

PCT

WELTORGANISATION FÜR GEISTIGES EIGENTUM
Internationales Büro



INTERNATIONALE ANMELDUNG VERÖFFENTLICHT NACH DEM VERTRAG ÜBER DIE
INTERNATIONALE ZUSAMMENARBEIT AUF DEM GEBIET DES PATENTWESENS (PCT)

(51) Internationale Patentklassifikation ⁷ : C12N 15/12, 15/86, 5/10, 1/21, 1/19, C07K 14/47, 16/18, A61K 38/17, 48/00, 39/395, G01N 33/68, C12Q 1/68	A2	(11) Internationale Veröffentlichungsnummer: WO 00/08143 (43) Internationales Veröffentlichungsdatum: 17. Februar 2000 (17.02.00)
--	-----------	--

(21) Internationales Aktenzeichen: PCT/DE99/02479

(22) Internationales Anmeldedatum: 6. August 1999 (06.08.99)

(30) Prioritätsdaten:
198 35 910.1 7. August 1998 (07.08.98) DE

(71) Anmelder (für alle Bestimmungsstaaten ausser US):
DEUTSCHES KREBSFORSCHUNGSZENTRUM
STIFTUNG DES ÖFFENTLICHEN RECHTS [DE/DE];
Im Neuenheimer Feld 280, D-69120 Heidelberg (DE).
MAX-PLANCK-INSTITUT FÜR MOLEKULARE
GENETIK [DE/DE]; Ihnestrasse 73, D-14195 Berlin (DE).

(72) Erfinder; und

(75) Erfinder/Anmelder (nur für US): SERANSKI, Peter
[DE/DE]; Humboldtstrasse 28, D-69120 Heidelberg (DE).
POUSTKA, Annemarie [DE/DE]; Werderstrasse 36,
D-69120 Heidelberg (DE). LEHRRACH, Hans [DE/DE];
Lützelsteiner Weg 50, D-14195 Berlin (DE). RADELOF,
Uwe [DE/DE]; Ahornstrasse 25, D-12163 Berlin (DE).

(74) Anwalt: SCHÜSSLER, Andrea; Truderinger Strasse 246,
D-81825 München (DE).

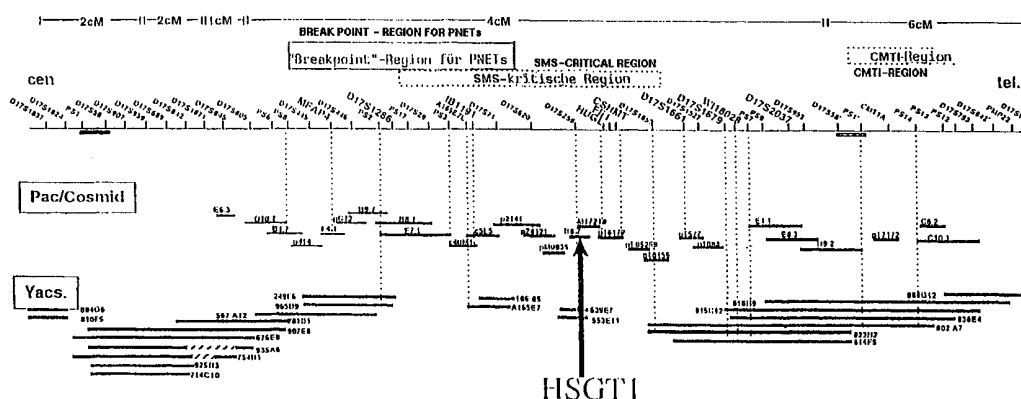
(81) Bestimmungsstaaten: AL, AM, AT, AU, AZ, BA, BB, BG,
BR, BY, CA, CH, CN, CU, CZ, DK, EE, ES, FI, GB, GD,
GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP,
KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK,
MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI,
SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW,
ARIPO Patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG,
ZW), eurasisches Patent (AM, AZ, BY, KG, KZ, MD, RU,
TJ, TM), europäisches Patent (AT, BE, CH, CY, DE, DK,
ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI
Patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR,
NE, SN, TD, TG).

Veröffentlicht

Ohne internationalen Recherchenbericht und erneut zu
veröffentlichen nach Erhalt des Berichts.

(54) Title: GENE ISOLATED ON THE SHORT ARM OF HUMAN CHROMOSOME 17

(54) Bezeichnung: GEN ISOLIERT AUF DEM KURZEN ARM DES MENSCHLICHEN CHROMOSOMS 17



(57) Abstract

The present invention relates to DNA molecules containing information for a human gene that was isolated on the short arm of chromosome 17. The invention also relates to vectors containing said DNA molecules. In one embodiment, said vectors are suitable for expression of the inventive DNA molecules in prokaryotes or eukaryotes. The invention further relates to medicaments and diagnostic compositions containing said DNA molecules or vectors, the protein coded thereby or an antibody that specifically recognizes said protein.

(57) Zusammenfassung

Die vorliegende Erfindung betrifft DNA-Moleküle, die die Information für ein menschliches Gen, das auf dem kurzen Arm von Chromosom 17 isoliert worden ist, enthalten. Die vorliegende Erfindung stellt ferner diese DNA-Moleküle enthaltende Vektoren bereit, wobei in einer bevorzugten Ausführungsform diese Vektoren zur Expression der erfindungsgemässen DNA-Moleküle in Prokaryonten oder Eukaryonten geeignet sind. Ferner betrifft die vorliegende Erfindung Arzneimittel und diagnostische Zusammensetzungen, die die vorstehenden DNA-Moleküle bzw. Vektoren enthalten, das davon codierte Protein oder einen dieses Protein spezifisch erkennenden Antikörper.

LEDIGLICH ZUR INFORMATION

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

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

Gen isoliert auf dem kurzen Arm des menschlichen Chromosoms 17

Die vorliegende Erfindung betrifft DNA-Moleküle, die die Information für ein menschliches Gen, das auf dem kurzen Arm von Chromsom 17 isoliert worden ist, enthalten. Die vorliegende Erfindung stellt ferner diese DNA-Moleküle enthaltende Vektoren bereit, wobei in einer bevorzugten Ausführungsform diese Vektoren zur Expression der erfindungsgemäßen DNA-Moleküle in Prokaryonten oder Eukaryonten geeignet sind. Ferner betrifft die vorliegende Erfindung Arzneimittel und diagnostische Zusammensetzungen, die die vorstehenden DNA-Moleküle bzw. Vektoren enthalten, das davon codierte Protein oder einen dieses Protein spezifisch erkennenden Antikörper.

Das Smith-Magenis-Syndrom (SMS) ist eine multiple kongenitale Anomalie. Die davon betroffenen Patienten leiden an einer mentalen Retardierung, Größenzunehmungsstörungen, Abnormalitäten der REM-Schlafphase und einem abnormalen Verhalten (Hamill et al., Ann. Gent. 31, S. 36-38 (1990)). Es wurde inzwischen herausgefunden, daß SMS auf einer internen Deletion/Mutation in der p11.2-Region von Chromsom 17 beruht. Generell wurde herausgefunden, daß auch andere Mutationen in dieser Region zu neurodegenerativen Erkrankungen führen, die teilweise von den oben genannten Symptomen begleitet sind.

Da die Gene in dieser Region noch nicht vollständig bekannt sind, konnten die molekularen Mechanismen, die zu obigen Symptomen, insbesondere in Verbindung mit Smith Magenies Syndrom führen, noch nicht aufgeklärt werden.

Somit liegt der Erfindung im wesentlichen das technische Problem zugrunde, die molekularen Grundlagen zu schaffen, um neurodegenerative Erkrankungen, die verglichen mit dem Smith Magenies Syndrom ähnliche oder gleiche Symptome hervorrufen, aufzuklären sowie eine geeignete Diagnostik und/oder Therapie dafür bereitzustellen.

Die Lösung dieses technischen Problems wurde durch die Bereitstellung der in den Patentansprüchen gekennzeichneten Ausführungsformen erreicht. Es konnte ein DNA-Molekül isoliert und identifiziert werden, das in Zusammenhang mit SMS steht.

5

Die Identifizierung eines Gens (HSGT1), das bei seiner teilweisen oder vollständigen Deletion mit SMS in Verbindung steht, ist von Interesse, da dies einen Beitrag zur Aufklärung der molekularen Pathogenese der SMS und somit verbesserter Therapien gewährleistet. Parallel dazu bildet die Klonierung des SMS-Gens die Grundlage zur Entwicklung diagnostischer Tests, um zukünftig eine zuverlässigere und frühzeitige Diagnose bei Betroffenen zu gewährleisten. Darüberhinaus werden funktionelle Analysen des Proteins zweifellos zum Verständnis der Neuronenentwicklung beitragen, da das Gen in einer Region im Genom liegt (humanes Chromosom 17p11.2; s. Fig. 4), die mit verschiedenen neurologischen Erkrankungen assoziiert ist. Es ist somit als ein Kandidatengen für Untersuchungen der Pathomechanismen anzusehen, die neurodegenerativen Krankheiten und entwicklungsbedingten Störungen der Differenzierung des Zentralnervensystems zugrunde liegen.

20 Somit betrifft die vorliegende Erfindung ein DNA-Molekül (HSGT1), die folgendes umfaßt:

- (a) die Nukleinsäure von Fig. 1, oder eine hiervon durch ein oder mehrere Basenpaare unterschiedliche DNA bzw. ein Fragment davon,
- 25 (b) eine mit der Nucleinsäure von (a) hybridisierende DNA, oder
- (c) eine mit der Nucleinsäure von (a) oder (b) über den degenerierten genetischen Code verwandte DNA.

Die unter (a) und (c) definierten Nucleinsäuremoleküle codieren für Proteine, Polypeptide oder Peptide, die mindestens noch eine der nachstehend beschriebenen biologischen Aktivitäten des von der Nukleinsäure gemäß Fig. 1 codierten Proteins aufweisen, beispielsweise als Differenzierungsfaktor bei der Neuronen-

30

entwicklung und/oder Intermediärfilament in Neuronen, oder dessen Veränderung oder Ausfall zu SMS führt oder führen kann.

Die unter (a) definierten DNA-Moleküle umfassen auch DNA-Moleküle, die sich gegenüber der in Figur 1 angegebenen Sequenz durch Deletion(en), Insertion(en), Austausch(e) und/oder andere im Stand der Technik bekannte Modifikationen, z.B. alternatives Splicing, unterscheiden bzw. ein Fragment des ursprünglichen Nucleinsäuremoleküls umfassen, wobei das durch diese DNA-Moleküle codierte Protein noch eine oder mehrere der vorstehend beschriebenen biologischen Aktivitäten aufweist. Dazu zählen auch Allelvarianten. Verfahren zur Erzeugung der vorstehenden Änderungen in der Nucleinsäuresequenz sind dem Fachmann bekannt und in Standardwerken der Molekularbiologie beschrieben, beispielsweise in Sambrook et al., Molecular Cloning: A Laboratory Manual, 2. Ausgabe, Cold Spring Harbor Laboratory Press, Cold Spring Harbor NY (1989). Der Fachmann ist auch in der Lage zu bestimmen, ob ein von einer so veränderten Nucleinsäuresequenz codiertes Protein noch über eine der biologischen Aktivitäten von HSGT1 verfügt. Dies kann beispielsweise dadurch erfolgen, daß untersucht wird, ob in Zellkulturmodellen das veränderte Protein die gleichen Funktionen übernimmt wie das Wildtyp-HSGT1-Protein.

Der Begriff "hybridisierende DNA" weist auf eine DNA hin, die unter üblichen Bedingungen, insbesondere bei 20°C unter dem Schmelzpunkt der DNA, mit einer DNA von (a) hybridisiert. Der Begriff "hybridisieren" bezieht sich dabei auf konventionelle Hybridisierungsbedingungen, vorzugsweise auf Hybridisierungsbedingungen, bei denen als Lösung 5xSSPE, 1% SDS, 1xDenhardts-Lösung verwendet wird und die Hybridisierungstemperaturen zwischen 35°C und 70°C, vorzugsweise bei 65°C liegen. Nach der Hybridisierung wird vorzugsweise zuerst mit 2xSSC, 1% SDS und danach mit 0,2xSSC bei Temperaturen zwischen 35°C und 70°C, vorzugsweise bei 65°C gewaschen (zur Definition von SSPE, SSC und Denhardts-Lösung siehe Sambrook et al., supra). Besonders bevorzugt sind stringente Hybridisierungsbedingungen, wie sie beispielsweise in Sambrook et al., supra, beschrieben sind.

In einer besonders bevorzugten Ausführungsform ist das erfindungsgemäße DNA-Molekül eine cDNA.

5 In einer weiteren bevorzugten Ausführungsform ist das erfindungsgemäße DNA-Molekül eine genomische DNA, die vorzugsweise von einem Säuger, beispielsweise einem Menschen stammt. Auf Nucleinsäurehybridisierung basierende Screening-Verfahren erlauben die Isolierung der erfindungsgemäßen genomischen DNA-Moleküle aus jedem Organismus bzw. abgeleiteten genomischen DNA-Banken, wobei Sonden verwendet werden, die die in Figur 1 angegebene Nucleinsäuresequenz oder einen Teil davon enthalten. Anhand der genomischen DNA, beispielsweise aus einem Menschen, können weitere Mutationen identifiziert werden, die an der Entstehung von neurodegenerativen Erkrankungen, insbesondere SMS, beteiligt sind, beispielsweise Mutationen, die zu Aminosäureaustauschen, zur Hemmung des Spleißens oder zu fehlerhaftem Spleißen führen.

15 Interessanterweise wurde im HSGT1-Gen ein Polymorphismus gefunden, der sich in der Protein-kodierenden Region des Gens befindet und für eine Polyglutaminanhäufung codiert, d.h. es kommt eine variable Wiederholung der Sequenz CAG (vgl. Fig. 1) vor. Polyglutaminanhäufungen sind bereits in verschiedenen Genen in Zusammenhang mit neuropathologischen Erkrankungen, z.B. Chorea Huntington, beschrieben.

