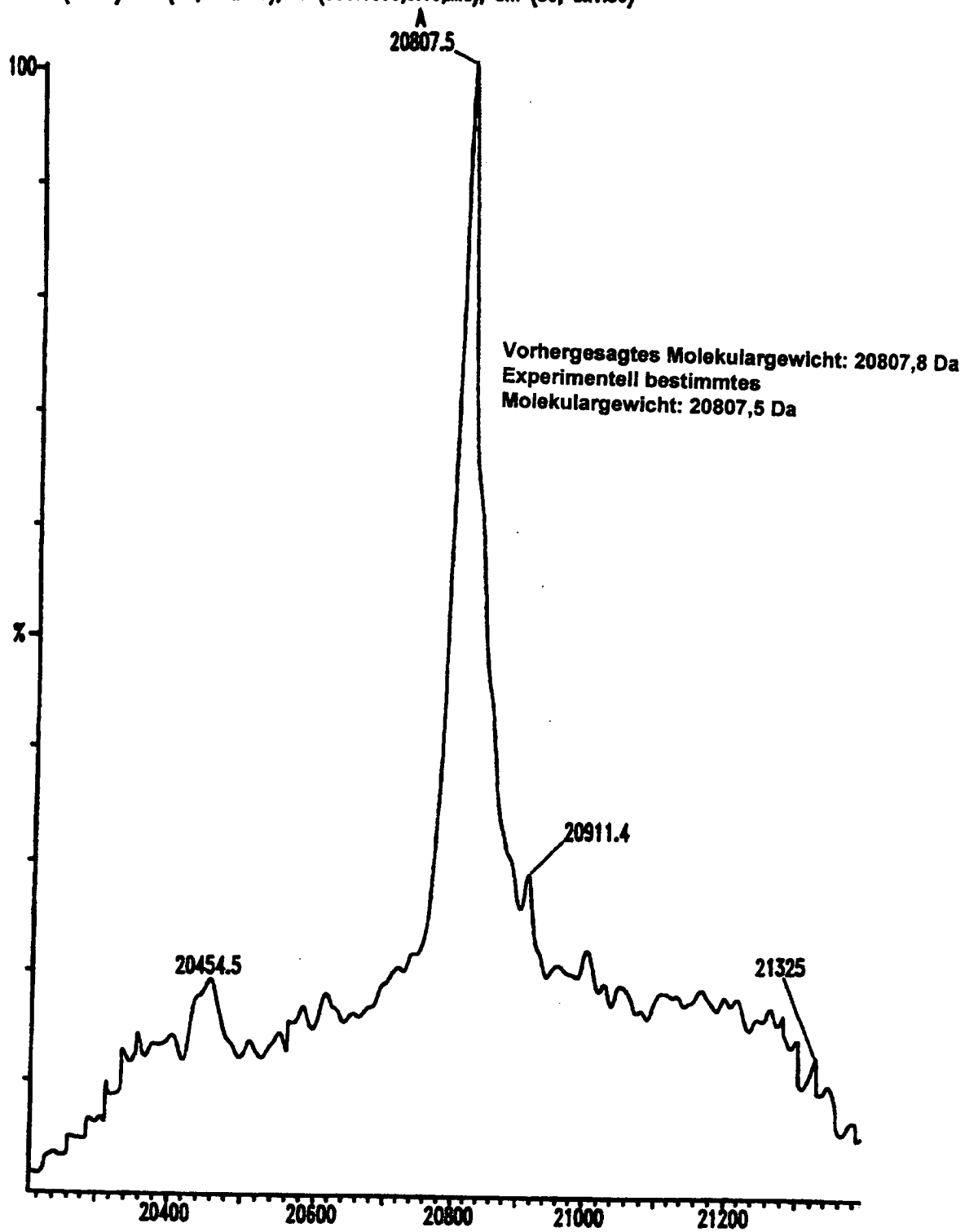


FIG. 70

ATG	GGC	CCC	CCG	GGT	CTG	GCG	GGC	CCT	CCG	GGT	GAA	AGC	GGT	CGT	GAA	GGC	GCG	CCG	GGT
		9			18			27			36			45			54		
GCC	GAA	CGC	AGC	CCA	GGC	CGC	GAC	GGT	AGC	CCG	CGG	GCC	AAA	GGG	GAT	CGT	GGT	GAA	ACC
		69			78			87			96			105			114		
GGC	CCG	GCG	CAC	CCC	CCG	GGT	GCA	CCG	GGC	GCG	CCG	GGT	GCC	CCA	GGC	CCG	GTG	GGC	CCG
		129			138			147			156			165			174		
GCG	GGC	AAA	AGC	GGT	GAT	CGT	GGT	CAG	ACC	GGT	CCG	GCG	GGC	CCG	GCC	GGT	CCG	GTG	GGC
		189			198			207			216			225			234		
CCA	GCG	GGC	GCC	CGT	GGC	CCG	GCC	GGT	CCG	CAG	GGC	CCG	CGG	GGT	GAC	AAA	GGT	GAA	ACG
		249			258			267			276			285			294		
GGC	GAA	CAG	GGC	GAC	CGT	GCC	ATT	AAA	GGC	CAC	CGT	GCC	TTC	AGC	GGC	CTG	CAG	GGT	CCA
		309			318			327			336			345			354		
CCG	GGC	CCG	CCG	GGC	AGT	CCG	GGT	GAA	CAG	GGT	CCG	TCC	GGA	GCC	AGC	GGG	CCG	GCG	GGC
		369			378			387			396			405			414		
