



(19)  
Bundesrepublik Deutschland  
Deutsches Patent- und Markenamt

(10) DE 698 31 737 T2 2006.06.29

(12)

## Übersetzung der europäischen Patentschrift

(97) EP 0 977 871 B1

(21) Deutsches Aktenzeichen: 698 31 737.8

(86) PCT-Aktenzeichen: PCT/IB98/00625

(96) Europäisches Aktenzeichen: 98 913 981.1

(87) PCT-Veröffentlichungs-Nr.: WO 98/049322

(86) PCT-Anmeldetag: 24.04.1998

(87) Veröffentlichungstag

der PCT-Anmeldung: 05.11.1998

(97) Erstveröffentlichung durch das EPA: 09.02.2000

(97) Veröffentlichungstag

der Patenterteilung beim EPA: 28.09.2005

(47) Veröffentlichungstag im Patentblatt: 29.06.2006

(51) Int Cl.<sup>8</sup>: C12N 15/57 (2006.01)

C12N 9/64 (2006.01)

A61K 31/70 (2006.01)

A61K 38/48 (2006.01)

C12Q 1/37 (2006.01)

A61K 48/00 (2006.01)

C07K 16/40 (2006.01)

C12N 15/00 (2006.01)

A01K 67/027 (2006.01)

C12Q 1/68 (2006.01)

(30) Unionspriorität:

96697 26.04.1997 CH

(73) Patentinhaber:

Sonderegger, Peter, Zürich, CH

(74) Vertreter:

Haußingen, P., Ing. Faching. f.  
Schutzrechtswesen, Pat.-Anw., 06526  
Sangerhausen

(84) Benannte Vertragsstaaten:

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

(72) Erfinder:

Sonderegger, Peter, 8057 Zürich, CH

(54) Bezeichnung: NEUROTRYPSIN

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 eingelebt, 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

### Technisches Gebiet

**[0001]** Die vorliegende Erfindung betrifft Neurotrypsin und eine pharmazeutische Zusammensetzung, welche diese Substanzen enthält oder auf diese Substanzen einwirkt.

### Beschreibung der Erfindung

**[0002]** Neurotrypsin ist eine neu entdeckte Serinprotease, welche vor allem im Gehirn und in der Lunge exprimiert wird; die Expression im Gehirn findet fast ausschliesslich in Nervenzellen statt.

**[0003]** Neurotrypsin hat eine bisher nicht gefundene Domänenzusammensetzung: neben der Protease-Domäne findet man 3 oder 4 SRCR (Scavenger Receptor Cysteine-Rich)-Domänen und eine Kringel-Domäne. Es ist hervorzuheben, dass die Kombination von Kringel- und SRCR-Domänen bisher noch nie in Proteinen gefunden worden ist. Am Aminoterminus des Neurotrypsin-Proteins befindet sich ein Segment von über 60 Aminosäuren, welches einen ausserordentlich hohen Anteil von Prolin und basischen Aminosäuren (Arginin und Histidin) aufweist.

**[0004]** Die Erfindung ist durch die Merkmale in den unabhängigen Ansprüchen gekennzeichnet. Bevorzugte Ausführungsformen sind in den abhängigen Ansprüchen definiert.

**[0005]** Die neu gefundenen Neurotrypsine

- Neurotrypsin des Menschen (Verbindung der Formel I),
- Neurotrypsin der Maus (Verbindung der Formel II)

unterscheiden sich strukturell sehr stark von den bisher bekannten Serinproteasen.

**[0006]** Die Serinprotease, deren Protease-Domäne strukturell am nächsten mit der Protease-Domäne der neuen Verbindungen verwandt ist, nämlich Plasmin (des Menschen), weist eine Aminosäuresequenz-Identität von nur 44% auf.

**[0007]** Das Prolin-reiche, basische Segment am Aminoterminus weist eine gewisse Ähnlichkeit auf mit den basischen Segmenten der Netrine und der Semaphorine/Collapsine. Aufgrund dieses Segmentes ist es wahrscheinlich, dass Neurotrypsin mittels Heparin-Affinitätschromatographie angereichert werden kann.

**[0008]** Die Neurotrypsine des Menschen (Verbindung der Formel I) und der Maus (Verbindung der Formel II) weisen unter sich eine sehr hohe strukturelle Ähnlichkeit auf.

**[0009]** Die Identität der Aminosäuresequenzen der nativen Proteine der Verbindungen der Formeln I oder II beträgt 81%.

**[0010]** Das Neurotrypsin des Menschen (Verbindung der Formel I) hat eine kodierende Sequenz von 2625 Nukleotiden. Das kodierte Peptid der Verbindung der Formel I ist 875 Aminosäuren lang und enthält ein Signalpeptid von 20 Aminosäuren. Das Neurotrypsin der Maus (Verbindung der Formel II) hat eine kodierende Sequenz von 2283 Nukleotiden. Das kodierte Protein der Verbindung der Formel II ist 761 Aminosäuren lang und enthält ein Signalpeptid von 21 Aminosäuren. Der Grund für die grössere Länge des Neurotrypsins des Menschen liegt darin, dass das menschliche Neurotrypsin 4 SRCR-Domänen aufweist, während dem das Neurotrypsin der Maus nur 3 SRCR-Domänen hat.

**[0011]** Die Domänen, welche bei beiden Verbindungen (Verbindung der Formel I und Verbindung der Formel II) vorhanden sind, weisen einen hohen Grad von Sequenzähnlichkeit auf. Die einander entsprechenden SRCR-Domänen der Verbindungen der Formeln I und II weisen eine Aminosäuresequenzidentität von 81% bis 91% auf. Die entsprechenden Kringel-Domänen haben eine Aminosäuresequenzidentität von 75%. Ein hoher Grad von Ähnlichkeit besteht auch in der enzymatisch aktiven (d.h. proteolytischen) Domäne (90% Aminosäuresequenzidentität).

**[0012]** Die Proteasedomänen der Neurotrypsine des Menschen (Verbindung der Formel I) und der Maus (Verbindung der Formel II) sind im Folgenden gegeneinander aufgeregelt, um den hohen Grad von Sequenzidentität zu illustrieren.

CGLRLLHRRQKRIGGKNSLRGGWPWQVSLRLKSSHGDGRLLCGATLLSS 50  
|||||·|||:·|||:·|||:·|||:·|||:·|||:·|||:  
CGLRLLHRRQKRIGGNNSLRGAWPWQASLRLRSAHGDRLLCGATLLSS

**I K S V T K L** 258  
| | | | · |  
**I K S V T S L**

**[0013]** Von 258 Aminosäuresequenzpositionen, welche in den Vergleich einbezogenen worden sind, sind 233 Aminosäuren in beiden Verbindungen identisch (obere Sequenz: Verbindung der Formel I; untere Sequenz: Verbindung der Formel II; identische Aminosäuren sind mit senkrechten Strichen gekennzeichnet).