20 Die erfindungsgemäßen Nucleinsäuren können auch in einen Vektor bzw. Expressionsvektor inseriert werden. Somit umfaßt die vorliegende Erfindung auch diese Nucleinsäuremoleküle enthaltende Vektoren. Beispiele solcher sind dem Fachmann bekannt. Im Falle eines Expressionsvektors für E. coli sind dies z.B. pGEMEX, pUC-Derivate (z.B. pUC8), pBR322, pBlueScript, pGEX-2T, pET3b und pQE-8. Für die Expression in Hefe sind z.B. pY100 und Ycpad1 zu nennen, während für die Expression in tierischen Zellen z.B. pKCR, pEFBOS, cDM8 und pCEV4, anzugeben sind. Für die Expression in Insektenzellen eignet sich besonders der Baculovirus-Expressionsvektor pAcSGHisNT-A. In einer bevorzugten Ausführungsform ist das erfindungsgemäße Nucleinsäuremolekül im Vektor mit regulatorischen

Elementen funktionell verknüpft, die dessen Expression in prokaryotischen oder eukaryotischen Wirtszellen erlauben. Solche Vektoren enthalten neben den regulatorischen Elementen, beispielsweise einem Promotor, typischerweise einen Replikationsursprung und spezifische Gene, die die phänotypische Selektion einer transformierten Wirtszelle erlauben. Zu den regulatorischen Elementen für die Expression in Prokaryonten, beispielsweise E.coli, zählen der lac-, trp-Promotor oder T7-Promotor, und für die Expression in Eukaryonten der AOX1- oder GAL1-Promotor in Hefe, und der CMV-, SV40-, RVS-40-Promotor, CMV- oder SV40-Enhancer für die Expression in tierischen Zellen. Weitere Beispiele für geeignete Promotoren sind der Metallothionein I- und der Polyhedrin-Promotor.

Zu geeigneten Vektoren zählen insbesondere auch T7 basierende Expressionsvektoren für die Expression in Bakterien (Rosenberg et al., Gene 56(1987), 125) oder pMSXND für die Expression in Säugerzellen (Lee und Nathans, J.Biol.Chem. 263 (1988), 3521).

In einer bevorzugten Ausführungsform ist der die erfindungsgemäßen DNA-Moleküle enthaltene Vektor ein Virus, beispielsweise ein Adenovirus, Vaccinia-Virus oder ein AAV-Virus, der bei einer Gentherapie von Nutzen ist. Besonders bevorzugt sind Retroviren. Beispiele für geeignete Retroviren sind MoMuLV, HaMuSV, MuMTV, RSV oder GaLV. Für Zwecke der Gentherapie können die erfindungsgemäßen DNA-Moleküle auch in Form von kolloidalen Dispersionen zu den Zielzellen transportiert werden. Dazu zählen beispielsweise Liposomen oder Lipoplexe (Mannino et al., Biotechniques 6 (1988), 682).

Allgemeine auf dem Fachgebiet bekannte Verfahren können zur Konstruktion von Expressionsvektoren, die die erfindungsgemäßen DNA-Moleküle und geeignete Kontrollsequenzen enthalten, verwendet werden. Zu diesen Verfahren zählen beispielsweise in vitro-Rekombinationstechniken, synthetische Verfahren, sowie in vivo-Rekombinationsverfahren, wie sie beispielsweise in Sambrook et al., supra, beschrieben sind.

Die vorliegende Erfindung betrifft auch die vorstehend beschriebenen Vektoren enthaltende Wirtszellen. Zu diesen Wirtszellen zählen Bakterien, Hefe, Insekten- und Tierzellen, vorzugsweise Säugerzellen. Bevorzugt sind die *E. coli* Stämme HB101, DH1, x1776, JM101, JM109, BL21, XL1Blue und SG 13009, der Hefestamm *Saccharomyces cerevisiae* und die tierischen Zellen L, 3T3, FM3A, CHO, COS, Vero, HeLa sowie die Insektenzellen sf9. Verfahren zur Transformation dieser Wirtszellen, zur phänotypischen Selektion von Transformanten und zur Expression der erfindungsgemäßen DNA-Moleküle unter Verwendung der vorstehend beschriebenen Vektoren sind auf dem Fachgebiet bekannt.

Ein Klon, der für eine erfindungsgemäße Nukleinsäure kodiert, wurde bei der DSMZ, Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH, Mascheroder Weg 1b, 38124 Braunschweig gemäß dem Budapester Vertrag unter der Hinterlegungsnummer DSM 12357 (Hinterlegungsdatum: 31. Juli 1998) bzw. DSM 12472 (Hinterlegungsdatum: 16. September 1998) [= *E. coli*/PSB8.7-l13281] hinterlegt.

Die vorliegende Erfindung betrifft auch ein Verfahren zur Herstellung eines Proteins, das durch obige Nukleinsäure kodiert wird, umfassend die Kultivierung der vorstehend beschriebenen Wirtszellen unter Bedingungen, die die Expression des Proteins erlauben (vorzugsweise stabile Expression), und Gewinnung des Proteins aus der Kultur. Geeignete Verfahren zur rekombinanten Herstellung des Proteins sind allgemein bekannt (siehe beispielsweise Holmgren, *Annu.Rev.Biochem.* 54 (1985), 237; LaVallie et al., *Bio/Technology* 11 (1993), 187; Wong, *Curr.Opin.Biotech.* 6 (1995), 517; Romanos, *Curr.Opin.Biotech.* 6 (1995), 527; Williams et al., *Curr. Opin. Biotech.* 6 (1995), 538; und Davies, *Curr. Opin.Biotech.* 6 (1995), 543. Auch geeignete Reinigungsverfahren (beispielsweise präparative Chromatographie, Affinitätschromatographie, beispielsweise Immunoaffinitätschromatographie, HPLC etc.) sind allgemein bekannt.

In einer weiteren Ausführungsform betrifft die vorliegende Erfindung ein von den erfindungsgemäßen DNA-Molekülen (HSGT1) codiertes oder nach dem vor-

stehenden Verfahren erhaltenes Protein. Die Aminosäuresequenz eines solchen Proteins ist in Fig. 2 gezeigt. In diesem Zusammenhang soll erwähnt werden, daß das erfindungsgemäße Protein gemäß üblicher, auf dem Fachgebiet bekannten, Verfahren modifiziert sein kann. Zu diesen Modifikationen zählen Austausche, Insertionen oder Deletionen von Aminosäuren, die die Struktur des Proteins modifizieren, wobei seine biologische Aktivität im wesentlichen erhalten bleibt. Zu den Austauschen zählen vorzugsweise "konservative" Austausche von Aminosäureresten, d.h. Austausche gegen biologisch ähnliche Reste, z.B. die Substitution eines hydrophoben Rests (z.B. Isoleucin, Valin, Leucin, Methionin) gegen einen anderen hydrophoben Rest, oder die Substitution eines polaren Rests gegen einen anderen polaren Rest (z.B. Arginin gegen Lysin, Glutaminsäure gegen Asparaginsäure etc.). Deletionen können zur Erzeugung von Molekülen führen, die eine deutlich geringere Größe aufweisen, d.h., denen beispielsweise Aminosäuren am N- oder C-Terminus fehlen. Um noch die Funktion eines erfindungsgemäßen Proteins zu erfüllen, sollte auf Aminosäureebene zu der Sequenz von Fig. 2 eine Homologie von mind. 50% vorliegen.

Die vorliegende Erfindung betrifft auch Antikörper, die das vorstehend beschriebene Protein HSGT1 spezifisch erkennen. Die Antikörper können monoclonale, polyclonale oder synthetische Antikörper sein oder Fragmente davon, beispielsweise Fab-, Fv-oder scFv-Fragmente. Vorzugsweise handelt es sich dabei um monoclonale Antikörper. Für die Herstellung ist es günstig, Tiere, insbesondere Kaninchen oder Hühner für einen polyclonalen und Mäuse für einen monoclonalen Antikörper, mit einem vorstehenden (Fusions)protein oder Fragmenten davon zu immunisieren. Weitere "Booster" der Tiere können mit dem gleichen (Fusions)protein oder Fragmenten davon erfolgen. Der polyclonale Antikörper kann dann aus dem Serum bzw. Eigelb der Tiere erhalten werden. Die erfindungsgemäßen Antikörper können gemäß Standardverfahren hergestellt werden, wobei das von den erfindungsgemäßen DNA-Molekülen codierte Protein oder ein synthetisches Fragment davon als Immunogen dienen. Monoclonale Antikörper können beispielsweise durch das von Köhler und Milstein (Nature 256 (1975), 495) und Galfré (Meth. Enzymol. 73 (1981), 3) beschriebene Verfahren hergestellt wer-

den, wobei Maus-Myelomzellen mit von immunisierten Säugern stammenden Milzzellen fusioniert werden. Diese Antikörper können beispielsweise zur Immunpräzipitation der vorstehend diskutierten Proteine oder zur Isolierung verwandter Proteine aus cDNA-Expressionsbanken verwendet werden. Die Antikörper können
5 beispielsweise in Immunoassays in Flüssigphase oder an einen festen Träger gebunden werden. Dabei können die Antikörper auf verschiedene Art und Weise markiert sein. Geeignete Marker und Markierungsverfahren sind dem Fachgebiet bekannt. Beispiele für Immunassays sind ELISA und RIA.

10 Die vorliegende Erfindung betrifft ferner die Verwendung der vorstehend beschriebenen DNA-Moleküle, Vektoren, Proteine und/oder Antikörper. Vorzugsweise werden diese zur Herstellung eines Arzneimittels zur Prävention oder Behandlung von neurodegenerativen Erkrankungen, insbesondere SMS, verwendet. Dabei kann das Arzneimittel in der Gentherapie Verwendung finden, wobei die vorstehend
15 beschriebenen Verfahren bzw. Vektoren zur Einschleusung der erfindungsgemäßen DNA-Moleküle Anwendung finden können. Andererseits kann das von den erfindungsgemäßen DNA-Molekülen codierte Protein direkt verabreicht werden, um so in Zellen, die keine funktionalen Kopien des erfindungsgemäßen Gens mehr besitzen, ausreichende Mengen des biologisch aktiven Proteins zur Verfügung zu
20 stellen. Das erfindungsgemäße Arzneimittel kommt bevorzugt dann zur Anwendung, wenn in dem betreffenden Patienten biologisch aktives Protein nicht oder in zu geringen Mengen gebildet werden. Es gibt Hinweise darauf, daß dieses Gen bei der Entstehung von Autismus eine Rolle spielt.

25 Diese Arzneimittel enthalten gegebenenfalls zusätzlich einen pharmazeutisch verträglichen Träger. Geeignete Träger und die Formulierung derartiger Arzneimittel sind dem Fachmann bekannt. Zu geeigneten Trägern zählen beispielsweise Phosphat-gepufferte Kochsalzlösungen, Wasser, Emulsionen, beispielsweise Öl/Wasser-Emulsionen, Netzmittel, sterile Lösungen etc. Die Verabreichung der
30 Arzneimittel kann oral oder parenteral erfolgen. Zu den Verfahren für die parenterale Verabreichung gehören die topische, intra-arterielle, intramuskuläre, subkutane, intramedulläre, intrathekale, intraventrikuläre, intravenöse, intraperitoneale

oder intranasale Verabreichung. Die geeignete Dosierung wird von dem behandelnden Arzt bestimmt und hängt von verschiedenen Faktoren ab, beispielsweise von dem Alter, dem Geschlecht, dem Gewicht des Patienten, dem Stadium der Erkrankung, der Art der Verabreichung etc.

5

Die vorliegende Erfindung betrifft ferner eine diagnostische Zusammensetzung, die das vorstehend beschriebene DNA-Molekül oder den Antikörper enthält oder Kombinationen davon, gegebenenfalls zusammen mit einem geeigneten Nachweismittel.

10

Das erfindungsgemäße, wie vorstehend definierte DNA-Molekül kann auch als Sonde verwendet werden, um DNA-Moleküle zu isolieren, die beispielsweise von einer anderen Spezies oder einem anderen Organismus stammen und ein Protein codieren mit einer ebensolchen biologischen Aktivität. Vorzugsweise weist dazu die Sonde eine Länge von mindestens 10, besonders bevorzugt mindestens 15 Basen auf. Geeignete, auf Hybridisierung basierende Nachweisverfahren sind dem Fachmann bekannt. Geeignete Markierungen für die Sonde sind dem Fachmann ebenfalls bekannt und dazu zählen beispielsweise Markierung mit Radioisotopen, Biolumineszenz-, Chemilumineszenz-, Fluoreszenzmarkern, Metallchelaten, Enzymen etc..

20

Darüber hinaus kann dies auch durch eine PCR (Wiedmann et al., PCR Methods Appl. 3, S. 551-564 (1994); Saiki et al., Nature 324, S. 163-166 (1986)) oder "Ligase chain reaction" (LCR) (Taylor et al., Curr. Opin. Biotechnol. 6, S. 24-29 (1995); Rouwendal et al., Methods Mol. Biol. S. 149-156 (1996)) erfolgen, wobei die Primer von der Sequenz in Figur 1 abgeleitet sind und wobei geeignete Primer (hinsichtlich Länge, Komplementarität zur Matritze, dem zu amplifizierenden Bereich etc.) vom Fachmann gemäß üblicher Verfahren entworfen werden können.

25

30

Die vorstehend beschriebenen Verfahren erlauben auch die Identifizierung weiterer Mutationen, die in Zusammenhang mit neurodegenerativen Erkrankungen stehen, wobei die Sequenz eines Patienten mit SMS oder dazu ähnlichen Symptomen mit

gesunden Patienten verglichen wird.

Außerdem betrifft die vorliegende Erfindung ein Verfahren zur Diagnose von neurodegenerativen Erkrankungen, insbesondere SMS, in vitro, die folgenden Schritte umfassend:

- Isolierung von Nucleinsäure aus dem Patienten,
- Durchführung einer LCR oder PCR mit geeigneten Primern oder einer Hybridisierungsanalyse mit einer oder mehreren geeigneten Sonden,
- Vergleich der Längen der amplifizierten bzw. hybridisierenden Fragmente mit den Längen von Kontrollfragmenten aus dem nichtmutierten HSGT1-Gen, oder
- Sequenzierung der amplifizierten Fragmente und Vergleich mit der Sequenz von Kontrollfragmenten aus dem nichtmutierten HSGT1-Gen, wobei Längenunterschiede oder Mutationen, die zu Substitutionen, Deletionen oder Additionen von Aminosäuren oder zu vorzeitiger Termination der Translation führen, indikativ für neurodegenerative Erkrankungen, insbesondere SMS, sind.