CCA	CGC	GGT	CCG	CCG	GGC	AGC	GCG	GGC	GCG	CCG	GGC	AAA	GAC	GGT	CTG	AAC	GGT	CTG	CCG
		429			438			447			456			465			474		
GGC	CCG	ATC	GGC	CCG	CCG	GGC	CCA	CGC	GCG	CCG	ACC	GGT	GAT	GCG	GGT	CCG	GTG	GGT	CCC
		489			498			507			516			525			534		
CCG	GGC	CCG	CCG	GGC	CCG	CCA	GGC	CCG	CCG	GGA	CCG	CCG	AGC	GCG	GGT	TTC	GAC	TTC	AGC
		549			558			567			576			585			594		
TTC	CTG	CCG	CAG	CCG	CCG	CAG	GAG	AAA	GCG	CAC	GAC	GCG	GGT	CGC	TAC	TAC	CGT	GCG	TAA
		609			618			627			636			645			654		
...	...	669			678			687			696			705			714		
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...

FIG. 71

10	20	30	40	50	60
MGPFGLAGPP	GLSGREGAPG	AEGSPGRDUS	PGAKGDRGET	GPAGPPGAPG	APGAPGPVGP
70	80	90	100	110	120
AGKSGDRGET	GPAGPAGPVG	PAGARGPAGP	QGPRGDKGET	GEQGDRGIKG	HRGFSGLQGP
130	140	150	160	170	180
PGPFGSPCEQ	GP'GASGPAG	FRGPFGSAGA	PGKDGLNGLP	GPIGPPGPRG	RTGDAGPVG
190	200	210	220	230	240
PGPPGPFGPP	GPPEAGFDFS	FLPQPFQEKA	HGGGYRA*		