**[0014]** Die erfindungsgemässen Neurotrypsine sind verglichen mit den bekannten Serinproteasen einzigartig, weil sie gemäss gegenwärtigen Erkenntnissen in ausgeprägtem Masse von Nervenzellen exprimiert werden. Ein anderes Organ mit starker Expression von Neurotrypsin ist die Lunge (siehe Gschwend et al., Mol. Cell. Neurosci. 9, Seiten 207–219, 1997).

**[0015]** Die den Strukturen der Verbindungen der Formeln I oder II am strksten gleichenden Proteine sind Serinproteasen, wie etwa Gewebe-Plasminogenaktivator (tPA), Urokinase-Typ-Plasminogenaktivator (uPA), Plasmin, Trypsin, Apolipoprotein (a), Coagulation-Factor XI, Neuripsin, und Acrosin.

**[0016]** Im erwachsenen Gehirn werden die erfindungsgemäßen Verbindungen vorwiegend in der Grosshirnrinde, dem Hippocampus, und der Amygdala exprimiert.

**[0017]** Im erwachsenen Hirnstamm und Rückenmark werden die erfindungsgemäßen Verbindungen vorwiegend in den motorischen Nervenzellen exprimiert. Eine etwas schwächere Expression ist in den Nervenzellen der oberflächlichen Schichten des Hinterhorns des Rückenmarks zu finden.

**[0018]** Im erwachsenen peripheren Nervensystem werden die erfindungsgemäßen Verbindungen in einer Subpopulation der Spinalganglienurone exprimiert.

**[0019]** Die erfindungsgemässen Verbindungen wurden im Rahmen einer Studie gefunden, welche zum Ziel hatte, Trypsin-ähnliche Serinproteasen im Nervensystem aufzuspüren.

**[0020]** Die erste Verbindung, die gefunden und charakterisiert wurde, war die Verbindung der Formel II (siehe

Gschwend et al., Mol. Cell. Neurosci., 9, Seiten 207– 219, 1997).

**[0021]** Durch ein "Alignment" der Proteasedomänen von 7 bekannten Serinproteasen (Gewebe-Typ Plasminogenaktivator, Urokinase-Typ Plasminogenaktivator, Thrombin, Plasmin, Trypsin, Chymotrypsin und pankreatische Elastase) in der Nähe des Histidins und des Serins der katalytischen Triade der aktiven Stelle wurden die Sequenzen von so genannten "Primer-Oligonukleotiden" für die Polymerasen-Kettenreaktion ermittelt.

**[0022]** Die Primer-Oligonukleotiden wurden in einer Polymerasen-Kettenreaktion (PCR) zusammen mit ss-cDNA aus Total-RNA aus Gehirnen von 10 Tage alten Mäusen eingesetzt und führten zur Amplifizierung eines cDNA-Fragments mit einer Länge von ungefähr 500 Basenpaaren.

**[0023]** Dieses cDNA-Fragment wurde erfolgreich zur Isolierung von weiteren cDNA-Fragmenten mittels der Durchsuche von im Handel erhältlichen cDNA-Bibliotheken eingesetzt. Zusammen erstreckten sich die isolierten cDNA-Fragmente über die volle Länge des kodierenden Teils der Verbindung der Formel II.

**[0024]** Durch herkömmliche DNA-Sequenzierung wurden die vollständige Nukleotidsequenz und die davon abgeleitete Aminosäuresequenz erhalten.

**[0025]** Die Verbindung der Formel I wurde aufgrund ihrer ausgeprägten Ähnlichkeit mit der Verbindung der Formel II kloniert.

**[0026]** Die eingesetzten Primer-Oligonukleotide wurden gemäss der bekannten Sequenz der Verbindung der Formel II synthetisiert.

**[0027]** Die Klonierung der Verbindung der Formel I wurde mittels zweier im Handel erhältlicher cDNA-Bibliotheken aus fötalem menschlichem Gehirn durchgeführt.

**[0028]** Diese Art der Klonierung kann auch zur Isolierung der homologen Verbindung anderer Spezies, wie Ratte, Kaninchen, Meerschweinchen, Rind, Schaf, Schwein, Primaten, Vögel, Zebrafisch (*Brachydanio rerio*), *Drosophila melanogaster*, *Caenorhabditis elegans* etc., verwendet werden.

**[0029]** Die kodierenden Nukleotidsequenzen können eingesetzt werden zur Erzeugung von Proteinen mit den kodierten Aminosäuresequenzen der Verbindungen der Formeln I oder II. Ein in unserem Labor praktiziertes Verfahren erlaubt die Produktion von rekombinanten Proteinen in Myelomazellen als Fusionsproteine mit einer Immunoglobulin-Domäne (konstante Domäne der Leichten-Kette-Kappa). Das Konstruktionsprinzip ist im Detail beschrieben durch Rader et al. (Rader et al., Eur. J. Biochem. 215, Seiten 133–141, 1993). Das so von den Myelomazellen synthetisierte Fusionsprotein wurde durch Immunoaffinitätschromatographie mittels eines monoklonalen Antikörpers gegen die Ig-Domäne der Leichten-Kette-Kappa isoliert. Mit der gleichen Expressionsmethode kann auch das native Protein einer Verbindung, ausgehend von der kodierenden Sequenz, produziert werden.

**[0030]** Die kodierenden Sequenzen der Verbindungen der Formeln I oder II können als Ausgangsverbindungen dienen für das Aufspüren und die Isolierung von Allelen der Verbindungen der Formeln I oder II. Sowohl die Polymerasen-Kettenreaktion als auch die Nukleinsäure-Hybridisierung können zu diesem Zweck eingesetzt werden.

**[0031]** Die kodierenden Sequenzen der Verbindungen der Formeln I oder II können als Ausgangsverbindungen dienen für so genannte "site-directed mutagenesis", um Nukleotidsequenzen zu generieren, welche die durch die Verbindungen der Formeln I oder II kodierten Proteine, oder Teile davon, kodieren, aber deren Nukleotidsequenz im Bezug auf die Verbindungen der Formeln I oder II degeneriert sind, bedingt durch die Verwendung alternativer Codons.

**[0032]** Die kodierenden Sequenzen der Verbindungen der Formeln I oder II können verwendet werden als Ausgangsverbindungen für die Herstellung von Sequenzvarianten durch so genannte "site-directed mutagenesis".

## Beste Arten zur Ausführung der Erfindung (Beispiele)

## cDNA-Klonierung der Verbindung der Formel II (Neurotrypsin der Maus)

**[0033]** Totale RNA wurde aus dem Gehirn von 10 Tage alten Mäusen (ICR-ZUR) gemäss der Methode von Chomczynski und Sacchi (1987) isoliert. Die Herstellung von einzelsträngiger cDNA erfolgte unter Benützung von Oligo(dT)-Primer" und einer RNA-abhängigen DNA-Polymerase (Superscript RNase H<sup>-</sup>-Reverse Transcriptase; Gibco-BRL, Gaithersburg, MD) gemäss den Instruktionen des Herstellers. Für die Durchführung der Polymerasen-Kettenreaktion wurden ein vorwärts-gerichteter "Primer", gemäss der Aminosäuresequenz der Region des konservierten Histidins der katalytischen Triade, und ein rückwärts-gerichteter "Primer", gemäss der Aminosäuresequenz der Region des konservierten Serins der katalytischen Triade der Serinproteasen, synthetisiert. Die Aminosäuresequenzen, welche für die Bestimmung der Oligonukleotid-„Primer“ verwendet wurden, wurden von 7 bekannten Serinproteasen entnommen. Sie sind im Folgenden dargestellt.