Alternativ kann die Diagnose auch dadurch erfolgen, daß Lymphoblasten isoliert werden und Metaphase-Präparate hergestellt werden, die einer FISH-Analyse unterworfen werden, um festzustellen, ob bestimmte Bereiche des Genoms überhaupt vorhanden sind.

Bei oben genanntem Verfahren können dem Fachmann bekannte Methoden bezüglich der Präparation von DNA oder RNA aus biologischen Proben, des Restriktionsverdaus der DNA, der Auftrennung der Restriktionsfragmente auf nach Größe auftrennenden Gelen, beispielsweise Agarosegelen, der Herstellung und Markierung der Sonde und des Nachweises der Hybridisierung, beispielsweise über "Southern-Blot"; angewandt werden.

Der vorstehende Nachweis kann auch über PCR oder LCR durchgeführt werden. Dabei werden Primer verwendet, die die erfindungsgemäße Sequenz oder geeignete Teilbereiche flankieren. Diagnostisch von Bedeutung sind dabei Amplifikationsprodukte von DNA aus dem fraglichen Gewebe, die sich, beispielsweise hinsichtlich ihrer Länge und/oder der Sequenz, von den Amplifikationsprodukten von DNA aus gesundem Gewebe unterscheiden.

In einer alternativen Ausführungsform kann ein Verfahren angewendet werden, das folgende Schritte umfaßt:

- Isolierung von RNA aus dem Patienten,
- Durchführung einer Northern-Analyse mit einer oder mehreren geeigneten Sonden,
- Vergleich der Konzentration und/oder Länge HSGT1-mRNA der Patientenprobe mit einer mRNA einer gesunden Person, wobei Längenunterschiede oder das Fehlen bzw. eine geringere Konzentration der HSGT1-mRNA (beispielsweise bedingt durch Mutationen in regulatorischen Sequenzen, die mit der Transkription in Zusammenhang stehen) im Vergleich zur Kontroll-mRNA indikativ für eine neurogenerative Erkrankungen, insbesondere für SMS, ist.

Bei diesem Verfahren können dem Fachmann bekannte Methoden bezüglich der Präparation von Gesamt-RNA bzw. poly(A)+RNA aus biologischen Proben, der Auftrennung der RNAs auf nach Größe auftrennenden Gelen, beispielsweise denaturierenden Agarosegelen, der Herstellung und Markierung der Sonde und des Nachweises über "Northern-Blot" angewandt werden.

In einer weiteren alternativen Ausführungsform kann eine mögliche Erkrankung auch durch ein Verfahren diagnostiziert werden, das folgende Schritte umfaßt:

- Gewinnung einer Zellprobe von dem Patienten,

- Inkontaktbringen der so erhaltenen Zellprobe mit dem vorstehend beschriebenen erfindungsgemäßen Antikörper als Sonde unter Bedingungen, die die Bindung des Antikörpers erlauben, wobei eine nichtstattgefundene oder schwache Bindung des Antikörpers auf eine mangelnde oder erniedrigte Expression von HSGT1 hinweist, was indikativ für eine Erkrankung ist.

5

10

Dieser Nachweis kann ebenfalls unter Anwendung von dem Fachmann bekannten Standardtechniken durchgeführt werden. Diesem sind auch Zellaufschlußverfahren bekannt, die die Isolierung des Proteins auf eine solche Weise erlauben, daß dieses mit dem Antikörper in Kontakt gebracht werden kann. Der Nachweis des gebundenen Antikörpers erfolgt vorzugsweise über Immunassays, beispielsweise Western-Blot, ELISA oder RIA.

15

Die erfindungsgemäßen Diagnoseverfahren können auch als pränatale Diagnoseverfahren gemäß dem Fachmann bekannten Techniken durchgeführt werden.

20

Schließlich betrifft die vorliegende Erfindung Kits zur Durchführung der erfindungsgemäßen Diagnoseverfahren, die den erfindungsgemäßen Antikörper oder ein Fragment davon, ein erfindungsgemäßes DNA-Molekül als Sonde oder ein für PCR oder LCR geeignetes, auf der Sequenz des erfindungsgemäßen DNA-Moleküls basierendes Primerpaar enthalten, gegebenenfalls in Kombination mit einem geeigneten Nachweismittel.

25

Je nach Ausgestaltung des mit den erfindungsgemäßen Kits durchzuführenden Diagnoseverfahrens können die in dem Kit enthaltenden DNA-Moleküle, Antikörper oder Fragmente davon an einem geeigneten Träger immobilisiert sein.

30

Die Erfindung wird nun weiter anhand der Figuren beschrieben, welche zeigen:

Fig.1: Nucleinsäuresequenz der cDNA des HSGT1-Gens

Der CAG-Trinukleotidrepeat ist fett dargestellt. Die Primer-Bin-

ungsstellen für die PCR, die einen Polymorphismus amplifizieren, sind unterstrichen.

Fig. 2: von Fig. 1 abgeleitete Aminosäuresequenz

Fig. 3: Northern Blot-Analyse, bei der RNAs aus verschiedenen Geweben verwendet wurde. Als Hybridierungsprobe wurde ein RT-PCR-Produkt aus fötaler Hirn-RNA vom 5'-Ende einschließlich der polymorphen Repeats eingesetzt.

Fig. 4: Genomkarte mit der genauen Lokalisation des Gens HSGT1

Die Erfindung wird nun nachfolgend mit Bezug auf die Beispiele beschrieben.

Beispiel 1

Northern Blots

Northern-Blots mit RNAs aus verschiedenen Geweben wurden bei 42 °C über Nacht in einem Hybridisierungspuffer mit 2xSSC, Denhardt-Lösung, Formamid und Heringssperma-DNA sowie der spezifischen Sonde (RT-PCR-Produkt aus fötaler Hirn-RNA aus dem 5'-Ende einschl. des polymorphen Repeats) hybridisiert. Die Filter wurden zweimal bis dreimal bei 65 °C in 2xSSC/0,1% SDS gewaschen und mit diesen bei -70 °C ein Kodak-Film oder Fuji-Film 24 bis 48 Stunden belichtet.

Das Ergebnis des Northern Blots ist in Fig. 3 gezeigt. Es ist sichtbar, daß das HSGT1-Gen ein 8 kB großes Transkript besitzt und ubiquitär schwach exprimiert wird.

Beispiel 2:

Herstellung und Reinigung eines erfindungsgemäßen HSGT1-Proteins

Zur Herstellung eines erfindungsgemäßen HSGT1-Proteins wird die DNA von Fig. 1 mit Bam HI-Linkern versehen, mit Bam HI nachgeschnitten und in den mit Bam

HI gespaltenen Expressionsvektor pQE-8 (Diagen) inseriert. Es wird das Expressionsplasmid pQ/HSGT1 erhalten. Ein solches kodiert für ein Fusionsprotein aus 6 Histidin-Resten (N-Terminuspartner) und dem erfindungsgemäßen HSGT1-Protein von Fig. 2 (C-Terminuspartner). pQ/HSGT1 wird zur Transformation von E.coli SG 13009 (vgl. Gottesmann, S. et al., J. Bacteriol. 148, (1981), 265-273) verwendet. Die Bakterien werden in einem LB-Medium mit 10 µg/ml Ampicillin und 25 µg/ml Kanamycin kultiviert und 4 h mit 60 µM Isopropyl-β-D-Thiogalactopyranosid (IPTG) induziert. Durch Zugabe von 6 M Guanidinhydrochlorid wird eine Lyse der Bakterien erreicht, anschließend wird mit dem Lysat eine Chromatographie (Ni-NTA-Resin) in Gegenwart von 8 M Harnstoff entsprechend der Angaben des Herstellers (Diagen) des Chromatographie-Materials durchgeführt. Das gebundene Fusionsprotein wird in einem Puffer mit pH 3,5 eluiert. Nach seiner Neutralisierung wird das Fusionsprotein einer 18 % SDS-Polyacrylamid-Gelelektrophorese unterworfen und mit Coomassie-Blau angefärbt (vgl. Thomas, J.O. und Kornberg, R.D., J.Mol.Biol. 149 (1975), 709-733).

Es zeigt sich, daß ein erfindungsgemäßes (Fusions)protein in hochreiner Form hergestellt werden kann.

Beispiel 3:

Herstellung und Nachweis eines erfindungsgemäßen Antikörpers

Ein erfindungsgemäßes Fusionsprotein von Beispiel 3 wird einer 18 % SDS-Polyacrylamid-Gelelektrophorese unterzogen. Nach Anfärbung des Gels mit 4 M Natriumacetat wird eine ca. 200 kD Bande aus dem Gel herausgeschnitten und in Phosphat gepufferter Kochsalzlösung inkubiert. Gel-Stücke werden sedimentiert, bevor die Proteinkonzentration des Überstandes durch eine SDS-Polyacrylamid-Gelelektrophorese, der eine Coomassie-Blau-Färbung folgt, bestimmt wird. Mit dem Gel-gereinigten Fusionsprotein werden Tiere wie folgt immunisiert:

Immunisierungsprotokoll für polyklonale Antikörper im Kaninchen

Pro Immunisierung werden 35 µg Gel-gereinigtes Fusionsprotein in 0,7 ml PBS und 0,7 ml komplettem bzw. inkomplettem Freund's Adjuvans eingesetzt.

5

- Tag 0: 1. Immunisierung (komplettes Freund's Adjuvans)
Tag 14: 2. Immunisierung (inkomplettes Freund's Adjuvans; icFA)
Tag 28: 3. Immunisierung (icFA)
Tag 56: 4. Immunisierung (icFA)
10 Tag 80: Ausbluten

15

Das Serum des Kaninchens wird im Immunoblot getestet. Hierzu wird ein erfindungsgemäßes Fusionsprotein von Beispiel 3 einer SDS-Polyacrylamid-Gelelektrophorese unterzogen und auf ein Nitrocellulosefilter übertragen (vgl. Khyse-Andersen), J., J.Biochem.Biophys. Meth. 10, (1984), 203-209). Die Western Blot-Analyse wird wie in Bock, C.-T. et al., Virus Genes 8, (1994), 215-229, beschrieben, durchgeführt. Hierzu wird das Nitrocellulosefilter 1 h bei 37°C mit einem ersten Antikörper inkubiert. Dieser Antikörper ist das Serum des Kaninchens (1:10000 in PBS). Nach mehreren Waschschritten mit PBS wird das Nitrocellulosefilter mit einem zweiten Antikörper inkubiert. Dieser Antikörper ist ein mit alkalischer Phosphatase gekoppelter monoklonaler Ziege Anti-Kaninchen-IgG-Antikörper (Dianova) (1:5000) in PBS. Nach 30-minütiger Inkubation bei 37°C folgen mehrere Waschschriffe mit PBS und anschließend die alkalische Phosphatase-Nachweisreaktion mit Entwicklerlösung (36 µm 5' Bromo-4-chloro-3-indolylphosphat, 400 µM Nitroblau-tetrazolium, 100 mM Tris-HCl, pH 9,5, 100 mM NaCl, 5 mM MgCl₂) bei Raumtemperatur, bis Banden sichtbar werden.

20

25

Es zeigt sich, daß erfindungsgemäße, polyklonale Antikörper hergestellt werden können.

30

Immunisierungsprotokoll für polyklonale Antikörper im Huhn

Pro Immunisierung werden 40 μ g Gel-gereinigtes Fusionsprotein in 0,8 ml PBS und 0,8 ml kompletten bzw. inkomplettem Freund's Adjuvans eingesetzt.

5

- Tag 0: 1. Immunisierung (komplettes Freund's Adjuvans)
Tag 28: 2. Immunisierung (inkomplettes Freund's Adjuvans; icFA)
Tag 50: 3. Immunisierung (icFA).

- 10 Aus Eigelb werden Antikörper extrahiert und im Western Blot getestet. Es werden erfindungsgemäße polyklonale Antikörper nachgewiesen.

Immunisierungsprotokoll für monoklonale Antikörper der Maus

- 15 Pro Immunisierung werden 12 μ g Gel-gereinigtes Fusionsprotein in 0,25 ml PBS und 0,25 ml komplettem bzw. inkomplettem Freund's Adjuvans eingesetzt; bei der 4. Immunisierung ist das Fusionsprotein in 0,5 ml (ohne Adjuvans) gelöst.

- Tag 0: 1. Immunisierung (komplettes Freund's Adjuvans)
20 Tag 28: 2. Immunisierung (inkomplettes Freund's Adjuvans; icFA)
Tag 56: 3. Immunisierung (icFA)
Tag 84: 4. Immunisierung (PBS)
Tag 87: Fusion

- 25 Überstände von Hybridomen werden im Western Blot getestet. Erfindungsgemäße, monoklonale Antikörper werden nachgewiesen.

Patentansprüche

5

1. Nukleinsäure, die für das Protein HSGT1 codiert, wobei die Nukleinsäure umfaßt:

10

- (a) die Nukleinsäure von Fig. 1 oder eine hiervon durch ein oder mehrere Basenpaare unterschiedliche Nukleinsäure bzw. ein Fragment davon,
- (b) eine mit der Nukleinsäure von (a) hybridisierende Nukleinsäure oder
- (c) eine mit der Nukleinsäure von (a) oder (b) über den degenerierten genetischen Code verwandte Nukleinsäure.

15

2. Nukleinsäure nach Anspruch 1, wobei sie eine cDNA oder genomische DNA ist.

3. Vektor, umfassend die Nukleinsäure nach Anspruch 1.

20

4. Vektor nach Anspruch 3, wobei die Nukleinsäure mit regulatorischen Elementen funktionell verknüpft ist, die die Expression in prokaryontischen oder eukaryontischen Wirtszellen erlauben.

25

5. Vektor nach Anspruch 3 oder 4, der ein RNA-Virus oder Retrovirus ist.

6. Transformante, enthaltend das Expressionsplasmid nach Anspruch 3.

7. Transformante nach Anspruch 6, die ein Bakterium, eine Hefe-, Insekten-, Tier- oder Säugerzelle ist.

30

8. Protein umfassend die Aminosäuresequenz von Fig. 2 oder ein hiervon durch ein oder mehrere Aminosäureaustausche unterschiedliches Protein

bzw. ein Fragment dieser Proteine.