FIG. 72

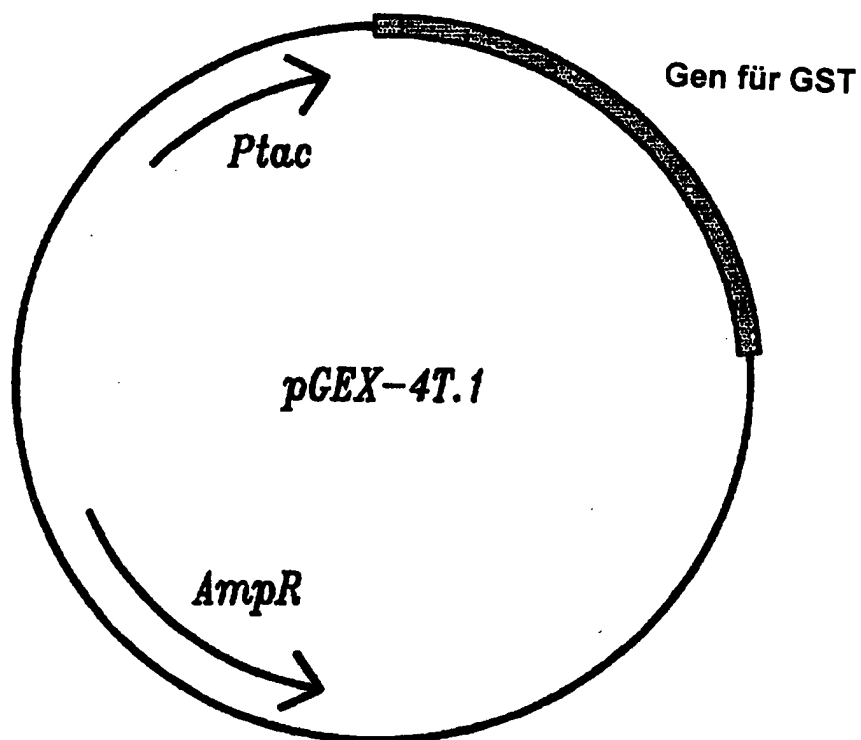


FIG. 73

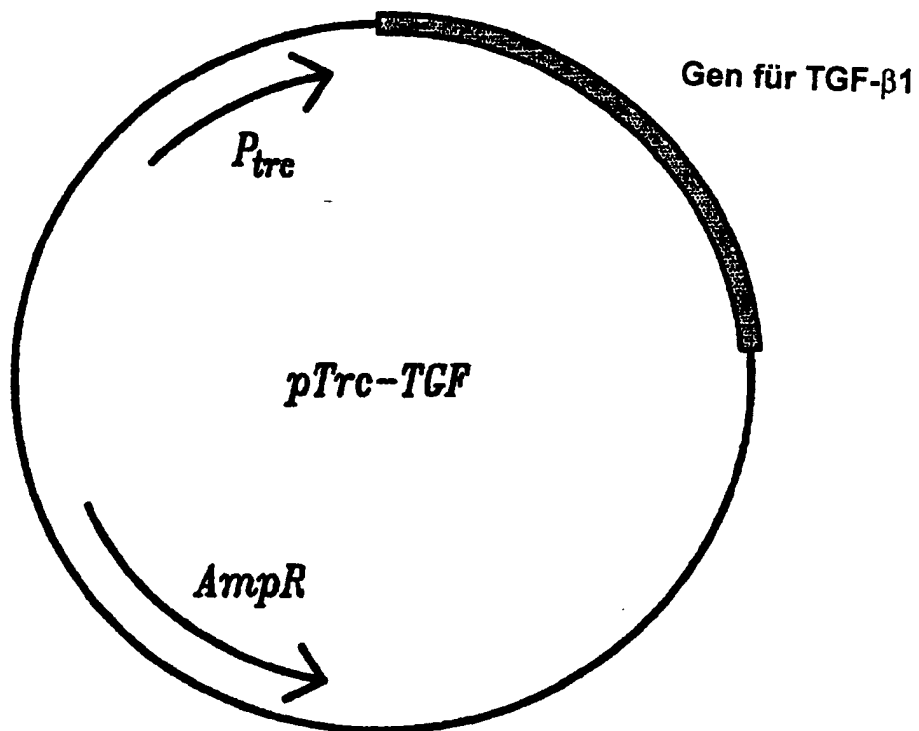


FIG. 74

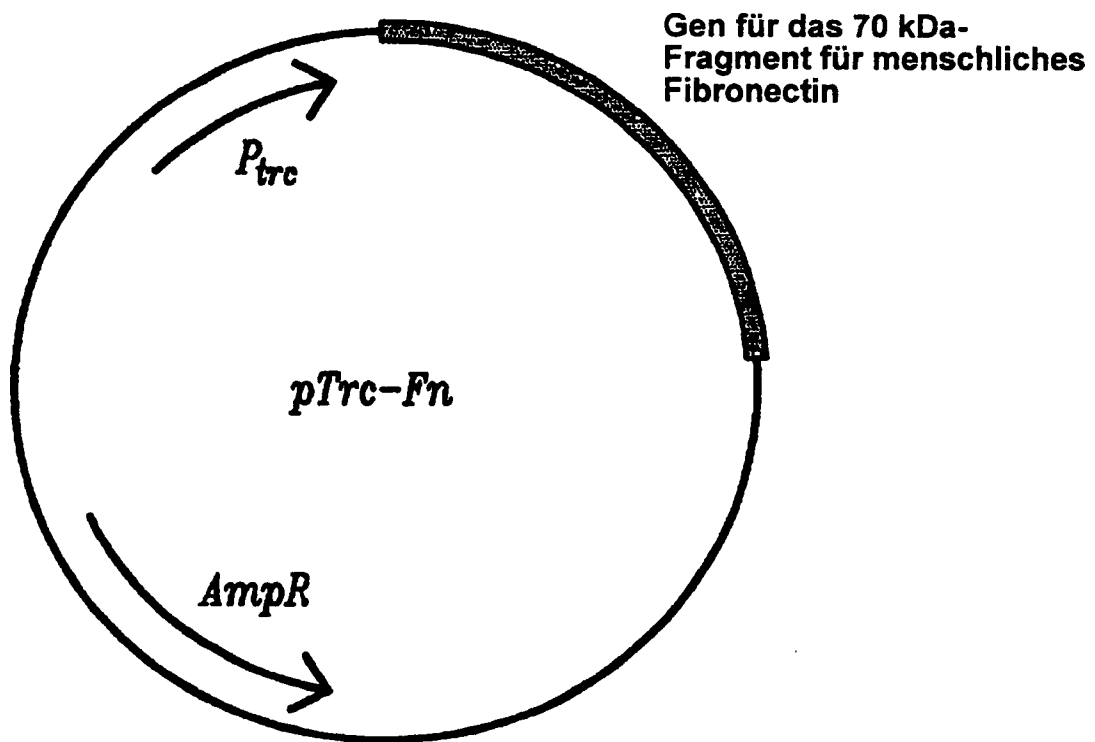


FIG. 75

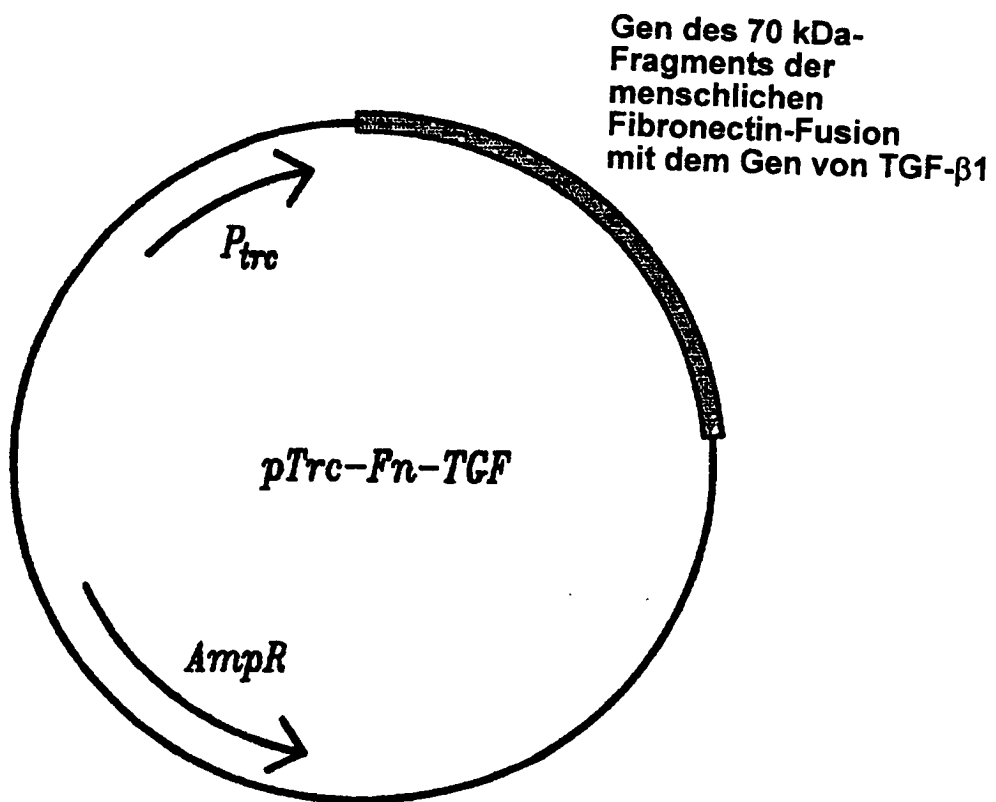


FIG. 76

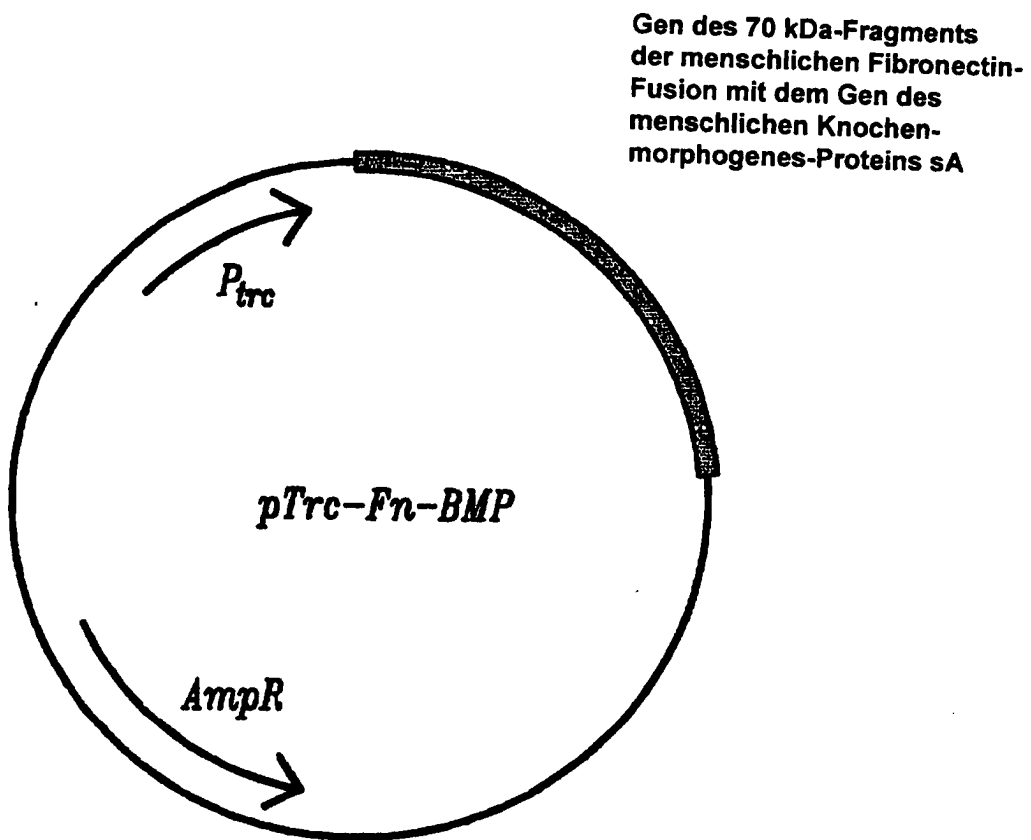


FIG. 77

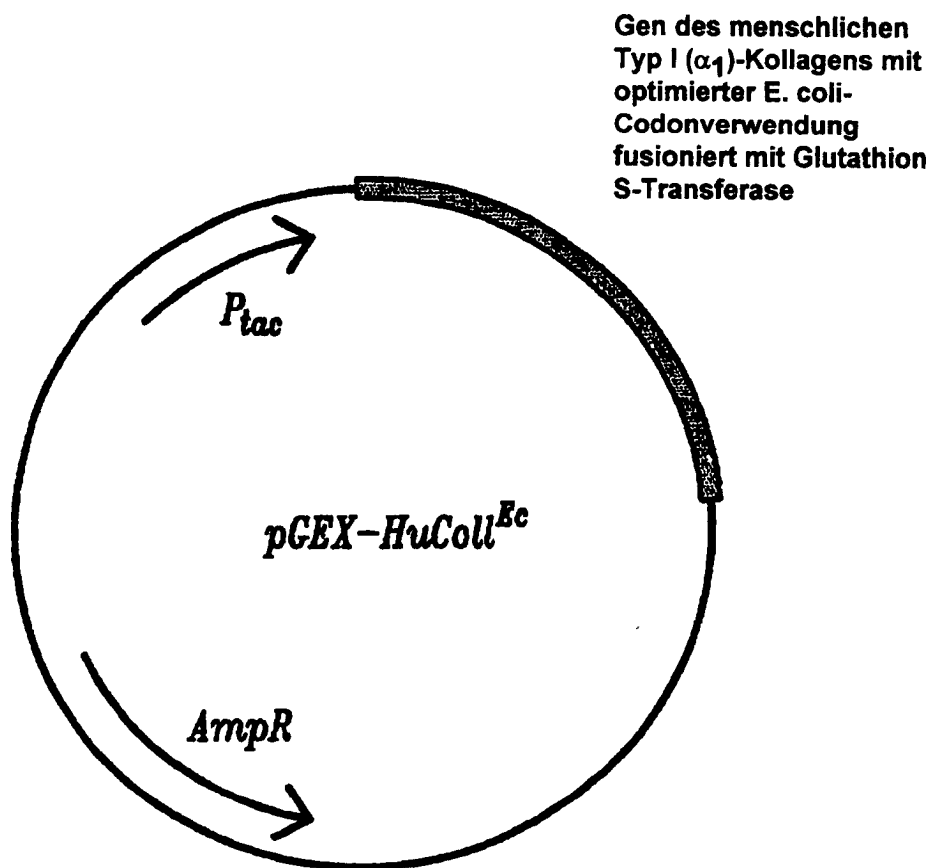


FIG. 78

ATGGGCTCTCCGGGTGTTAACGGCGCCCCTGGTGAAGCGGGCCGCGACGGCAATCCGGG
CAACGATGGTCCGCCGGGTCGTGATGGTCAGCCGGGTCACAAAGGTGAGCGTGGCTACC