Protease-Domäne	N-	I	II	C
tPA (m)	..SSC	W V L S A A H C	FLE.....HDA	C Q G D S G G PLV..
uPA (m)	..SPC	W V A S A A H C	FIQ.....TDS	C K G D S G G PLI..
Thrombin (m)	..SDR	W V L T A A H C	ILY.....GDA	C E G D S G G PFV..
Plasmin (m)	..APE	W V L T A A H C	LKS.....VDS	C Q G D S G G PLV..
Trypsin (m)	..NDQ	W V V S A A H C	YKY.....KDS	C Q G D S G G PVV..
Chymotrypsin (r)	..SED	W V V T A A H C	GVK.....VSS	C M G D S G G PLV..
pankr. Elast. II (m)	..ANN	W V L T A A H C	LSN.....TSS	C N G D S G G PLN..

Primer (I) 5'-TGG GTI SYI WSI GCI GCI CAT TG' (II) 3'-ACR BTY CCI CTR WSI CCI CC-5'

**[0034]** Die Protease-Domänen von 7 bekannten Serinproteasen (Gewebe-Typ-Plasminogenaktivator, Urokinase-typ Plasminogenaktivator, Thrombin, Plasmin, Trypsin, Chymotrypsin und pankreatische Elastase) wurden im Bereich des konservierten Histidins und Serins der katalytischen Triade der aktiven Stelle aufgereiht. Die in diesen Regionen konservierten Aminosäuren wurden als Basis für die Bestimmung der degenerierten "Primer" benutzt. Die Primersequenzen sind nach der Empfehlung der IUB-Nomenklatur (Nomenclature Committee, 1985) angegeben.

**[0035]** Die in der PCR eingesetzten Primer trugen zur Erleichterung einer späteren Klonierung die Restriktionsstellen für EcoRI und BamHI an ihren 5'-Enden.

**[0036]** Folgende Primer wurden eingesetzt:

**[0037]** In Leserichtung (sense primers):  
5'-GGGAATTCTGGGT(C/G)(T/C)I(T/A)(G/C)IGCIGCICA(T/C)TG-3'

**[0038]** In Gegenrichtung (antisense primers):  
5'-GGGGATCCCCICCI(G/C)(A/T)(A/G)TCICC(C/T)T(G/C/T)(G/A)CA-3'.

**[0039]** Die Polymerasen-Kettenreaktion wurde unter Standard-Bedingungen mittels der DNA-Polymerase AmpliTaq (Perkin Elmer) gemäss den Empfehlungen des Produzenten durchgeführt. Das folgende PCR-Profil wurde eingesetzt: 93°C für 3 Minuten, gefolgt von 35 Zyklen von 93°C für 1 Minute, 48°C für 2 Minuten und 72°C für 2 Minuten. Im Anschluss an den letzten Zyklus wurde die Inkubation bei 72°C während weiteren 10 Minuten fortgesetzt.

**[0040]** Die amplifizierten Fragmente hatten eine ungefähre Länge von 500 Basenpaaren. Sie wurden mit EcoRI und BamHI geschnitten und in einen Bluescript-Vektor eingefügt (Bluescript SK(-), Stratagene). Die resultierenden Klone wurden durch DNA-Sequenzbestimmung mittels der Didesoxy-Kettenterminations-Methode (Sanger et al., Proc. Natl. Acad. Sci. USA 77, Seiten 2163–2167, 1977) auf einem automatisierten DNA-Se-

quenziergerät (LI-COR, Modell 4000L; Lincoln, NE) unter Benützung eines kommerziellen Sequenzierkits (SequiTherm long-read cycle sequencing kit-LC; Epicentre Technologies, Madison, WI) analysiert. Die Analyse führte zu einer Sequenz von 474 Basenpaaren der katalytischen Region der Serinprotease-Domäne der Verbindung der Formel II.

**[0041]** Das 474 Basenpaar lange PCR-Fragment wurde zum Absuchen einer Oligo(dT)-"primed" Uni-ZAP-XR-cDNA-Bibliothek aus dem Gehirn von 20 Tage alten Mäusen (Stratagene; Cat. No. 937 319) eingesetzt. Es wurden  $3 \times 10^6$  Lambda-Plaques mittels des radioaktiv markierten PCR-Fragments als Sonde unter hochstringenten Bedingungen (Sambrook et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, 1989) abgesucht und es wurden 24 positive Klone gefunden.

**[0042]** Aus den positiven Lambda-Uni-ZAP-XR-Phagemid-Klonen wurde das entsprechende Bluescript-Plasmid nach der Standardmethode gemäss den Empfehlungen des Herstellers (Stratagene) durch in vivo-Exzision herausgeschnitten. Um die Länge der eingefügten Fragmente zu bestimmen, wurden die entsprechenden Bluescript-Plasmid-Klone mit SacI und KpnI verdaut. Die Klone, welche die längsten Fragmente enthielten, wurden mittels DNA-Sequenzierung (wie oben beschrieben) analysiert, und für die anschliessende Daten-Auswertung wurde die GCG-Software (Version 8.1, Unix; Silicon Graphics, Inc.) verwendet.

**[0043]** Da keiner der Klone die kodierende Sequenz in voller Länge enthielt, wurde eine zweite cDNA-Bibliothek abgesucht. Die in dieser Absuche eingesetzte Bibliothek war eine Oligo(dT)- und "Random-Primed" cDNA Bibliothek in einem Lambda-Phagen (Lambda gt10), welche auf mRNA aus 15 Tage alten Maus-Embryonen basierte (oligo(dT)- and random-primed Lambda gt10 cDNA library; Clontech, Palo Alto, CA; Kat. No. ML 3002a). Als Sonde wurde ein radioaktiv markiertes DNA-Fragment (Aval/AatII) vom 5'-Ende des längsten Klones der ersten Suche eingesetzt, und es wurden ungefähr  $2 \times 10^6$  Plaques abgesucht. Diese Absuche ergab 14 positive Klone. Die cDNA-Fragmente wurden mittels EcoRI herausgeschnitten und in den Bluescript-Vektor (KS(+); Stratagene) kloniert. Die Sequenzanalyse wurde wie oben beschrieben ausgeführt.

**[0044]** Man erhielt so die Nukleotidsequenz über die volle Länge der cDNA von 2361 resp. 2376 Basenpaaren der Verbindung der Formel II. Mit dem beschriebenen Verfahren der PCR-Klonierung ist es möglich, auch Varianten-Formen der Verbindungen der Formeln I und II zu finden und zu isolieren, beispielsweise deren Allele, oder deren Splice-Varianten. Das beschriebene Verfahren des Absuchens einer cDNA-Bibliothek ermöglicht auch das Auffinden und die Isolierung von Verbindungen, welche unter stringenten Bedingungen an die kodierenden Sequenzen der Formeln I und II hybridisieren.