- 5 9. Verfahren zur Herstellung des Proteins nach Anspruch 8, umfassend die Kultivierung der Transformante nach Anspruch 6 oder 7 unter Bedingungen, die die Expression des Proteins erlauben, und Gewinnung des Proteins aus der Kultur.
10. Antikörper, gerichtet gegen das Protein nach Anspruch 8.
- 10 11. Verwendung des Proteins nach Anspruch 8 als Reagens zur Diagnose und/oder Therapie von neurodegenerativen Erkrankungen, insbesondere des Smith Magenis Syndroms.
- 15 12. Verwendung der Nukleinsäure nach Anspruch 1 oder 2 als Reagens zur Diagnose und/oder Therapie von neurodegenerativen Erkrankungen, insbesondere des Smith Magenis Syndroms.
- 20 13. Verwendung des Antikörpers nach Anspruch 10 als Reagenz zur Diagnose und/oder Therapie von neurodegenerativen Erkrankungen, insbesondere des Smith Magenis Syndroms.

CDNA Sequenz des HSGT1 Gens

Der CAG-Trinukleotidrepeat ist fett dargestellt, die Primer-Bindungsstellen für die PCR, die den Polymorphismus amplifiziert, sind unterstrichen.

[illegible]

Fig. 1
ERSATZBLATT (REGEL 26)

2/6

GAAGGAGGAGGCAGGTGGGCTGCTGCAGTGCCCCGAGGTGGCCAAGGCTGACCGGTGGCTGGAGGACAGC
CGGCACTGCTGTTCCACCGCCGACTTCGGGGACCTCCCACTGCTGCCACCCACCAGCAGGAAGGAGGACC
TGGAAGCTGAGGAGGAGTACTCCTCCCTATGTGAGCTCCTGGGCAGCCCCGAGCAGAGGCCTGGCATGCA
GGACCCGCTGTCACCCAAGGCCCCACTCATCTGCACCAAGGAGGAGGTGGAGGAGGTGCTGGACTCCAAG
GCGGGCTGGGGCTCTCCGTGCCACCTCTCAGGGGAGTCCGTATCCTGCTGGGCCCTACAGTGGGCACCG
AGTCAAAGGTCCAGAGCTGGTTTGAGTCCTCTCTGTACACATGAAGCCAGGTGAAGAGGGGGCCTGATGG
GGAGCGAGCTCCAGGGGATTCCACCACCTCGGACGCCTCTCTGGCCCAGAAGCCCAACAAGCCTGCTGTG
CCCGAGGCGCCCATCGCAAAGAAAGAGCCTGTGCCACGGGGCAAAGCTTACGGAGCCGTGCGGTGCACC
GGGGGCTGCCCAGGCCCAGGACTCCCCATGCAGGGCACCAAGTGTGCCAAAGACCTCTTGCTCCCTGA
ATCCTGCACAGGGCCCCCCCCAGGGACAGATGGAAGGGGCTGGAGCCCCAGGCCGGGGGGCCTCGGAAGGG
CTCCCCAGGATGTGTACTCGTTCTCTCAGGGCCTGAGTGAGCCCCGCACGCCCGGACCCCCAGGCCTGA
CCACCACCCCTGCACCCCCAGACAACTGGGGGGCAAAGCAGCGAGCCGCCTTCAAGTCGGGCAAGCGGGT
GGGGAAGCCCTCACCCAAGGCTGCCTCCAGCCCCAGCAACCCGGCCGCCCTGCCTGTGGCCTCCGACAGC
AGCCCGATGGGCTCCAAGACCAAGGAGACAGACTCACCCAGCACGCCTGGCAAGGACCAGCGCTCCATGA
TCCTTCGGTCACGCACCAAAACCCAGGAGATCTTCCACTCCAAGCGGCGGAGGCCCTCTGAGGGCCGGCT
CCCCAACTGCCGTGCCACCAAGAAGCTCCTCGACAACAGCCACTTGCCCGCCACATTCAAGGTCTCCAGC
AGCCCCCAGAAGGAGGGCAGGGTGAGCCAGCGGGCAAAGGTCCCCAAACCTGGTGACGGCAGCAAGCTCT
CTGACCGGGCCCTCCATGCGCTCAAAAGGAAGTCGGCCTTCATGGCGCCGGTCCCCACCAAGAAGCGGAA
CCTGGTCTTGCGGAGCCGCAGCAGCAGCAGCAACGCCAGTGGCAATGGGGGAGATGGGAAGGAGGAG
AGGCCTGAGGGTTCCCCACCCCTCTTCAAGAGGATGTCTTCTCCCAAGAAAGCCAAGCCCACCAAGGGCA
ATGGCGAGCCTGCCACAAAGCTCCCACCCCCGGAGACCCCCGATGCCTGCCTCAAGCTCGCCTCTCGGGC
AGCCTTCAGGGGGCCATGAAGACCAAGGTGCTGCCACCCCGGAAGGGCCGGGGCCTGAAGCTGGAAGCC
ATCGTGAGAAGATCACCTCGCCCAGCCTCAAGAAGTTCGCATGTAAAGCGCCAGGGGCCTCTCCTGGTA
ATCCTCTGAGCCCATCCCTTTCCGACAAAGACCGTGGGCTCAAGGGTGCTGGGGGCAGCCCAGTGGGGGT
GGAAGAAGGCCCTGGTAAATGTGGGCACCGGGCAGAAGCTCCCAACTTCTGGGGCTGATCCGTTATGCAGA
AATCCAACCAACAGATCCTTAAAAGGCCAACTCATGAACACAGGGCCAGTGGCACACAAGGGGCCAGTGA
GGACAACTCTGGTGGAGGAGGCAAGAAGCCAAAGATGGAGGAGCTGGGCCTGGCCTCCAGCCCCCGGAG
GGCAGGCCCTGCCAGCCCCAGACAAGGGCACAGAAACAGCCAGGCCACCAACTACAGCAGCTATTCCA
AGCGGAAGCGCCTCACTCGGGGGCCGGGCCAAGAACACCACCTTTACCCCTGTAAGGGGGCTGCCAAGCG
ACGACGACAGCAGCAGGTGCTGCCCCTGGATCCCGCAGAGCCTGAAATCCGCCTCAAGTACATTTCTCT
TGCAAGCGGCTGAGGTGAGACAGCCGGACCCCCGCTTCTCACCTTCGTGCGGGTGGAGAAGCGAGACG
CGTTACCAACCATATGCACTGTTGTCAACTCCCCTGGAGATGCGCCCAAGCCCCACAGGAAGCCTTCCTC
CTCTGCCTCCTCTTCCTCATCCTCGTCTCTCTTGGATGCAGCCGGGGCCTCCCTGGCCACACTC
CCTGGAGGCTCCATCCTGCAGCCGCGGCCCTCCTTGCCCCTCTCCTCCACGATGCACTTGGGGCCTGTGG
TTTCCAAGGCCCTGAGTACCTCTTGCTTGTGCTGCTCTGCCAAAACCCGGCCAACTTCAAGGACCT
TGGGGACCTCTGTGGGCCCTACTACCCTGAACACTGCCTCCCCAAAAAGGCCAAAACCTCAAGGAGAAG
GTGCGGCCAGAAGGCACCTGTGAGGAGGCCTCGCTGCCGCTTGAGAGAACTCAAGGTCCCGAGTGTG
CAGCTGCCGCCACTGCCGGGAAGCCCCCAGGCCTGACGGCCAGCTGACCCGGCCAAGCAGGGCCCACT
GCGCACAGTGCCCGGGGCTGTCCCGGAGGCTGCAGAGCTGCTACTGCTGTGATGGCCGGGAGGATGGG
GGCGAGGAGGCAGCCCCAGCCGACAAGGCACTGGGTGCATGAGGCCTGTGCCGTGTGGACCGGGCGCTC

Fig. 1 (Forts.)

ERSATZBLATT (REGEL 26)

3/6

TACCTGGTGGCCGGGAAGCTCTTTGGGCTGCAGGAGGCCATGAAGGTGGCCGTGGACACACGGTGGCACT
GTGGCTCTGGCGCCTGGTGACTTCTCCCTCCCCGGCCTCAGGTTTGCTTCACTCTTCCAAGGGCCTTCTT
GGTGTGACTGTCCTGTCCTGGCCACGTCCACTCCAAGCAGCTGGTCCAGATGTGTTCCAGCTGCCAAGAA
GCGGGGGCCACCATTTGGGTGCTGCCACAAAGGATGCCTCCACACCTACCACTACCCGTGTGCCAGCGATG
C

Fig. 1 (Forts.)

4/6

MQSFRERCGFHGKQQNYQQT SQETSRL ENYRQPSQAGLSCDRQRLLAKDYNNPQPYPSYE
GGAGTPSGTAAAVAADKYHRGSKALPTQQGLQGRPAFPGYGVQDSSPYPGRYAGEESLQA
WGAPQPPPPQPQLPAGVAKYDENLMKKTAVPPSRQYAEQGAQVPFRTHSLHVQQPPPPQ
QPLAYPKLQRQKLQNDIASPLPFPGWWAGGHSYKSCTAPTAQPHDRPLTASSSLAPGQR
VQNLHAYQSGRLSYDQQQQQQQQQQQQQALQSRHHAQETLHYQNLAKYQHYGQQGQGYC
QPDAAVRTPEQYYQTFSPSSSHSPARSVGRSPSYSSSTPSPLMPNLENFPYSQQPLSTGAF
PAGITDHSFMPLLNPSPTDATSSPADLTALTSQVENISNTVQQLLSKAAVPQKKGVK
LVSRTP EQHKSQHCSPEGSGYSAEPAGTLPSEPPSSTPQSTHAEPQEADYLSGSEDPLER
SFLYCNQARGSPARVNSNSKAKPESVSTCSVTSPDDMSTKSDDSFQSLHGSLPLDSFSKF
VAGERDCPRLLLSALAQEDLASEILGLQEAI GEKADKA WAEAPSLVKDSSKPPFSLENHS
ACLDVAKSAWPRPGEALPDSLQLDKGGNAKDFSPGLFEDPSVAFATPDPKKTTGPLS
FGTKPTLGVPA PDPTTA AFD CFPDTTA ASSAD SANPFA WPEENLG DACPRWGLHPGELTK
GLEQGGKASDGISKGD THEASACLGFEEDPPGEKVASLPGDFKQEEVGGVKEEAGGLLQ
CPEVAKADRWLED SRHCCSTADFGDLPLLPPTSRKEDLEAEEYSSLCCELLGSPEQRPGM
QDPLSPKAPLIC TKEEVEEVLDSKAGW GSPCHLSGESVILLGPTVGTESKVQSWFESSLS
HMKPGEEGPDGERAPGDSTTSDASLAQKPNKPAVPEAPIAKKEPVPRGKSLRSRRVHRGL
PEAEDSPCRAPVL PKD LLLPESCTGPPQGM EGAGAPGRGASEGLPRMCTRSLTALSEPR
TPGPPGLTTTPAPDKLGGKQRAAFKSGKR VGKPSPKAASSPSNPAALPVASDSSPMGSK
TKETD SPSTPGKDQ RSMILRSRTKTQEIFH SKRRRPSEGRLPNCRA TKKLLDNSHL PATF
KVSSSPQKEGRVSQRARVPKPGAGSKLSDRPLHALKRKSAFMAPVPTKKRNLVLRSSSS
SSNASGN GGDGKEERPEGSPTLFKRMSSPKKAKPTKGNGEPATKLPPPETPDACLKLASR
AAFQGAMKTKVLP RKGRGLKLEAIVQKITSPSLKKFACKAPGASPGNPLSPSLSDKDRG
LKGAGGSPVGVEEGLVNVGTGQKLPTSGADPLCRNPTNRS LKGKLMNNRASGTQGASEDN
SGGGGKKPKMEELGLASQPPEGRPCQPQTRAQKQPGHTNYSSYSKRKRLTRGRAKNTTSS
PCKGRAKRRRQQQVLPLDPAEPEIRLKYISSCKRLRSDSRTPAFSPFVRVEKRDAFTTIC
TVVNSPGDAPKPHRKPSSSASSSSSSSSSFLDAAGASLATLPGGSILQPRPSLPLSSTMH
LGPVVSKALSTSLVCCLCQNPANFKDLGDLGCPYYPEHCLPKKKPKLKEKVRPEGTCEE
ASLPLERTLKGPECAAAATAGKPPRPDGPADPAKQGPLRTSARGLSRRLQSCYCCDGRED
GGEEAAPADKGHWVHEACAVWTGGVYLVAGKLFGLQEAMKVAVDMHGGTVALAPGDFSLP
GLRFASLFQGPSWCDPVLATSTPSSWSRCVPAAKKPGPPLGAATKDASTPTTTRVPAMQ