CGGGTAACATCGGTCCGGTTGGTGCGGCCGGCGCTCCGGGTCCGCACGGTCCGGTAGGC
CCAGCCGGCAAACACGGTAACCGTGGTGAACCGGGTCCGTCCGGTCCGGTAGGTCCGGC
GGGTGCTGTTGGTCCACGCGGCCCGTCCGGCCCCGCAGGGTATTGCGGGTGACAAAGGCG
AACC GGCGAAAAAGGTCCGCGTGGTCTGCCGGGCCTTAAGGGCCACAACGGTCTGCAA
GGTCTGCCGGGTATCGCGGGTCACCACGGTGATCAGGGTGCTCCGGGTTCCGTTGGTCCG
GCCGGTCCGCGTGGCCCCGGCTGGTCCGTCTGGTCCGGCCGGTAAAGACGGCCGTACGGG
CCACCCGGGTACGGTGGGTCCGGCCGGCATTGCGGGTCCGCAAGGTACACAGGGTCCGG
CGGGTCCGCCGGGTCCGCCGGGTCCGCCGGGTCCGCCGGGTGTTAGCGGTGGCGGTTAT
GATTTTGGTTATGACGGTGATTTCTATCGTGCGTAA

FIG. 79

MGPPGLAGPPGESGREGAPGAEGSPGRDGSPGAKGDRGETGPAGPPGAPGAPGAPGPVGPA
GKSGDRGETGPAGPAGPVGPAGARGPAGPQGPRGDKGETGEQGDRGIKGHRGFSLQGPPG
PPGSPGEQGPPSGASGPAGPRGPPGSAGAPGKDGLNGLPGPIGPPGPRGRTGDAGPVGPPGPPG
PPGPPGPPSAGFDFSFLPQPPQEK AHDGG RYYRA

FIG. 80

Oligo N4-1

5'GGAATTCTCCCATGGGCCCCGCCGGGTCTGGCGGGCCCTCCGGGTGAAAGCGGTCGTGA
AGGCGCGCCGGGTGCCGAAGGCAGCCCAGGCCGCGAC

Oligo N4-2

3'CTTCCGTCGGGTCCGGCGCTGCCATCGGGCCCCCGGTTTCCCCTAGCACCACCTTGGCC
GGGCCGCCCCGGGGGGCCCACGTGGCATTATTCGAACCC

Oligo N4-3

5'GGAATTCGGTGCACCGGGCGCGCCGGGTGCCCCAGGCGCGGTGGGCCCCGGCGGGCAAA
AGCGGTGATCGTGGCGAGACCGGTCCGGCGGGC

Oligo N4-4

3'CTCTGGCCAGGCCGCCCCGGGCCGGCCAGGCCACCCGGGTGCCCCGCGGGCACC GGCC
GGCCAGGCGTCCCGGGCGCCATTATTCGAACCC

FIG. 8I