#### Klonierung der cDNA der Verbindung der Formel I (Neurotrycin des Menschen)

**[0045]** Die Klonierung der cDNA der Verbindung der Formel I wurde auf der Grundlage der Nukleotidsequenz der Verbindung der Formel II durchgeführt. Als erster Schritt wurde ein Fragment der Verbindung der Formel I mittels Polymerasenkettreaktion (PCR) amplifiziert. Als Matrize wurde die DNA verwendet, welche aus einer cDNA-Bibliothek vom Gehirn eines menschlichen Fötus (17.–18. Schwangerschaftswoche) erhalten wurde, welche auf dem Markt erhältlich ist (Oligo(dT)- and random-primed, human fetal brain cDNA library in the Lambda ZAP II vector, Cat. No. 936206, Stratagene). Die synthetischen PCR-Primer enthielten, zur Erleichterung der nachfolgenden Klonierung, die Restriktionsstellen HindIII und Xhol am 5'-Ende.

**[0046]** In Leserichtung (sense primers):

5'-GGGAAGCTTGGICA(A/G)TGGGGIACI(A/G)TITG(C/T)GA(C/T)-3'

**[0047]** In Gegenrichtung (antisense primer):

5'-GGGCTCGAGCCCCAACCTGTTATGTAAIAGTTG-3'

**[0048]** Die PCR wurde unter Standard-Bedingungen mittels der DNA-Polymerase AmpliTaq (Perkin Elmer) gemäss den Empfehlungen des Produzenten durchgeführt. Das entstandene Fragment von 1116 Basenpaaren wurde in den Bluescript-Vektor (Bluescript SK(-), Stratagene) eingefügt. Ein 600 Basenpaare-langes HindIII/StuI-Fragment, entsprechend der 5'-Hälfte des 1116 Basenpaare-langen PCR-Fragments, wurde zum Absuchen einer Lambda-cDNA-Bibliothek aus menschlichem fötalem Gehirn (Human Fetal Brain 5'-STRETCH PLUS cDNA library; Lambda gt10; Cat. No. HL3003a; Clontech) eingesetzt. Es wurden  $2 \times 10^6$  Lambda-Plaques mittels des radioaktiv markierten PCR-Fragments unter hochstringenten Bedingungen (Sambrook et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, 1989) abgesucht, und es wurden 23 positive Klone gefunden und isoliert.

**[0049]** Aus den positiven Lambda-gt10-Klonen wurden die entsprechenden cDNA-Fragmente mit EcoRI herausgeschnitten und in einen Bluescript-Vektor (Bluescript KS(+), Stratagene) eingefügt. Die Sequenzierung erfolgte mittels der Didesoxy-Ketteterminations-Methode (Sanger et al., Proc. Natl. Acad. Sci. USA 77, Seiten 2163–2167, 1977), unter Verwendung eines kommerziellen Sequenzierkits (Sequi Therm long-read cycle sequencing kit-LC; Epicentre Technologies, Madison, WI) und Bluescript-spezifischen Primern.

**[0050]** In einer alternativen Sequenzier-Strategie wurden die cDNA-Fragmente der positiven Lambda-gt10-Klonen, unter Verwendung Lambda-spezifischer Primer, mittels PCR amplifiziert. Die Sequenzierung wurde wie oben beschrieben durchgeführt.

**[0051]** Die computerisierte Analyse der Sequenzen wurde mittels des Programmpakets GCG (Version 8.1, Unix; Silicon Graphics Inc.) durchgeführt.

**[0052]** Man erhält so die Nukleotidsequenz über die volle Länge der cDNA von 3350 Basenpaaren. Mit dem beschriebenen Verfahren der PCR-Klonierung ist es möglich, auch Varianten-Formen der Verbindungen der Formeln I oder II zu finden und zu isolieren, beispielsweise deren Allele, oder deren Splice-Varianten. Das beschriebene Verfahren des Absuchens einer cDNA-Bibliothek ermöglicht auch das Auffinden und die Isolierung von Verbindungen, welche unter stringenten Bedingungen mit den codierenden Sequenzen der Formeln I oder II hybridisieren.

#### Visualisierung der codierten Sequenzen der Verbindungen I oder II mittels Antikörpern

**[0053]** Das mehr als 60 Aminosäuren lange Prolin-reiche, basische Segment am Aminoterminus der codierten Sequenz der Verbindungen der Formeln I oder II eignet sich gut für die Herstellung von Antikörpern mittels der Synthese von Peptiden und deren Einsatz zur Immunisierung. Wir haben aus dem Prolin-reichen, basischen Segment am Aminoterminus der kodierten Sequenz der Verbindung der Formel II zwei Peptidsequenzen mit einer Länge von 19 und 13 Aminosäuren zur Erzeugung von Antikörpern ausgewählt. Die Peptide hatten die folgenden Sequenzen:

Peptid 1: H<sub>2</sub>N-SRS PLH RPH PSP PRS QX-CO NH<sub>2</sub>

Peptid 2: H<sub>2</sub>N-LPS SRR PPR TPR F-COOH

**[0054]** Die beiden Peptide wurden chemisch synthetisiert, an eine makromolekulare Trägersubstanz (Keyhole Limpet Hemocyanin) gekoppelt, und zur Immunisierung in 2 Kaninchen injiziert. Die erzeugten Antiseren wiesen einen hohen Antikörper-Titer auf und konnten erfolgreich sowohl zur Identifizierung von nativem Neutrordin aus Gehirnextrakt der Maus als auch zur Identifizierung von rekombinantem Neutrordin verwendet werden. Das angewandte Verfahren zur Erzeugung von Antikörpern kann auch zur Erzeugung von Antikörpern gegen die kodierte Sequenz der Verbindung der Formel I angewendet werden.

**[0055]** Die resultierenden Antikörper gegen die Teilequenzen der kodierten Sequenzen der Formeln I oder II können zur Aufspürung und zur Isolierung von Varianten-Formen der Verbindungen der Formeln I oder II, wie beispielsweise Allele oder Splice-Varianten, eingesetzt werden. Solche Antikörper können auch verwendet werden für die Auffindung und Isolierung von gentechnisch erzeugten Varianten der Verbindungen der Formeln I oder II.

#### Reinigung der kodierten Sequenzen der Verbindungen der Formeln I oder II

**[0056]** Neben konventionellen chromatographischen Methoden, wie beispielsweise Ionenaustauscher-Chromatographie, kann die Reinigung der kodierten Sequenzen der Verbindungen der Formeln I oder II, auch erreicht werden unter der Verwendung von zwei affinitätschromatographischen Reinigungsverfahren. Eine affinitätschromatographische Reinigungsprozedur basiert auf der Verfügbarkeit von Antikörpern. Durch Kopplung der Antikörper an eine chromatographische Matrix resultiert ein Reinigungsverfahren, in welchem ein sehr hoher Grad an Reinheit der entsprechenden Verbindung in einem Schritt erzielt werden kann.