Fig. 2

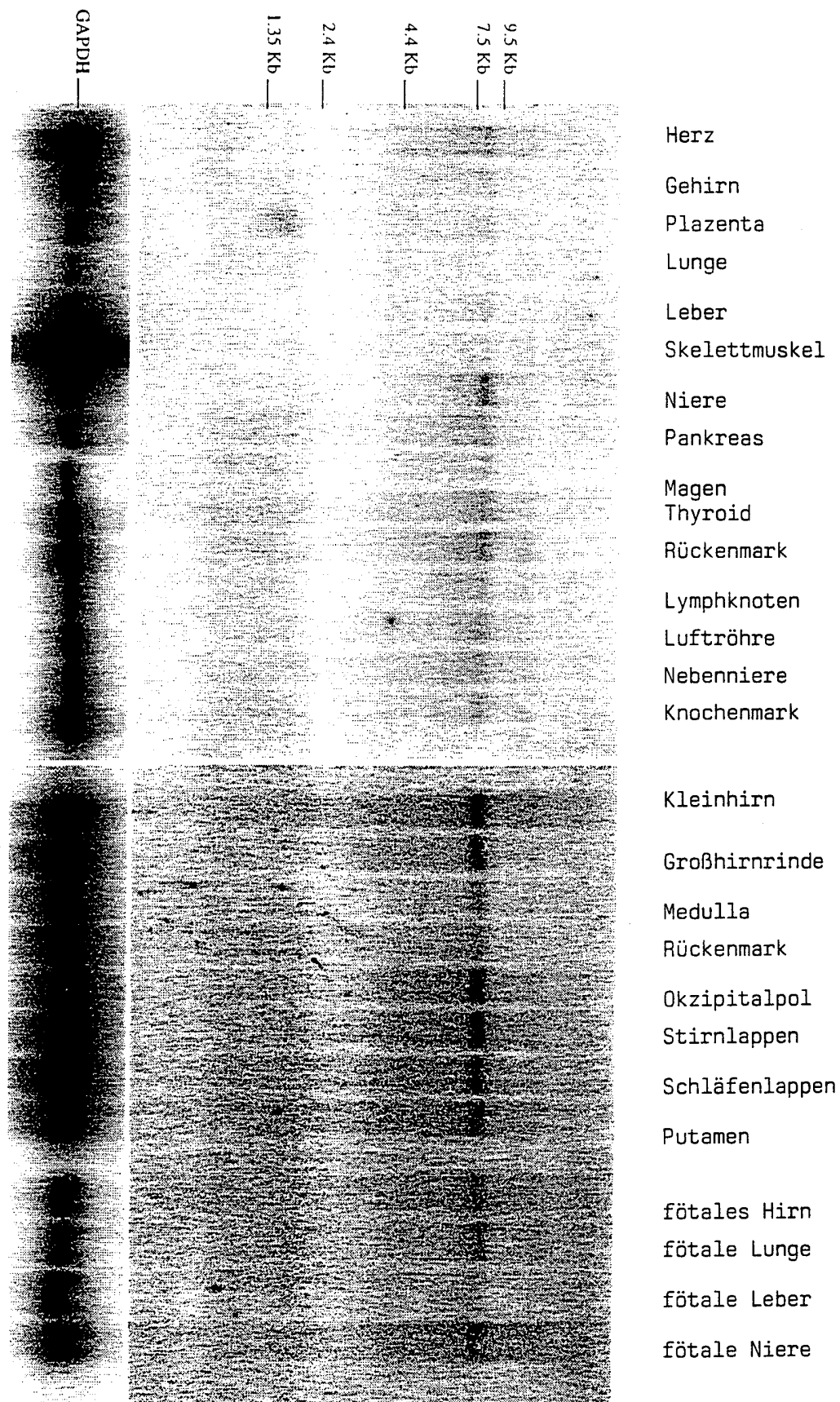


Fig. 3

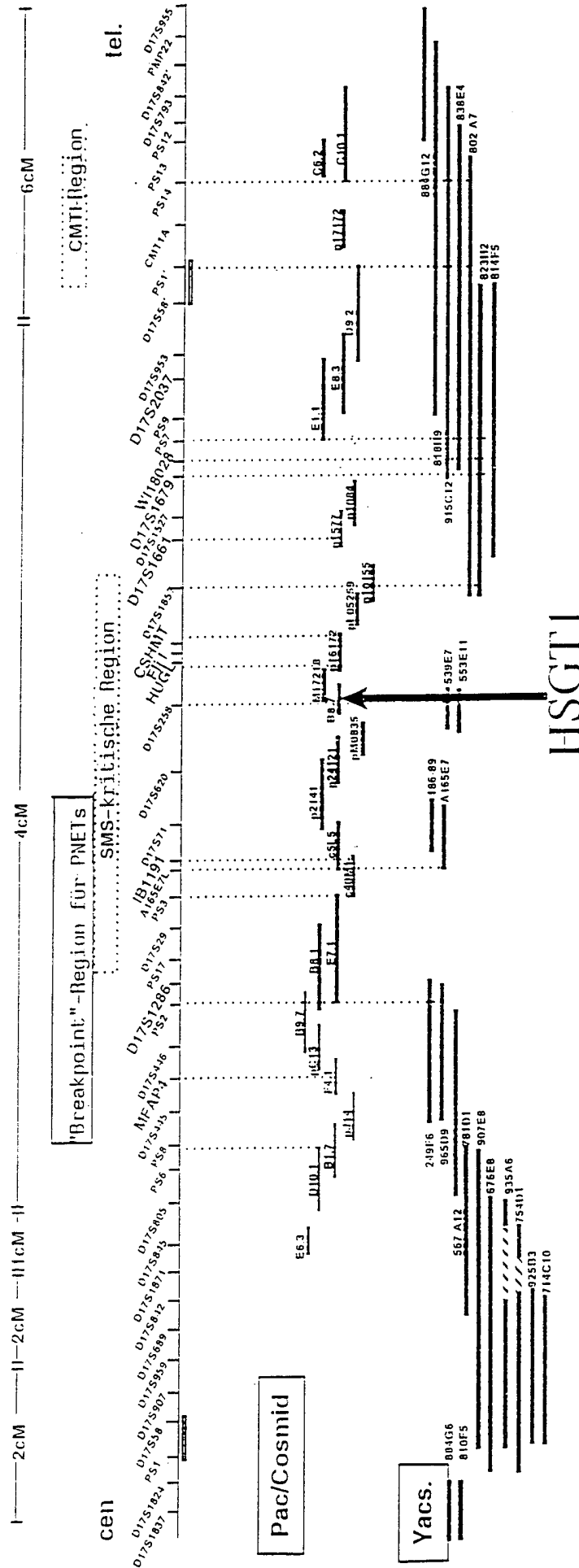


Fig. 4

SEQUENZPROTOKOLL

(1) ALLGEMEINE ANGABEN:

(i) ANMELDER:

- (A) NAME: Deutsches Krebsforschungszentrum
- (B) STRASSE: Im Neuenheimer Feld 280
- (C) ORT: Heidelberg
- (E) LAND: Deutschland
- (F) POSTLEITZAHL: 69120

(ii) BEZEICHNUNG DER ERFINDUNG: Gen isoliert auf dem kurzen Arm des menschlichen Chromosoms 17

(iii) ANZAHL DER SEQUENZEN: 2

(iv) COMPUTER-LESBARE FASSUNG:

- (A) DATENTRÄGER: Floppy disk
- (B) COMPUTER: IBM PC compatible
- (C) BETRIEBSSYSTEM: PC-DOS/MS-DOS
- (D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPA)

(v) DATEN DER JETZIGEN ANMELDUNG:
NOCH NICHT BEKANNT

(vi) DATEN DER VORANMELDUNG:
ANMELDENUMMER: DE 198 35 910.1
ANMELDETAG: 7-AUG-1998

(2) ANGABEN ZU SEQ ID NO: 1:

(i) SEQUENZKENNZEICHEN:

- (A) LÄNGE: 5395 Basenpaare
- (B) ART: Nucleotid
- (C) STRANGFORM: nicht bekannt
- (D) TOPOLOGIE: nicht bekannt

(ii) ART DES MOLEKÜLS: cDNA

(ix) MERKMAL:

- (A) NAME/SCHLÜSSEL: CDS
- (B) LAGE: 1..5394

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 1:

ATG CAG TCT TTT CGA GAA AGG TGT GGT TTC CAT GGC AAA CAA CAG AAC	48
Met Gln Ser Phe Arg Glu Arg Cys Gly Phe His Gly Lys Gln Gln Asn	
1 5 10 15	
TAC CAG CAG ACC TCG CAG GAA ACA TCA CGC CTA GAG AAT TAC AGG CAG	96
Tyr Gln Gln Thr Ser Gln Glu Thr Ser Arg Leu Glu Asn Tyr Arg Gln	
20 25 30	
CCG AGT CAG GCC GGG CTA AGC TGC GAC CGG CAG CGG CTG CTC GCC AAG	144
Pro Ser Gln Ala Gly Leu Ser Cys Asp Arg Gln Arg Leu Leu Ala Lys	
35 40 45	
GAC TAT TAT AAC CCG CAG CCT TAC CCG AGC TAT GAG GGT GGC GCT GGC	192
Asp Tyr Tyr Asn Pro Gln Pro Tyr Pro Ser Tyr Glu Gly Gly Ala Gly	
50 55 60	
ACG CCC TCT GGC ACT GCA GCC GCG GTG GCC GCC GAC AAG TAC CAC CGA	240
Thr Pro Ser Gly Thr Ala Ala Ala Val Ala Ala Asp Lys Tyr His Arg	
65 70 75 80	

GGC Gly	AGC Ser	AAG Lys	GCC Ala	CTG Leu 85	CCC Pro	ACA Thr	CAG Gln	CAA Gln	GGC Gly 90	CTG Leu	CAG Gln	GGG Gly	AGG Arg	CCG Pro 95	GCT Ala	288
TTC Phe	CCT Pro	GGC Gly	TAC Tyr 100	GGC Gly	GTC Val	CAG Gln	GAC Asp	AGC Ser 105	AGC Ser	CCC Pro	TAC Tyr	CCA Pro	GGC Gly 110	CGC Arg	TAT Tyr	336
GCT Ala	GGT Gly	GAG Glu 115	GAG Glu	AGC Ser	CTT Leu	CAG Gln	GCT Ala 120	TGG Trp	GGG Gly	GCC Ala	CCA Pro	CAG Gln 125	CCA Pro	CCA Pro	CCC Pro	384
CCA Pro	CAG Gln 130	CCG Pro	CAG Gln	CCA Pro	CTA Leu	CCT Pro 135	GCA Ala	GGG Gly	GTG Val	GCC Ala	AAG Lys 140	TAT Tyr	GAT Asp	GAG Glu	AAC Asn	432
TTG Leu 145	ATG Met	AAA Lys	AAG Lys	ACA Thr 150	GCA Ala	GTG Val	CCC Pro	CCC Pro	AGC Ser 155	AGG Arg	CAG Gln	TAT Tyr	GCA Ala	GAG Glu	CAG Gln 160	480
GGC Gly	GCC Ala	CAG Gln	GTG Val	CCC Pro 165	TTT Phe	CGG Arg	ACT Thr	CAC His	TCC Ser 170	CTG Leu	CAC His	GTC Val	CAG Gln	CAG Gln 175	CCA Pro	528
CCG Pro	CCG Pro	CCC Pro	CAG Gln 180	CAG Gln	CCC Pro	CTG Leu	GCA Ala	TAC Tyr 185	CCC Pro	AAG Lys	CTC Leu	CAA Gln	AGG Arg 190	CAG Gln	AAG Lys	576
CTG Leu	CAG Gln 195	AAC Asn	GAC Asp	ATT Ile	GCC Ala	TCC Ser 200	CCT Pro	CTG Leu	CCC Pro	TTC Phe	CCC Pro	CAG Gln 205	GGG Gly	TGG Trp	TGG Trp	624
GCA Ala 210	GGG Gly	GGC Gly	CAC His	TCC Ser	TAT Tyr	AAG Lys 215	AGT Ser	TGC Cys	ACA Thr	GCA Ala	CCG Pro 220	ACT Thr	GCC Ala	CAG Gln	CCC Pro	672
CAT His 225	GAC Asp	AGG Arg	CCG Pro	CTG Leu	ACT Thr 230	GCC Ala	AGC Ser	TCC Ser	AGC Ser	CTG Leu 235	GCC Ala	CCG Pro	GGG Gly	CAG Gln	CGG Arg 240	720
GTC Val	CAG Gln	AAT Asn	CTT Leu	CAT His 245	GCC Ala	TAC Tyr	CAG Gln	TCG Ser	GGC Gly 250	CGC Arg	CTC Leu	AGC Ser	TAT Tyr	GAC Asp 255	CAG Gln	768
CAG Gln	CAG Gln	CAG Gln	CAG Gln 260	CAG Gln	CAG Gln	CAG Gln	CAG Gln	CAG Gln 265	CAG Gln	CAG Gln	CAG Gln	CAA Gln	GCC Ala 270	CTT Leu	CAG Gln	816
AGC Ser	CGG Arg	CAC His 275	CAT His	GCC Ala	CAG Gln	GAA Glu	ACC Thr 280	CTC Leu	CAT His	TAC Tyr	CAA Gln	AAC Asn 285	CTC Leu	GCC Ala	AAG Lys	864
TAT Tyr 290	CAG Gln	CAC His	TAC Tyr	GGG Gly	CAG Gln	CAA Gln 295	GGC Gly	CAG Gln	GGC Gly	TAC Tyr	TGC Cys 300	CAG Gln	CCG Pro	GAC Asp	GCA Ala	912
GCC Ala 305	GTC Val	CGG Arg	ACC Thr	CCA Pro	GAG Glu 310	CAG Gln	TAC Tyr	TAC Tyr	CAG Gln 315	ACC Thr	TTC Phe	AGC Ser	CCC Pro	AGC Ser	TCC Ser 320	960
AGC Ser	CAC His	TCA Ser	CCC Pro	GCC Ala 325	CGC Arg	TCC Ser	GTG Val	GGC Gly	CGC Arg 330	TCA Ser	CCT Pro	TCC Ser	TAC Tyr	AGT Ser 335	TCC Ser	1008
ACA Thr	CCG Pro	TCG Ser	CCG Pro 340	CTG Leu	ATG Met	CCA Pro	AAC Asn	CTG Leu 345	GAG Glu	AAC Asn	TTT Phe	CCC Pro	TAC Tyr 350	AGC Ser	CAG Gln	1056