**[0057]** Ein anderes wichtiges Merkmal, welches für die Reinigung der kodierten Sequenzen der Verbindungen der Formeln I und II verwendet werden kann, ist das Prolin-reiche, basische Segment am Aminoterminus. Es ist zu erwarten, dass, aufgrund der hohen Dichte an positiven Ladungen, dieses Segment die Bindung der kodierten Sequenzen der Verbindungen der Formeln I oder II an Heparin und Heparin-ähnliche Affinitätsmatrices vermittelt. Dieses Prinzip ermöglicht auch die Isolierung, oder zumindest die Anreicherung, von Varianten-Formen der kodierten Sequenzen der Formeln I oder II, beispielsweise deren Allele oder Splice-Varianten. Gleichermaßen kann die Heparin-Affinitätschromatographie auch zur Isolierung, oder zumindest zur Anrei-

cherung, von spezieshomologen Proteinen der Verbindungen der Formeln I oder II eingesetzt werden.

#### Industrielle Anwendbarkeit

**[0058]** Die kodierenden Sequenzen der Verbindungen der Formeln I und II können verwendet werden für die Herstellung der kodierten Proteine, oder Teilen davon, der Formeln I und II. Die Herstellung der kodierten Proteine kann in prokaryotischen oder eukaryotischen Expressionssystemen erzielt werden.

**[0059]** Das Gen-Expressionsmuster der erfindungsgemäßen Verbindungen im Gehirn ist äusserst interessant, weil diese Moleküle im adulten Nervensystem vor allem in Nervenzellen derjenigen Regionen exprimiert werden, denen eine wichtige Rolle bei Lern- und Gedächtnisfunktionen zugeschrieben wird. Zusammen mit der kürzlich gefundenen Evidenz für eine Rolle von extrazellulären Proteasen bei der neuralen Plastizität, lässt das Expressionsmuster vermuten, dass die proteolytische Wirkung von Neurotrypsin eine Rolle innehat bei strukturellen Reorganisationen im Zusammenhang mit Lern- und Gedächtnis-Operationen, zum Beispiel Operationen, welche an der Verarbeitung und Speicherung von erlernten Verhaltensweisen, erlernten Gefühlen oder von Gedächtnisinhalten beteiligt sind. Die erfindungsgemäßen Verbindungen können deshalb Ziele für pharmazeutische Interventionen bei Funktionsstörungen des Gehirns sein.

**[0060]** Das Gen-Expressionsmuster der erfindungsgemäßen Verbindungen in der Grosshirnrinde (vor allem Schichten V und VI) ist äusserst interessant, weil eine Reduktion der zellulären Differenzierung in der Grosshirnrinde in Assoziation mit Schizophrenie gefunden wurde. Die erfindungsgemäßen Verbindungen können deshalb Ziele für pharmazeutische Interventionen bei Schizophrenie und verwandten psychiatrischen Krankheiten sein.

**[0061]** Es ist gefunden worden, dass die kodierenden Sequenzen der erfindungsgemäßen Verbindungen in Neuronen erhöht sind, welche an das beschädigte Gewebe eines fokalen ischämischen Hirnschlags angrenzen, was darauf hinweist, dass die erfindungsgemäßen Verbindungen eine Rolle in der Gewebereaktion in verletztem zerebralem Gewebe spielen. Die erfindungsgemäßen Verbindungen können deshalb Ziele für pharmazeutische Interventionen nach einem ischämischen Hirnschlag und anderen Formen von Beschädigungen von neuralem Gewebe sein.

**[0062]** Vom Gewebe-Typ Plasminogenaktivator, eine Serinprotease, welche mit den erfindungsgemäßen Verbindungen verwandt ist, wurde kürzlich gefunden, dass er in Excitotoxizitäts-vermittelten neuronalen Zelltod involviert ist. Eine ähnliche Funktion ist denkbar für die erfindungsgemäßen Verbindungen und, folglich, stellen die erfindungsgemäßen Verbindungen ein mögliches Ziel für pharmakologische Interventionen bei Krankheiten dar, in welchen der Zelltod auftritt.

**[0063]** Das Gen-Expressionsmuster der erfindungsgemäßen Verbindungen im Rückenmark und in den sensorischen Ganglien ist interessant, weil diese Moleküle im adulten Nervensystem in Neuronen derjenigen Gehirnregionen exprimiert werden, denen eine Rolle bei der Verarbeitung von Schmerz, sowie bei der Entstehung pathologischer Schmerzzustände, zugeschrieben wird. Die erfindungsgemäßen Verbindungen können deshalb Ziele für pharmazeutische Interventionen bei pathologischem Schmerz sein.

**[0064]** Im folgenden Teil werden Angaben bezüglich der Verbindungen der Formeln I oder II gemacht:

(1) ANGABEN ZUR VERBINDUNG DER FORMEL I (Neurotrypsin des Menschen)

(i) SEQUENZKENNZEICHEN:

- (A) LÄNGE: 3350 Basenpaare
- (B) ART: Nukleinsäure
- (C) STRANGFORM: Einzelstrang
- (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: c-DNA zu m-RNA

(vi) URSPRÜNGLICHE HERKUNKFT:

- (A) ORGANISMUS: Homo sapiens
- (D) ENTWICKLUNGSSTADIUM: Fötalstadium
- (F) GEWEBETYP: Gehirn

(vii) UNMITTELBARE HERKUNFT:

(A) BIBLIOTHEK: human fetal brain 5'-stretch plus cDNA library in the lambda gt10 vector; catalog No. HL 3003a; Clontech, Palo Alto, CA, USA.

(B) CLONE: cDNA-Klone No.:

3-1, 3-2, 3-6, 3-7, 3-8, 3-10, 3-11, 3-12

(ix) MERKMAL:

- (A) NAME/SCHLÜSSEL: Signalpeptid
- (B) LAGE: 44 .. 103

(ix) MERKMAL:

- (A) NAME/SCHLÜSSEL: reifes Peptid
- (B) LAGE: 104 .. 2668

(ix) MERKMAL:

- (A) NAME/SCHLÜSSEL: codierende Sequenz
- (B) LAGE: 44 .. 2668

(ix) MERKMAL:

- (A) NAME/SCHLÜSSEL: Prolin-reiches, basisches Segment
- (B) LAGE: 104 .. 319

(ix) MERKMAL:

- (A) NAME/SCHLÜSSEL: Kringel-Domäne
- (B) LAGE: 320 .. 538

(ix) MERKMAL:

- (A) NAME/SCHLÜSSEL: SRCR-Domäne 1
- (B) LAGE: 551 .. 856

(ix) MERKMAL:

- (A) NAME/SCHLÜSSEL: SRCR-Domäne 2
- (B) LAGE: 881 .. 1186

(ix) MERKMAL:

- (A) NAME/SCHLÜSSEL: SRCR-Domäne 3
- (B) LAGE: 1202 .. 1504

(ix) MERKMAL:

- (A) NAME/SCHLÜSSEL: SRCR-Domäne 4
- (B) LAGE: 1541 .. 1846

(ix) MERKMAL:

- (A) NAME/SCHLÜSSEL: proteolytische Domäne
- (B) LAGE: 1898 .. 2668

(ix) MERKMAL:

- (A) NAME/SCHLÜSSEL: Histidin der katalytischen Triade
- (B) LAGE: 2069 - 2071

(ix) MERKMAL:

- (A) NAME/SCHLÜSSEL: Asparaginsäure der katalytischen Triade
- (B) LAGE: 2219 - 2221

(ix) MERKMAL:

- (A) NAME/SCHLÜSSEL: Serin der katalytischen Triade
- (B) LAGE: 2516 .. 2518

(ix) MERKMAL:

- (A) NAME/SCHLÜSSEL: PolyA-Signal
- (B) LAGE: 2873 .. 2878

(ix) MERKMAL

- (A) NAME/SCHLÜSSEL: PolyA-Signal
- (B) LAGE: 3034 .. 3039

(ix) MERKMAL:

- (A) NAME/SCHLÜSSEL: PolyA-Signal
- (B) LAGE: 3215 .. 3220

(ix) MERKMAL

- (A) NAME/SCHLÜSSEL: 3'UTR
- (B) LAGE: 2669 .. 3350

(ix) MERKMAL

- (A) NAME/SCHLÜSSEL: 5'UTR
- (B) LAGE: 1 .. 43

## Verbindung der Formel I (Neurotrypsin des Menschen)

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

## DE 698 31 737 T2 2006.06.29

AGC CAC TGG GAT GAT TCT GAT GCA TCA GTC ATT TGT CAC CAG CTG CAG Ser His Trp Asp Asp Ser Asp Ala Ser Val Ile Cys His Gln Leu Gln 180 185 190	679
CTG GGA GGA AAA GGA ATA GCA AAA CAA ACC CCG TTT TCT GGA CTG GGC Leu Gly Gly Lys Gly Ile Ala Lys Gln Thr Pro Phe Ser Gly Leu Gly 195 200 205	727
CTT ATT CCC ATT TAT TGG AGC AAT GTC CGT TGC CGA GGA GAT GAA GAA Leu Ile Pro Ile Tyr Trp Ser Asn Val Arg Cys Arg Gly Asp Glu Glu 210 215 220	775
AAT ATA CTG CTT TGT GAA AAA GAC ATC TGG CAG GGT GGG GTG TGT CCT Asn Ile Leu Leu Cys Glu Lys Asp Ile Trp Gln Gly Gly Val Cys Pro 225 230 235 240	823
CAG AAG ATG GCA GCT GTC ACG TGT AGC TTT TCC CAT GGC CCA ACG Gln Lys Met Ala Ala Ala Val Thr Cys Ser Phe Ser His Gly Pro Thr 245 250 255	871
TTC CCC ATC ATT CGC CTT GCT GGA GGC AGC AGT GTG CAT GAA GGC CGG Phe Pro Ile Ile Arg Leu Ala Gly Gly Ser Ser Val His Glu Gly Arg 260 265 270	919
GTG GAG CTC TAC CAT GCT GGC CAG TGG GGA ACC GTT TGT GAT GAC CAA Val Glu Leu Tyr His Ala Gly Gln Trp Gly Thr Val Cys Asp Asp Gln 275 280 285	967
TGG GAT GAT GCC GAT GCA GAA GTG ATC TGC AGG CAG CTG GGC CTC AGT Trp Asp Asp Ala Asp Ala Glu Val Ile Cys Arg Gln Leu Gly Leu Ser 290 295 300	1015
GGC ATT GCC AAA GCA TGG CAT CAG GCA TAT TTT GGG GAA GGG TCT GGC Gly Ile Ala Lys Ala Trp His Gln Ala Tyr Phe Gly Glu Gly Ser Gly 305 310 315 320	1063
CCA GTT ATG TTG GAT GAA GTA CGC TGC ACT GGG AAT GAG CTT TCA ATT Pro Val Met Leu Asp Glu Val Arg Cys Thr Gly Asn Glu Leu Ser Ile 325 330 335	1111
GAG CAG TGT CCA AAG AGC TCC TGG GGA GAG CAT AAC TGT GGC CAT AAA Glu Gln Cys Pro Lys Ser Ser Trp Gly Glu His Asn Cys Gly His Lys 340 345 350	1159
GAA GAT GCT GGA GTG TCC TGT ACC CCT CTA ACA GAT GGG GTC ATC AGA Glu Asp Ala Gly Val Ser Cys Thr Pro Leu Thr Asp Gly Val Ile Arg 355 360 365	1207
CTT GCA GGT GGG AAA GGC AGC CAT GAG GGT CGC TTG GAG GTA TAT TAC Leu Ala Gly Gly Lys Gly Ser His Glu Gly Arg Leu Glu Val Tyr Tyr 370 375 380	1255
AGA GGC CAG TGG GGA ACT GTC TGT GAT GGC TGG ACT GAG CTG AAT Arg Gly Gln Trp Gly Thr Val Cys Asp Asp Gly Trp Thr Glu Leu Asn 385 390 395 400	1303

## DE 698 31 737 T2 2006.06.29

ACA TAC GTG GTT TGT CGA CAG TTG GGA TTT AAA TAT GGT AAA CAA GCA Thr Tyr Val Val Cys Arg Gln Leu Gly Phe Lys Tyr Gly Lys Gln Ala 405	410	415	1351	
TCT GCC AAC CAT TTT GAA GAA AGC ACA GGG CCC ATA TGG TTG GAT GAC Ser Ala Asn His Phe Glu Glu Ser Thr Gly Pro Ile Trp Leu Asp Asp 420	425	430	1399	
GTC AGC TGC TCA GGA AAG GAA ACC AGA TTT CTT CAG TGT TCC AGG CGA Val Ser Cys Ser Gly Lys Glu Thr Arg Phe Leu Gln Cys Ser Arg Arg 435	440	445	1447	
CAG TGG GGA AGG CAT GAC TGC AGC CAC CGC GAA GAT GTT AGC ATT GCC Gln Trp Gly Arg His Asp Cys Ser His Arg Glu Asp Val Ser Ile Ala 450	455	460	1495	
TGC TAC CCT GGC GGC GAG GGA CAC AGG CTC TCT CTG GGT TTT CCT GTC Cys Tyr Pro Gly Gly Glu Gly His Arg Leu Ser Leu Gly Phe Pro Val 465	470	475	480	1543
AGA CTG ATG GAT GGA GAA AAT AAG AAA GAA GGA CGA GTG GAG GTT TTT Arg Leu Met Asp Gly Glu Asn Lys Lys Glu Gly Arg Val Glu Val Phe 485	490	495	1591	
ATC AAT GGC CAG TGG GGA ACA ATC TGT GAT GAT GGA TGG ACT GAT AAG Ile Asn Gly Gln Trp Gly Thr Ile Cys Asp Asp Gly Trp Thr Asp Lys 500	505	510	1639	
GAT GCA GCT GTG ATC TGT CGT CAG CTT GGC TAC AAG GGT CCT GCC AGA Asp Ala Ala Val Ile Cys Arg Gln Leu Gly Tyr Lys Gly Pro Ala Arg 515	520	525	1687	
GCA AGA ACC ATG GCT TAC TTT GGA GAA GGA AAA GGA CCC ATC CAT GTG Ala Arg Thr Met Ala Tyr Phe Gly Glu Gly Lys Gly Pro Ile His Val 530	535	540	1735	
GAT AAT GTG AAG TGC ACA GGA AAT GAG AGG TCC TTG GCT GAC TGT ATC Asp Asn Val Lys Cys Thr Gly Asn Glu Arg Ser Leu Ala Asp Cys Ile 545	550	555	560	1783
AAG CAA GAT ATT GGA AGA CAC AAC TGC CGC CAC AGT GAA GAT GCA GGA Lys Gln Asp Ile Gly Arg His Asn Cys Arg His Ser Glu Asp Ala Gly 565	570	575	1831	
GTT ATT TGT GAT TAT TTT GGC AAG AAG GCC TCA GGT AAC AGT AAT AAA Val Ile Cys Asp Tyr Phe Gly Lys Lys Ala Ser Gly Asn Ser Asn Lys 580	585	590	1879	
GAG TCC CTC TCA TCT GTT TGT GGC TTG AGA TTA CTG CAC CGT CGG CAG Glu Ser Leu Ser Ser Val Cys Gly Leu Arg Leu Leu His Arg Arg Gln 595	600	605	1927	
AAG CGG ATC ATT GGT GGG AAA AAT TCT TTA AGG GGT GGT TGG CCT TGG Lys Arg Ile Ile Gly Gly Lys Asn Ser Leu Arg Gly Gly Trp Pro Trp 610	615	620	1975	

## DE 698 31 737 T2 2006.06.29

CAG GTT TCC CTC CGG CTG AAG TCA TCC CAT GGA GAT GGC AGG CTC CTC	2023
Gln Val Ser Leu Arg Leu Lys Ser Ser His Gly Asp Gly Arg Leu Leu	
625 630 635 640	
TGC GGG GCT ACG CTC CTG AGT AGC TGC TGG GTC CTC ACA GCA GCA CAC	2071
Cys Gly Ala Thr Leu Leu Ser Ser Cys Trp Val Leu Thr Ala Ala His	
645 650 655	
TGT TTC AAG AGG TAT GGC AAC AGC ACT AGG AGC TAT GCT GTT AGG GTT	2119
Cys Phe Lys Arg Tyr Gly Asn Ser Thr Arg Ser Tyr Ala Val Arg Val	
660 665 670	
GGA GAT TAT CAT ACT CTG GTA CCA GAG GAG TTT GAG GAA GAA ATT GGA	2167
Gly Asp Tyr His Thr Leu Val Pro Glu Glu Phe Glu Glu Glu Ile Gly	
675 680 685	
GTT CAA CAG ATT GTG ATT CAT CGG GAG TAT CGA CCC GAC CGC AGT GAT	2215
Val Gln Gln Ile Val Ile His Arg Glu Tyr Arg Pro Asp Arg Ser Asp	
690 695 700	
TAT GAC ATA GCC CTG GTT AGA TTA CAA GGA CCA GAA GAG CAA TGT GCC	2263
Tyr Asp Ile Ala Leu Val Arg Leu Gln Gly Pro Glu Glu Gln Cys Ala	
705 710 715 720	
AGA TTC AGC AGC CAT GTT TTG CCA GCC TGT TTA CCA CTC TGG AGA GAG	2311
Arg Phe Ser Ser His Val Leu Pro Ala Cys Leu Pro Leu Trp Arg Glu	
725 730 735	
AGG CCA CAG AAA ACA GCA TCC AAC TGT TAC ATA ACA GGA TGG GGT GAC	2359
Arg Pro Gln Lys Thr Ala Ser Asn Cys Tyr Ile Thr Gly Trp Gly Asp	
740 745 750	
ACA GGA CGA GCC TAT TCA AGA ACA CTA CAA CAA GCA GCC ATT CCC TTA	2407
Thr Gly Arg Ala Tyr Ser Arg Thr Leu Gln Gln Ala Ala Ile Pro Leu	
755 760 765	
CTT CCT AAA AGG TTT TGT GAA GAA CGT TAT AAG GGT CGG TTT ACA GGG	2455
Leu Pro Lys Arg Phe Cys Glu Glu Arg Tyr Lys Gly Arg Phe Thr Gly	
770 775 780	
AGA ATG CTT TGT GCT GGA AAC CTC CAT GAA CAC AAA CGC GTG GAC AGC	2503
Arg Met Leu Cys Ala Gly Asn Leu His Glu His Lys Arg Val Asp Ser	
785 790 795 800	
TGC CAG GGA GAC AGC GGA GGA CCA CTC ATG TGT GAA CGG CCC GGA GAG	2551
Cys Gln Gly Asp Ser Gly Gly Pro Leu Met Cys Glu Arg Pro Gly Glu	
805 810 815	
AGC TGG GTG GTG TAT GGG GTG ACC TCC TGG GGG TAT GGC TGT GGA GTC	2599
Ser Trp Val Val Tyr Gly Val Thr Ser Trp Gly Tyr Gly Cys Gly Val	
820 825 830	
AAG GAT TCT CCT GGT GTT TAT ACC AAA GTC TCA GCC TTT GTA CCT TGG	2647
Lys Asp Ser Pro Gly Val Tyr Thr Lys Val Ser Ala Phe Val Pro Trp	
835 840 845	

DE 698 31 737 T2 2006.06.29

ATA AAA AGT GTC ACC AAA CTG TAA TTCTTCATGG AAAACTTCAAA GCAGCATT 2700  
Ile Lys Ser Val Thr Lys Leu \*  
850 855

AAACAAATGG AAAACTTTGA ACCCCCCACTA TTAGCACTCA GCAGAGATGA CAACAAATGG 2760

CAAGATCTGT TTTTGCTTTG TGTTGTGGTA AAAAATTGTG TACCCCTGC TGCTTTGAG 2820

AAATTTGTGA ACATTTTCAG AGGCCTCAGT GTAGTGGAAG TGATAATCCT TAAATGAACA 2880

TTTTCTACCC TAATTCACT GGAGTGACTT ATTCTAAGCC TCATCTATCC CCTACCTATT 2940

TCTCAAAATC ATTCTATGCT GATTTACAA AAGATCATT TTACATTGGA ACTGAGAAC 3000

CCTTTAATT GAATCAGTGG TGTCTGAAAT CATATTAAAT ACCCACATT GACATAAATG 3060

CGGTACCCTT TACTACACTC ATGAGTGGCA TATTTATGCT TAGGTCTTT CAAAAGACTT 3120

GACAAGAAAT CTTCATATTC TCTGTAGCCT TTGTCAAGTG AGGAAATCAG TGGTTAAAGA 3180

ATTCCACTAT AAACCTTTAG GCCTGAATAG GAGTAGTAAA GCCTCAAGGA CATCTGCCTG 3240

TCACAATATA TTCTCAAAGT GATCTGATAT TTGGAAACAA GTATCCTTGT TGAGTACCAA 3300

GTGCTACAGA AACCATAGA TAAAAATACT TTCTACCTAC AGCGTGCCCC 3350

(1) ANGABEN ZUR VERBINDUNG DER FORMEL II (Neurotrypsin der Maus)