CAG Gln	CCG Pro	CTC Leu	AGC Ser	ACC Thr	GGG Gly	GCC Ala	TTC Phe	CCC Pro	GCA Ala	GGG Gly	ATC Ile	ACT Thr	GAC Asp	CAC His	AGC Ser	1104
		355					360					365				
CAC His	TTC Phe	ATG Met	CCC Pro	CTG Leu	CTC Leu	AAT Asn	CCC Pro	TCC Ser	CCA Pro	ACG Thr	GAT Asp	GCC Ala	ACC Thr	AGC Ser	TCT Ser	1152
	370					375					380					
CCT Pro	GCA Ala	GAC Asp	CTC Leu	ACG Thr	GCG Ala	CTG Leu	ACC Thr	TCA Ser	CAG Gln	GTG Val	GAG Glu	AAC Asn	ATC Ile	TCC Ser	AAC Asn	1200
	385				390					395					400	
ACC Thr	GTC Val	CAG Gln	CAG Gln	CTG Leu	CTG Leu	CTC Leu	TCC Ser	AAG Lys	GCT Ala	GCT Ala	GTG Val	CCG Pro	CAG Gln	AAG Lys	AAA Lys	1248
				405					410						415	
GGT Gly	GTC Val	AAG Lys	AAC Asn	CTC Leu	GTG Val	TCC Ser	AGG Arg	ACC Thr	CCA Pro	GAG Glu	CAG Gln	CAT His	AAA Lys	AGC Ser	CAG Gln	1296
			420				425						430			
CAC His	TGC Cys	AGC Ser	CCC Pro	GAA Glu	GGG Gly	AGC Ser	GGC Gly	TAC Tyr	TCA Ser	GCC Ala	GAG Glu	CCC Pro	GCA Ala	GGC Gly	ACA Thr	1344
		435					440					445				
CCG Pro	CTG Leu	TCA Ser	GAG Glu	CCG Pro	CCG Pro	AGC Ser	AGC Ser	ACG Thr	CCA Pro	CAG Gln	TCC Ser	ACG Thr	CAT His	GCG Ala	GAG Glu	1392
	450					455					460					
CCG Pro	CAG Gln	GAG Glu	GCC Ala	GAC Asp	TAC Tyr	CTG Leu	AGC Ser	GGC Gly	TCC Ser	GAG Glu	GAC Asp	CCA Pro	CTG Leu	GAG Glu	CGC Arg	1440
	465				470					475					480	
AGC Ser	TTC Phe	CTC Leu	TAC Tyr	TGC Cys	AAC Asn	CAG Gln	GCC Ala	CGT Arg	GGC Gly	AGC Ser	CCT Pro	GCC Ala	AGG Arg	GTC Val	AAC Asn	1488
				485					490					495		
AGC Ser	AAC Asn	TCG Ser	AAG Lys	GCC Ala	AAG Lys	CCC Pro	GAG Glu	TCC Ser	GTG Val	TCC Ser	ACC Thr	TGT Cys	TCT Ser	GTG Val	ACC Thr	1536
			500					505					510			
TCT Ser	CCT Pro	GAC Asp	GAC Asp	ATG Met	TCC Ser	ACC Thr	AAA Lys	TCT Ser	GAC Asp	GAC Asp	TCC Ser	TTC Phe	CAG Gln	AGC Ser	CTA Leu	1584
		515					520					525				
CAC His	GGC Gly	AGT Ser	CTG Leu	CCG Pro	CTC Leu	GAC Asp	AGC Ser	TTC Phe	TCC Ser	AAG Lys	TTC Phe	GTG Val	GCG Ala	GGT Gly	GAG Glu	1632
	530					535					540					
CGG Arg	GAC Asp	TGT Cys	CCG Pro	CGG Arg	CTG Leu	CTG Leu	CTC Leu	AGC Ser	GCC Ala	CTG Leu	GCA Ala	CAG Gln	GAG Glu	GAC Asp	CTG Leu	1680
	545				550					555					560	
GCC Ala	TCC Ser	GAG Glu	ATC Ile	CTG Leu	GGG Gly	CTG Leu	CAG Gln	GAA Glu	GCC Ala	ATC Ile	GGT Gly	GAG Glu	AAG Lys	GCC Ala	GAC Asp	1728
				565				570						575		
AAA Lys	GCT Ala	TGG Trp	GCT Ala	GAA Glu	GCA Ala	CCC Pro	AGC Ser	CTG Leu	GTC Val	AAG Lys	GAC Asp	AGC Ser	AGC Ser	AAG Lys	CCA Pro	1776
		580						585					590			
CCC Pro	TTC Phe	TCG Ser	CTG Leu	GAG Glu	AAC Asn	CAC His	AGC Ser	GCC Ala	TGC Cys	CTG Leu	GAC Asp	TCT Ser	GTG Val	GCC Ala	AAG Lys	1824
		595					600					605				
AGT Ser	GCG Ala	TGG Trp	CCC Pro	CGG Arg	CCT Pro	GGG Gly	GAG Glu	CCG Pro	GAG Glu	GCC Ala	CTG Leu	CCC Pro	GAC Asp	TCC Ser	TTG Leu	1872
	610					615					620					

CAG Gln 625	CTG Leu	GAC Asp	AAG Lys	GGC Gly	GGC Gly	AAT Asn	GCC Ala	AAG Lys	GAC Asp	TTC Phe	AGC Ser	CCA Pro	GGG Gly	CTG Leu	TTT Phe	1920
					630					635					640	
GAA Glu	GAC Asp	CCT Pro	TCC Ser	GTG Val	GCC Ala	TTC Phe	GCT Ala	ACG Thr	CCT Pro	GAC Asp	CCC Pro	AAA Lys	AAG Lys	ACA Thr	ACT Thr	1968
				645					650						655	
GGT Gly	CCT Pro	CTC Leu	TCC Ser	TTT Phe	GGT Gly	ACC Thr	AAG Lys	CCC Pro	ACC Thr	CTT Leu	GGG Gly	GTT Val	CCT Pro	GCT Ala	CCA Pro	2016
			660					665					670			
GAC Asp	CCC Pro	ACT Thr	ACA Thr	GCA Ala	GCT Ala	TTT Phe	GAC Asp	TGT Cys	TTC Phe	CCG Pro	GAC Asp	ACA Thr	ACC Thr	GCT Ala	GCC Ala	2064
		675					680					685				
AGC Ser	TCA Ser	GCG Ala	GAC Asp	AGC Ser	GCC Ala	AAC Asn	CCC Pro	TTT Phe	GCC Ala	TGG Trp	CCA Pro	GAG Glu	GAA Glu	AAC Asn	CTG Leu	2112
	690					695					700					
GGG Gly	GAT Asp	GCT Ala	TGT Cys	CCC Pro	AGG Arg	TGG Trp	GGA Gly	TTG Leu	CAC His	CCT Pro	GGC Gly	GAG Glu	CTT Leu	ACC Thr	AAG Lys	2160
705					710					715					720	
GGC Gly	CTG Leu	GAG Glu	CAG Gln	GGT Gly	GGG Gly	AAG Lys	GCC Ala	TCA Ser	GAT Asp	GGC Gly	ATC Ile	AGC Ser	AAA Lys	GGG Gly	GAC Asp	2208
			725						730					735		
ACC Thr	CAT His	GAG Glu	GCT Ala	TCG Ser	GCC Ala	TGC Cys	CTG Leu	GGC Gly	TTC Phe	CAG Gln	GAG Glu	GAG Glu	GAC Asp	CCC Pro	CCT Pro	2256
			740					745					750			
GGG Gly	GAG Glu	AAG Lys	GTG Val	GCC Ala	TCG Ser	TTG Leu	CCC Pro	GGG Gly	GAC Asp	TTC Phe	AAG Lys	CAG Gln	GAG Glu	GAG Glu	GTG Val	2304
		755					760					765				
GGT Gly	GGG Gly	GTG Val	AAG Lys	GAG Glu	GAG Glu	GCA Ala	GGT Gly	GGG Gly	CTG Leu	CTG Leu	CAG Gln	TGC Cys	CCC Pro	GAG Glu	GTG Val	2352
	770					775					780					
GCC Ala	AAG Lys	GCT Ala	GAC Asp	CGG Arg	TGG Trp	CTG Leu	GAG Glu	GAC Asp	AGC Ser	CGG Arg	CAC His	TGC Cys	TGT Cys	TCC Ser	ACC Thr	2400
785					790					795					800	
GCC Ala	GAC Asp	TTC Phe	GGG Gly	GAC Asp	CTC Leu	CCA Pro	CTG Leu	CTG Leu	CCA Pro	CCC Pro	ACC Thr	AGC Ser	AGG Arg	AAG Lys	GAG Glu	2448
			805						810					815		
GAC Asp	CTG Leu	GAA Glu	GCT Ala	GAG Glu	GAG Glu	GAG Glu	TAC Tyr	TCC Ser	TCC Ser	CTA Leu	TGT Cys	GAG Glu	CTC Leu	CTG Leu	GGC Gly	2496
		820					825					830				
AGC Ser	CCC Pro	GAG Glu	CAG Gln	AGG Arg	CCT Pro	GGC Gly	ATG Met	CAG Gln	GAC Asp	CCG Pro	CTG Leu	TCA Ser	CCC Pro	AAG Lys	GCC Ala	2544
		835					840					845				
CCA Pro	CTC Leu	ATC Ile	TGC Cys	ACC Thr	AAG Lys	GAG Glu	GAG Glu	GTG Val	GAG Glu	GAG Glu	GTG Val	CTG Leu	GAC Asp	TCC Ser	AAG Lys	2592
	850					855					860					
GCC Ala	GGC Gly	TGG Trp	GGC Gly	TCT Ser	CCG Pro	TGC Cys	CAC His	CTC Leu	TCA Ser	GGG Gly	GAG Glu	TCC Ser	GTC Val	ATC Ile	CTG Leu	2640
865					870					875					880	
CTG Leu	GGC Gly	CCT Pro	ACA Thr	GTG Val	GGC Gly	ACC Thr	GAG Glu	TCA Ser	AAG Lys	GTC Val	CAG Gln	AGC Ser	TGG Trp	TTT Phe	GAG Glu	2688
				885					890					895		

TCC Ser	TCT Ser	CTG Leu	TCA Ser 900	CAC His	ATG Met	AAG Lys	CCA Pro	GGT Gly 905	GAA Glu	GAG Glu	GGG Gly	CCT Pro	GAT Asp 910	GGG Gly	GAG Glu	2736
CGA Arg	GCT Ala	CCA Pro 915	GGG Gly	GAT Asp	TCC Ser	ACC Thr 920	ACC Thr	TCG Ser	GAC Asp	GCC Ala	TCT Ser	CTG Leu 925	GCC Ala	CAG Gln	AAG Lys	2784
CCC Pro 930	AAC Asn	AAG Lys	CCT Pro	GCT Ala	GTG Val	CCC Pro 935	GAG Glu	GCG Ala	CCC Pro	ATC Ile	GCA Ala 940	AAG Lys	AAA Lys	GAG Glu	CCT Pro	2832
GTG Val 945	CCA Pro	CGG Arg	GGC Gly	AAA Lys	AGC Ser 950	TTA Leu	CGG Arg	AGC Ser	CGT Arg	CGG Arg 955	GTG Val	CAC His	CGG Arg	GGG Gly	CTG Leu 960	2880
CCC Pro	GAG Glu	GCC Ala	GAG Glu	GAC Asp 965	TCC Ser	CCA Pro	TGC Cys	AGG Arg	GCA Ala 970	CCA Pro	GTG Val	CTG Leu	CCC Pro	AAA Lys 975	GAC Asp	2928
CTC Leu	TTG Leu	CTC Leu	CCT Pro 980	GAA Glu	TCC Ser	TGC Cys	ACA Thr	GGG Gly 985	CCC Pro	CCC Pro	CAG Gln	GGA Gly 990	CAG Gln	ATG Met	GAA Glu	2976
GGG Gly	GCT Ala	GGA Gly 995	GCC Ala	CCA Pro	GGC Gly	CGG Arg	GGG Gly 1000	GCC Ala	TCG Ser	GAA Glu	GGG Gly	CTC Leu 1005	CCC Pro	AGG Arg	ATG Met	3024
TGT Cys	ACT Thr 1010	CGT Arg	TCT Ser	CTC Leu	ACG Thr	GCC Ala 1015	CTG Leu	AGT Ser	GAG Glu	CCC Pro	CGC Arg 1020	ACG Thr	CCC Pro	GGA Gly	CCC Pro	3072
CCA Pro 1025	GGC Gly	CTG Leu	ACC Thr	ACC Thr	ACC Thr 1030	CCT Pro	GCA Ala	CCC Pro	CCA Pro	GAC Asp 1035	AAA Lys	CTG Leu	GGG Gly	GGC Gly	AAG Lys 1040	3120
CAG Gln	CGA Arg	GCC Ala	GCC Ala	TTC Phe 1045	AAG Lys	TCG Ser	GGC Gly	AAG Lys	CGG Arg 1050	GTG Val	GGG Gly	AAG Lys	CCC Pro	TCA Ser 1055	CCC Pro	3168
AAG Lys	GCT Ala	GCC Ala	TCC Ser 1060	AGC Ser	CCC Pro	AGC Ser	AAC Asn	CCG Pro 1065	GCC Ala	GCC Ala	CTG Leu	CCT Pro	GTG Val 1070	GCC Ala	TCC Ser	3216
GAC Asp	AGC Ser	AGC Ser	CCG Pro 1075	ATG Met	GGC Gly	TCC Ser	AAG Lys 1080	ACC Thr	AAG Lys	GAG Glu	ACA Thr	GAC Asp 1085	TCA Ser	CCC Pro	AGC Ser	3264
ACG Thr 1090	CCT Pro	GGC Gly	AAG Lys	GAC Asp	CAG Gln	CGC Arg 1095	TCC Ser	ATG Met	ATC Ile	CTT Leu	CGG Arg 1100	TCA Ser	CGC Arg	ACC Thr	AAA Lys	3312
ACC Thr 1105	CAG Gln	GAG Glu	ATC Ile	TTC Phe	CAC His 1110	TCC Ser	AAG Lys	CGG Arg	CGG Arg	AGG Arg 1115	CCC Pro	TCT Ser	GAG Glu	GGC Gly	CGG Arg 1120	3360
CTC Leu	CCC Pro	AAC Asn	TGC Cys 1125	CGT Arg	GCC Ala	ACC Thr	AAG Lys	AAG Lys	CTC Leu 1130	CTC Leu	GAC Asp	AAC Asn	AGC Ser	CAC His 1135	TTG Leu	3408
CCC Pro	GCC Ala	ACA Thr	TTC Phe 1140	AAG Lys	GTC Val	TCC Ser	AGC Ser	AGC Ser	CCC Pro 1145	CAG Gln	AAG Lys	GAG Glu	GGC Gly 1150	AGG Arg	GTG Val	3456
AGC Ser	CAG Gln	CGG Arg	GCA Ala	AGG Arg	GTC Val	CCC Pro	AAA Lys 1160	CCT Pro	GGT Gly	GCA Ala	GGC Gly	AGC Ser 1165	AAG Lys	CTC Leu	TCT Ser	3504

GAC Asp 1170	CGG Arg 1170	CCC Pro 1170	CTC Leu 1170	CAT His 1170	GCG Ala 1175	CTC Leu 1175	AAA Lys 1175	AGG Arg 1175	AAG Lys 1175	TCG Ser 1180	GCC Ala 1180	TTC Phe 1180	ATG Met 1180	GCG Ala 1180	CCG Pro 1180	3552
GTC Val 1185	CCC Pro 1185	ACC Thr 1185	AAG Lys 1185	AAG Lys 1185	CGG Arg 1190	AAC Asn 1190	CTG Leu 1190	GTC Val 1190	TTG Leu 1195	CGG Arg 1195	AGC Ser 1195	CGC Arg 1195	AGC Ser 1200	AGC Ser 1200	AGC Ser 1200	3600
AGC Ser 1205	AGC Ser 1205	AAC Asn 1205	GCC Ala 1205	AGT Ser 1205	GGC Gly 1205	AAT Asn 1205	GGG Gly 1210	GGA Gly 1210	GAT Asp 1210	GGG Gly 1210	AAG Lys 1215	GAG Glu 1215	GAG Glu 1215	AGG Arg 1215	CCT Pro 1215	3648
GAG Glu 1220	GGT Gly 1220	TCC Ser 1220	CCC Pro 1220	ACC Thr 1220	CTC Leu 1225	TTC Phe 1225	AAG Lys 1225	AGG Arg 1225	ATG Met 1225	TCT Ser 1230	TCT Ser 1230	CCC Pro 1230	AAG Lys 1230	AAA Lys 1230	GCC Ala 1230	3696
AAG Lys 1235	CCC Pro 1235	ACC Thr 1235	AAG Lys 1235	GGC Gly 1235	AAT Asn 1240	GGC Gly 1240	GAG Glu 1240	CCT Pro 1240	GCC Ala 1245	ACA Thr 1245	AAG Lys 1245	CTC Leu 1245	CCA Pro 1245	CCC Pro 1245	CCG Pro 1245	3744
GAG Glu 1250	ACC Thr 1250	CCC Pro 1250	GAT Asp 1255	GCC Ala 1255	TGC Cys 1255	CTC Leu 1255	AAG Lys 1255	CTC Leu 1255	GCC Ala 1260	TCT Ser 1260	CGG Arg 1260	GCA Ala 1260	GCC Ala 1260	TTC Phe 1260	CAG Gln 1260	3792
GGG Gly 1265	GCC Ala 1265	ATG Met 1265	AAG Lys 1270	ACC Thr 1270	AAG Lys 1270	GTG Val 1270	CTG Leu 1275	CCA Pro 1275	CCC Pro 1275	CGG Arg 1275	AAG Lys 1275	GGC Gly 1280	CGG Arg 1280	GGC Gly 1280	CTG Leu 1280	3840
AAG Lys 1285	CTG Leu 1285	GAA Glu 1285	GCC Ala 1285	ATC Ile 1285	GTG Val 1290	CAG Gln 1290	AAG Lys 1290	ATC Ile 1290	ACC Thr 1290	TCG Ser 1295	CCC Pro 1295	AGC Ser 1295	CTC Leu 1295	AAG Lys 1295	AAG Lys 1295	3888
TTC Phe 1300	GCA Ala 1300	TGT Cys 1300	AAA Lys 1305	GCG Ala 1305	CCA Pro 1305	GGG Gly 1305	GCC Ala 1305	TCT Ser 1310	CCT Pro 1310	GGT Gly 1310	AAT Asn 1310	CCT Pro 1310	CTG Leu 1310	AGC Ser 1310	CCA Pro 1310	3936
TCC Ser 1315	CTT Leu 1315	TCC Ser 1315	GAC Asp 1320	AAA Lys 1320	GAC Asp 1320	CGT Arg 1320	GGG Gly 1320	CTC Leu 1325	AAG Lys 1325	GGT Gly 1325	GCT Ala 1325	GGG Gly 1325	GGC Gly 1325	AGC Ser 1325	CCA Pro 1325	3984
GTG Val 1330	GGG Gly 1330	GTG Val 1330	GAA Glu 1335	GAA Glu 1335	GGC Gly 1335	CTG Leu 1335	GTA Val 1335	AAT Asn 1340	GTG Val 1340	GGC Gly 1340	ACC Thr 1340	GGG Gly 1340	CAG Gln 1340	AAG Lys 1340	CTC Leu 1340	4032
CCA Pro 1345	ACT Thr 1345	TCT Ser 1345	GGG Gly 1350	GCT Ala 1350	GAT Asp 1350	CCG Pro 1350	TTA Leu 1355	TGC Cys 1355	AGA Arg 1355	AAT Asn 1355	CCA Pro 1355	ACC Thr 1360	AAC Asn 1360	AGA Arg 1360	TCC Ser 1360	4080
TTA Leu 1365	AAA Lys 1365	GGC Gly 1365	AAA Lys 1365	CTC Leu 1365	ATG Met 1370	AAC Asn 1370	AGG Arg 1370	GCC Ala 1370	AGT Ser 1375	GGC Gly 1375	ACA Thr 1375	CAA Gln 1375	GGG Gly 1375	GCC Ala 1375	AGT Ser 1375	4128
GAG Glu 1380	GAC Asp 1380	AAC Asn 1380	TCT Ser 1380	GGT Gly 1385	GGA Gly 1385	GGA Gly 1385	GGC Gly 1385	AAG Lys 1385	AAG Lys 1385	CCA Pro 1390	AAG Lys 1390	ATG Met 1390	GAG Glu 1390	GAG Glu 1390	CTG Leu 1390	4176
GGC Gly 1395	CTG Leu 1395	GCC Ala 1395	TCC Ser 1400	CAG Gln 1400	CCC Pro 1400	CCG Pro 1400	GAG Gly 1400	GGC Gly 1405	AGG Arg 1405	CCC Pro 1405	TGC Cys 1405	CAG Gln 1405	CCC Pro 1405	CAG Gln 1405	ACA Thr 1405	4224
AGG Arg 1410	GCA Ala 1410	CAG Gln 1410	AAA Lys 1415	CAG Gln 1415	CCA Pro 1415	GGC Gly 1415	CAC His 1415	ACC Thr 1420	AAC Asn 1420	TAC Tyr 1420	AGC Ser 1420	AGC Ser 1420	TAT Tyr 1420	TCC Ser 1420	AAG Lys 1420	4272
CGG Arg 1425	AAG Lys 1425	CGC Arg 1430	CTC Leu 1430	ACT Thr 1430	CGG Arg 1430	GGC Gly 1430	CGG Arg 1435	GCC Ala 1435	AAG Lys 1435	AAC Asn 1435	ACC Thr 1435	ACC Thr 1435	TCT Ser 1440	TCA Ser 1440	CCC Pro 1440	4320

TGT Cys	AAG Lys	GGG Gly	CGT Arg	GCC Ala	AAG Lys	CGA Arg	CGA Arg	CGA Arg	CAG Gln	CAG Gln	CAG Gln	GTG Val	CTG Leu	CCC Pro	CTG Leu	4368
				1445					1450					1455		
GAT Asp	CCC Pro	GCA Ala	GAG Glu	CCT Pro	GAA Glu	ATC Ile	CGC Arg	CTC Leu	AAG Lys	TAC Tyr	ATT Ile	TCC Ser	TCT Ser	TGC Cys	AAG Lys	4416
			1460					1465					1470			
CGG Arg	CTG Leu	AGG Arg	TCA Ser	GAC Asp	AGC Ser	CGG Arg	ACC Thr	CCC Pro	GCC Ala	TTC Phe	TCA Ser	CCC Pro	TTC Phe	GTG Val	CGG Arg	4464
		1475					1480					1485				
GTG Val	GAG Glu	AAG Lys	CGA Arg	GAC Asp	GCG Ala	TTC Phe	ACC Thr	ACC Thr	ATA Ile	TGC Cys	ACT Thr	GTT Val	GTC Val	AAC Asn	TCC Ser	4512
	1490					1495					1500					
CCT Pro	GGA Gly	GAT Asp	GCG Ala	CCC Pro	AAG Lys	CCC Pro	CAC His	AGG Arg	AAG Lys	CCT Pro	TCC Ser	TCC Ser	TCT Ser	GCC Ala	TCC Ser	4560
	1505				1510					1515					1520	
TCT Ser	TCC Ser	TCA Ser	TCC Ser	TCG Ser	TCC Ser	TCG Ser	TTC Phe	TCC Ser	TTG Leu	GAT Asp	GCA Ala	GCC Ala	GGG Gly	GCC Ala	TCC Ser	4608
				1525					1530					1535		
CTG Leu	GCC Ala	ACA Thr	CTC Leu	CCT Pro	GGA Gly	GGC Gly	TCC Ser	ATC Ile	CTG Leu	CAG Gln	CCG Pro	CGG Arg	CCC Pro	TCC Ser	TTG Leu	4656
			1540					1545					1550			
CCC Pro	CTC Leu	TCC Ser	TCC Ser	ACG Thr	ATG Met	CAC His	TTG Leu	GGG Gly	CCT Pro	GTG Val	GTT Val	TCC Ser	AAG Lys	GCC Ala	CTG Leu	4704
		1555					1560					1565				
AGT Ser	ACC Thr	TCT Ser	TGC Cys	CTT Leu	GTT Val	TGC Cys	TGC Cys	CTC Leu	TGC Cys	CAA Gln	AAC Asn	CCG Pro	GCC Ala	AAC Asn	TTC Phe	4752
	1570					1575				1580						
AAG Lys	GAC Asp	CTT Leu	GGG Gly	GAC Asp	CTC Leu	TGT Cys	GGG Gly	CCC Pro	TAC Tyr	TAC Tyr	CCT Pro	GAA Glu	CAC His	TGC Cys	CTC Leu	4800
	1585				1590				1595						1600	
CCC Pro	AAA Lys	AAG Lys	AAG Lys	CCA Pro	AAA Lys	CTC Leu	AAG Lys	GAG Glu	AAG Lys	GTG Val	CGG Arg	CCA Pro	GAA Glu	GGC Gly	ACC Thr	4848
				1605					1610					1615		
TGT Cys	GAG Glu	GAG Glu	GCC Ala	TCG Ser	CTG Leu	CCG Pro	CTT Leu	GAG Glu	AGA Arg	ACA Thr	CTC Leu	AAA Lys	GGT Gly	CCC Pro	GAG Glu	4896
			1620					1625					1630			
TGT Cys	GCA Ala	GCT Ala	GCC Ala	GCC Ala	ACT Thr	GCC Ala	GGG Gly	AAG Lys	CCC Pro	CCC Pro	AGG Arg	CCT Pro	GAC Asp	GGC Gly	CCA Pro	4944
		1635					1640					1645				
GCT Ala	GAC Asp	CCG Pro	GCC Ala	AAG Lys	CAG Gln	GGC Gly	CCA Pro	CTG Leu	CGC Arg	ACC Thr	AGT Ser	GCC Ala	CGG Arg	GGC Gly	CTG Leu	4992
	1650					1655					1660					
TCC Ser	CGG Arg	AGG Arg	CTG Leu	CAG Gln	AGC Ser	TGC Cys	TAC Tyr	TGC Cys	TGT Cys	GAT Asp	GGC Gly	CGG Arg	GAG Glu	GAT Asp	GGG Gly	5040
	1665				1670					1675					1680	
GGC Gly	GAG Glu	GAG Glu	GCA Ala	GCC Ala	CCA Pro	GCC Ala	GAC Asp	AAG Lys	GGA Gly	CAC His	TGG Trp	GTG Val	CAT His	GAG Glu	GCC Ala	5088
			1685					1690						1695		
TGT Cys	GCC Ala	GTG Val	TGG Trp	ACC Thr	GGC Gly	GGC Gly	GTC Val	TAC Tyr	CTG Leu	GTG Val	GCC Ala	GGG Gly	AAG Lys	CTC Leu	TTT Phe	5136
		1700						1705					1710			

GGG CTG CAG GAG GCC ATG AAG GTG GCC GTG GAC ATG CAC GGT GGC ACT Gly Leu Gln Glu Ala Met Lys Val Ala Val Asp Met His Gly Gly Thr 1715 1720 1725	5184
GTG GCT CTG GCG CCT GGT GAC TTC TCC CTC CCC GGC CTC AGG TTT GCT Val Ala Leu Ala Pro Gly Asp Phe Ser Leu Pro Gly Leu Arg Phe Ala 1730 1735 1740	5232
TCA CTC TTC CAA GGG CCT TCT TGG TGT GAC TGT CCT GTC CTG GCC ACG Ser Leu Phe Gln Gly Pro Ser Trp Cys Asp Cys Pro Val Leu Ala Thr 1745 1750 1755 1760	5280
TCC ACT CCA AGC AGC TGG TCC AGG TGT GTT CCA GCT GCC AAG AAG CCG Ser Thr Pro Ser Ser Trp Ser Arg Cys Val Pro Ala Ala Lys Lys Pro 1765 1770 1775	5328
GGG CCA CCA TTG GGT GCT GCC ACA AAG GAT GCC TCC ACA CCT ACC ACT Gly Pro Pro Leu Gly Ala Ala Thr Lys Asp Ala Ser Thr Pro Thr Thr 1780 1785 1790	5376
ACC CGT GTG CCA GCG ATG C Thr Arg Val Pro Ala Met 1795	5395

(2) ANGABEN ZU SEQ ID NO: 2:

(i) SEQUENZKENNZEICHEN:

- (A) LÄNGE: 1798 Aminosäuren
- (B) ART: Aminosäure
- (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: Protein

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 2:

Met 1	Gln	Ser	Phe	Arg 5	Glu	Arg	Cys	Gly	Phe 10	His	Gly	Lys	Gln	Gln	Asn 15
Tyr	Gln	Gln	Thr 20	Ser	Gln	Glu	Thr	Ser 25	Arg	Leu	Glu	Asn	Tyr 30	Arg	Gln
Pro	Ser	Gln	Ala	Gly	Leu	Ser	Cys 40	Asp	Arg	Gln	Arg	Leu 45	Leu	Ala	Lys
Asp	Tyr 50	Tyr	Asn	Pro	Gln	Pro	Tyr 55	Pro	Ser	Tyr	Glu 60	Gly	Gly	Ala	Gly
Thr 65	Pro	Ser	Gly	Thr	Ala 70	Ala	Ala	Val	Ala 75	Ala	Asp	Lys	Tyr	His	Arg 80
Gly	Ser	Lys	Ala	Leu 85	Pro	Thr	Gln	Gln	Gly 90	Leu	Gln	Gly	Arg	Pro	Ala 95
Phe	Pro	Gly	Tyr 100	Gly	Val	Gln	Asp	Ser 105	Ser	Pro	Tyr	Pro	Gly 110	Arg	Tyr
Ala	Gly	Glu	Glu	Ser	Leu	Gln	Ala 120	Trp	Gly	Ala	Pro	Gln	Pro	Pro	Pro 125
Pro	Gln	Pro	Gln	Pro	Leu	Pro	Ala 135	Gly	Val	Ala	Lys	Tyr	Asp	Glu	Asn 140
Leu 145	Met	Lys	Lys	Thr	Ala 150	Val	Pro	Pro	Ser	Arg 155	Gln	Tyr	Ala	Glu	Gln 160

Gly Ala Gln Val Pro Phe Arg Thr His Ser Leu His Val Gln Gln Pro
 165 170 175
 Pro Pro Pro Gln Gln Pro Leu Ala Tyr Pro Lys Leu Gln Arg Gln Lys
 180 185 190
 Leu Gln Asn Asp Ile Ala Ser Pro Leu Pro Phe Pro Gln Gly Trp Trp
 195 200 205
 Ala Gly Gly His Ser Tyr Lys Ser Cys Thr Ala Pro Thr Ala Gln Pro
 210 215 220
 His Asp Arg Pro Leu Thr Ala Ser Ser Ser Leu Ala Pro Gly Gln Arg
 225 230 235 240
 Val Gln Asn Leu His Ala Tyr Gln Ser Gly Arg Leu Ser Tyr Asp Gln
 245 250 255
 Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Ala Leu Gln
 260 265 270
 Ser Arg His His Ala Gln Glu Thr Leu His Tyr Gln Asn Leu Ala Lys
 275 280 285
 Tyr Gln His Tyr Gly Gln Gln Gly Gln Gly Tyr Cys Gln Pro Asp Ala
 290 295 300
 Ala Val Arg Thr Pro Glu Gln Tyr Tyr Gln Thr Phe Ser Pro Ser Ser
 305 310 315 320
 Ser His Ser Pro Ala Arg Ser Val Gly Arg Ser Pro Ser Tyr Ser Ser
 325 330 335
 Thr Pro Ser Pro Leu Met Pro Asn Leu Glu Asn Phe Pro Tyr Ser Gln
 340 345 350
 Gln Pro Leu Ser Thr Gly Ala Phe Pro Ala Gly Ile Thr Asp His Ser
 355 360 365
 His Phe Met Pro Leu Leu Asn Pro Ser Pro Thr Asp Ala Thr Ser Ser
 370 375 380
 Pro Ala Asp Leu Thr Ala Leu Thr Ser Gln Val Glu Asn Ile Ser Asn
 385 390 395 400
 Thr Val Gln Gln Leu Leu Leu Ser Lys Ala Ala Val Pro Gln Lys Lys
 405 410 415
 Gly Val Lys Asn Leu Val Ser Arg Thr Pro Glu Gln His Lys Ser Gln
 420 425 430
 His Cys Ser Pro Glu Gly Ser Gly Tyr Ser Ala Glu Pro Ala Gly Thr
 435 440 445
 Pro Leu Ser Glu Pro Pro Ser Ser Thr Pro Gln Ser Thr His Ala Glu
 450 455 460
 Pro Gln Glu Ala Asp Tyr Leu Ser Gly Ser Glu Asp Pro Leu Glu Arg
 465 470 475 480
 Ser Phe Leu Tyr Cys Asn Gln Ala Arg Gly Ser Pro Ala Arg Val Asn
 485 490 495
 Ser Asn Ser Lys Ala Lys Pro Glu Ser Val Ser Thr Cys Ser Val Thr
 500 505 510
 Ser Pro Asp Asp Met Ser Thr Lys Ser Asp Asp Ser Phe Gln Ser Leu
 515 520 525

His Gly Ser Leu Pro Leu Asp Ser Phe Ser Lys Phe Val Ala Gly Glu
 530 535 540
 Arg Asp Cys Pro Arg Leu Leu Ser Ala Leu Ala Gln Glu Asp Leu
 545 550 555 560
 Ala Ser Glu Ile Leu Gly Leu Gln Glu Ala Ile Gly Glu Lys Ala Asp
 565 570 575
 Lys Ala Trp Ala Glu Ala Pro Ser Leu Val Lys Asp Ser Ser Lys Pro
 580 585 590
 Pro Phe Ser Leu Glu Asn His Ser Ala Cys Leu Asp Ser Val Ala Lys
 595 600 605
 Ser Ala Trp Pro Arg Pro Gly Glu Pro Glu Ala Leu Pro Asp Ser Leu
 610 615 620
 Gln Leu Asp Lys Gly Gly Asn Ala Lys Asp Phe Ser Pro Gly Leu Phe
 625 630 635 640
 Glu Asp Pro Ser Val Ala Phe Ala Thr Pro Asp Pro Lys Lys Thr Thr
 645 650 655
 Gly Pro Leu Ser Phe Gly Thr Lys Pro Thr Leu Gly Val Pro Ala Pro
 660 665 670
 Asp Pro Thr Thr Ala Ala Phe Asp Cys Phe Pro Asp Thr Thr Ala Ala
 675 680 685
 Ser Ser Ala Asp Ser Ala Asn Pro Phe Ala Trp Pro Glu Glu Asn Leu
 690 695 700
 Gly Asp Ala Cys Pro Arg Trp Gly Leu His Pro Gly Glu Leu Thr Lys
 705 710 715 720
 Gly Leu Glu Gln Gly Gly Lys Ala Ser Asp Gly Ile Ser Lys Gly Asp
 725 730 735
 Thr His Glu Ala Ser Ala Cys Leu Gly Phe Gln Glu Glu Asp Pro Pro
 740 745 750
 Gly Glu Lys Val Ala Ser Leu Pro Gly Asp Phe Lys Gln Glu Glu Val
 755 760 765
 Gly Gly Val Lys Glu Glu Ala Gly Gly Leu Leu Gln Cys Pro Glu Val
 770 775 780
 Ala Lys Ala Asp Arg Trp Leu Glu Asp Ser Arg His Cys Cys Ser Thr
 785 790 795 800
 Ala Asp Phe Gly Asp Leu Pro Leu Leu Pro Pro Thr Ser Arg Lys Glu
 805 810 815
 Asp Leu Glu Ala Glu Glu Glu Tyr Ser Ser Leu Cys Glu Leu Leu Gly
 820 825 830
 Ser Pro Glu Gln Arg Pro Gly Met Gln Asp Pro Leu Ser Pro Lys Ala
 835 840 845
 Pro Leu Ile Cys Thr Lys Glu Glu Val Glu Glu Val Leu Asp Ser Lys
 850 855 860
 Ala Gly Trp Gly Ser Pro Cys His Leu Ser Gly Glu Ser Val Ile Leu
 865 870 875 880
 Leu Gly Pro Thr Val Gly Thr Glu Ser Lys Val Gln Ser Trp Phe Glu
 885 890 895

Ser Ser Leu Ser His Met Lys Pro Gly Glu Glu Gly Pro Asp Gly Glu
 900 905 910
 Arg Ala Pro Gly Asp Ser Thr Thr Ser Asp Ala Ser Leu Ala Gln Lys
 915 920 925
 Pro Asn Lys Pro Ala Val Pro Glu Ala Pro Ile Ala Lys Lys Glu Pro
 930 935 940
 Val Pro Arg Gly Lys Ser Leu Arg Ser Arg Arg Val His Arg Gly Leu
 945 950 955 960
 Pro Glu Ala Glu Asp Ser Pro Cys Arg Ala Pro Val Leu Pro Lys Asp
 965 970 975
 Leu Leu Leu Pro Glu Ser Cys Thr Gly Pro Pro Gln Gly Gln Met Glu
 980 985 990
 Gly Ala Gly Ala Pro Gly Arg Gly Ala Ser Glu Gly Leu Pro Arg Met
 995 1000 1005
 Cys Thr Arg Ser Leu Thr Ala Leu Ser Glu Pro Arg Thr Pro Gly Pro
 1010 1015 1020
 Pro Gly Leu Thr Thr Thr Pro Ala Pro Pro Asp Lys Leu Gly Gly Lys
 1025 1030 1035 1040
 Gln Arg Ala Ala Phe Lys Ser Gly Lys Arg Val Gly Lys Pro Ser Pro
 1045 1050 1055
 Lys Ala Ala Ser Ser Pro Ser Asn Pro Ala Ala Leu Pro Val Ala Ser
 1060 1065 1070
 Asp Ser Ser Pro Met Gly Ser Lys Thr Lys Glu Thr Asp Ser Pro Ser
 1075 1080 1085
 Thr Pro Gly Lys Asp Gln Arg Ser Met Ile Leu Arg Ser Arg Thr Lys
 1090 1095 1100
 Thr Gln Glu Ile Phe His Ser Lys Arg Arg Arg Pro Ser Glu Gly Arg
 1105 1110 1115 1120
 Leu Pro Asn Cys Arg Ala Thr Lys Lys Leu Leu Asp Asn Ser His Leu
 1125 1130 1135
 Pro Ala Thr Phe Lys Val Ser Ser Ser Pro Gln Lys Glu Gly Arg Val
 1140 1145 1150
 Ser Gln Arg Ala Arg Val Pro Lys Pro Gly Ala Gly Ser Lys Leu Ser
 1155 1160 1165
 Asp Arg Pro Leu His Ala Leu Lys Arg Lys Ser Ala Phe Met Ala Pro
 1170 1175 1180
 Val Pro Thr Lys Lys Arg Asn Leu Val Leu Arg Ser Arg Ser Ser Ser
 1185 1190 1195 1200
 Ser Ser Asn Ala Ser Gly Asn Gly Gly Asp Gly Lys Glu Glu Arg Pro
 1205 1210 1215
 Glu Gly Ser Pro Thr Leu Phe Lys Arg Met Ser Ser Pro Lys Lys Ala
 1220 1225 1230
 Lys Pro Thr Lys Gly Asn Gly Glu Pro Ala Thr Lys Leu Pro Pro Pro
 1235 1240 1245
 Glu Thr Pro Asp Ala Cys Leu Lys Leu Ala Ser Arg Ala Ala Phe Gln
 1250 1255 1260

Gly Ala Met Lys Thr Lys Val Leu Pro Pro Arg Lys Gly Arg Gly Leu
 1265 1270 1275 1280
 Lys Leu Glu Ala Ile Val Gln Lys Ile Thr Ser Pro Ser Leu Lys Lys
 1285 1290 1295
 Phe Ala Cys Lys Ala Pro Gly Ala Ser Pro Gly Asn Pro Leu Ser Pro
 1300 1305 1310
 Ser Leu Ser Asp Lys Asp Arg Gly Leu Lys Gly Ala Gly Gly Ser Pro
 1315 1320 1325
 Val Gly Val Glu Glu Gly Leu Val Asn Val Gly Thr Gly Gln Lys Leu
 1330 1335 1340
 Pro Thr Ser Gly Ala Asp Pro Leu Cys Arg Asn Pro Thr Asn Arg Ser
 1345 1350 1355 1360
 Leu Lys Gly Lys Leu Met Asn Arg Ala Ser Gly Thr Gln Gly Ala Ser
 1365 1370 1375
 Glu Asp Asn Ser Gly Gly Gly Gly Lys Lys Pro Lys Met Glu Glu Leu
 1380 1385 1390
 Gly Leu Ala Ser Gln Pro Pro Glu Gly Arg Pro Cys Gln Pro Gln Thr
 1395 1400 1405
 Arg Ala Gln Lys Gln Pro Gly His Thr Asn Tyr Ser Ser Tyr Ser Lys
 1410 1415 1420
 Arg Lys Arg Leu Thr Arg Gly Arg Ala Lys Asn Thr Thr Ser Ser Pro
 1425 1430 1435 1440
 Cys Lys Gly Arg Ala Lys Arg Arg Arg Gln Gln Gln Val Leu Pro Leu
 1445 1450 1455
 Asp Pro Ala Glu Pro Glu Ile Arg Leu Lys Tyr Ile Ser Ser Cys Lys
 1460 1465 1470
 Arg Leu Arg Ser Asp Ser Arg Thr Pro Ala Phe Ser Pro Phe Val Arg
 1475 1480 1485
 Val Glu Lys Arg Asp Ala Phe Thr Thr Ile Cys Thr Val Val Asn Ser
 1490 1495 1500
 Pro Gly Asp Ala Pro Lys Pro His Arg Lys Pro Ser Ser Ser Ala Ser
 1505 1510 1515 1520
 Ser Ser Ser Ser Ser Ser Ser Phe Ser Leu Asp Ala Ala Gly Ala Ser
 1525 1530 1535
 Leu Ala Thr Leu Pro Gly Gly Ser Ile Leu Gln Pro Arg Pro Ser Leu
 1540 1545 1550
 Pro Leu Ser Ser Thr Met His Leu Gly Pro Val Val Ser Lys Ala Leu
 1555 1560 1565
 Ser Thr Ser Cys Leu Val Cys Cys Leu Cys Gln Asn Pro Ala Asn Phe
 1570 1575 1580
 Lys Asp Leu Gly Asp Leu Cys Gly Pro Tyr Tyr Pro Glu His Cys Leu
 1585 1590 1595 1600
 Pro Lys Lys Lys Pro Lys Leu Lys Glu Lys Val Arg Pro Glu Gly Thr
 1605 1610 1615
 Cys Glu Glu Ala Ser Leu Pro Leu Glu Arg Thr Leu Lys Gly Pro Glu
 1620 1625 1630

Cys Ala Ala Ala Ala Thr Ala Gly Lys Pro Pro Arg Pro Asp Gly Pro
 1635 1640 1645
 Ala Asp Pro Ala Lys Gln Gly Pro Leu Arg Thr Ser Ala Arg Gly Leu
 1650 1655 1660
 Ser Arg Arg Leu Gln Ser Cys Tyr Cys Cys Asp Gly Arg Glu Asp Gly
 1665 1670 1675 1680
 Gly Glu Glu Ala Ala Pro Ala Asp Lys Gly His Trp Val His Glu Ala
 1685 1690 1695
 Cys Ala Val Trp Thr Gly Gly Val Tyr Leu Val Ala Gly Lys Leu Phe
 1700 1705 1710
 Gly Leu Gln Glu Ala Met Lys Val Ala Val Asp Met His Gly Gly Thr
 1715 1720 1725
 Val Ala Leu Ala Pro Gly Asp Phe Ser Leu Pro Gly Leu Arg Phe Ala
 1730 1735 1740
 Ser Leu Phe Gln Gly Pro Ser Trp Cys Asp Cys Pro Val Leu Ala Thr
 1745 1750 1755 1760
 Ser Thr Pro Ser Ser Trp Ser Arg Cys Val Pro Ala Ala Lys Lys Pro
 1765 1770 1775
 Gly Pro Pro Leu Gly Ala Ala Thr Lys Asp Ala Ser Thr Pro Thr Thr
 1780 1785 1790
 Thr Arg Val Pro Ala Met
 1795