(i) SEQUENZKENNZEICHEN:

- (A) LÄNGE: 2376 Basenpaare
- (B) ART: Nukleinsäure
- (C) STRANGFORM: Einzelstrang
- (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: c-DNA zu m-RNA

(vi) URSPRÜNGLICHE HERKUNKFT:

- (A) ORGANISMUS: Mus musculus
- (D) ENTWICKLUNGSSTADIUM: Postnatal-Tag 10
- (F) GEWEBETYP: Gehirn
- (G) ZELLTYP: Neuronen

(vii) UNMITTELBARE HERKUNFT:

- (A) BIBLIOTHEK: mouse brain cDNA library in the lambda Uni-ZAP-XR vector,  
oligo (dT)-primed, from Balb c mice, postnatal day 20,  
Cat. No.. 937 319; Stratagene, La Jolla, CA, USA
- (B) CLONE: cDNA-Klon No. 16

(vii) UNMITTELBARE HERKUNFT:

(A) BIBLIOTHEK: mouse brain cDNA library in the Lambda gt10 vector,  
oligo(dT)- and random-primed, embryonic day 15,  
Cat. No. ML 3002a; Clontech, Palo Alto, CA, USA

(B) CLONE: cDNA-Klon No. 25

(ix) MERKMAL:

(A) NAME/SCHLÜSSEL: Signalpeptid  
(B) LAGE: 24 .. 86

(ix) MERKMAL:

(A) NAME/SCHLÜSSEL: reifes Peptid  
(B) LAGE: 87 .. 2306

(ix) MERKMAL:

(A) NAME/SCHLÜSSEL: kodierende Sequenz  
(B) LAGE: 24 .. 2306

(ix) MERKMAL:

(A) NAME/SCHLÜSSEL: Prolin-reiches, basisches Segment  
(B) LAGE: 90 .. 275

(ix) MERKMAL:

(A) NAME/SCHLÜSSEL: Kringel-Domäne

(B) LAGE: 276 .. 494

(ix) MERKMAL:

(A) NAME/SCHLÜSSEL: SRCR-Domäne 1

(B) LAGE: 519 .. 824

(ix) MERKMAL:

(A) NAME/SCHLÜSSEL: SRCR-Domäne 2

(B) LAGE: 840 .. 1142

(ix) MERKMAL:

(A) NAME/SCHLÜSSEL: SRCR-Domäne 3

(B) LAGE: 1179 .. 1484

(ix) MERKMAL:

(A) NAME/SCHLÜSSEL: proteolytische Domäne

(B) LAGE: 1536 .. 2306

(ix) MERKMAL:

(A) NAME/SCHLÜSSEL: Histidin der katalytischen Triade

(B) LAGE: 1707 .. 1709

(ix) MERKMAL:

- (A) NAME/SCHLÜSSEL: Asparaginsäure der katalytischen Triade
- (B) LAGE: 1857 .. 1859

(ix) MERKMAL:

- (A) NAME/SCHLÜSSEL: Serin der katalytischen Triade
- (B) LAGE: 2154 .. 2156

(ix) MERKMAL:

- (A) NAME/SCHLÜSSEL: PolyA-Signal
- (B) LAGE: 2324 .. 2329 und 2331 .. 2336

(ix) MERKMAL:

- (A) NAME/SCHLÜSSEL: PolyA-Segment
- (B) LAGE: 2357 .. 2376

(ix) MERKMAL:

- (A) NAME/SCHLÜSSEL: 3'UTR
- (B) LAGE: 2307 .. 2341 oder 2307 .. 2356

(ix) MERKMAL:

- (A) NAME/SCHLÜSSEL: 5'UTR
- (B) LAGE: 1 .. 23

## Verbindung der Formel II (Neurotrypsin der Maus)

GGACCACACT CGGCGCCGCA	GCC ATG GCG CTC GCC CGC TGC GTG CTG GCT GTG	53
Met Ala Leu Ala Arg Cys Val Leu Ala Val		
-20	-15	
ATT TTA GGG GCA CTG TCT GTA GTG GCC CGC GCT GAT CCG GTC TCG CGC	101	
Ile Leu Gly Ala Leu Ser Val Val Ala Arg Ala Asp Pro Val Ser Arg		
-10	-5	1
TCT CCC CTT CAC CGC CCG CAT CCG TCC CCA CCG CGT TCC CAA CAC GCG	149	
Ser Pro Leu His Arg Pro His Pro Ser Pro Pro Arg Ser Gln His Ala		
10	15	20
CAC TAC CTT CCC AGC TCG CGG CCA CCC AGG ACC CCG CGC TTC CCG	197	
His Tyr Leu Pro Ser Ser Arg Arg Pro Pro Arg Thr Pro Arg Phe Pro		
25	30	35
CTC CCG CTG CGG ATC CCC GCT GCC CAG CGC CCG CAG GTC CTC AGC ACC	245	
Leu Pro Leu Arg Ile Pro Ala Ala Gln Arg Pro Gln Val Leu Ser Thr		
40	45	50
GGG CAC ACG CCC CCG ACG ATT CCA CGC CGC TGC GGG GCA GGA GAG TCG	293	
Gly His Thr Pro Pro Thr Ile Pro Arg Arg Cys Gly Ala Gly Glu Ser		
55	60	65
TGG GGC AAT GCC ACC AAC CTC GGC GTC CCG TGT CTA CAC TGG GAC GAG	341	
Trp Gly Asn Ala Thr Asn Leu Gly Val Pro Cys Leu His Trp Asp Glu		
70	75	80
GTG CCG CCC TTC CTG GAG CGG TCG CCC CCG GCC AGT TGG GCT GAG CTG	389	
Val Pro Pro Phe Leu Glu Arg Ser Pro Pro Ala Ser Trp Ala Glu Leu		
90	95	100
CGA GGG CAG CCG CAC AAC TTC TGC CGG AGC CCG GAT GGC TCG GGC AGA	437	
Arg Gly Gln Pro His Asn Phe Cys Arg Ser Pro Asp Gly Ser Gly Arg		
105	110	115
CCT TGG TGC TTC TAT CGG AAT GCC CAG GGC AAA GTA GAC TGG GGC TAC	485	
Pro Trp Cys Phe Tyr Arg Asn Ala Gln Gly Lys Val Asp Trp Gly Tyr		
120	125	130
TGC GAT TGT GGT CAA GGC CCG GCG TTG CCC GTC ATT CGC CTT GTT GGT	533	
Cys Asp Cys Gly Gln Gly Pro Ala Leu Pro Val Ile Arg Leu Val Gly		
135	140	145
GGG AAC AGT GGG CAT GAA GGT CGA GTG GAG CTG TAC CAC GCT GGC CAG	581	
Gly Asn Ser Gly His Glu Gly Arg Val Glu Leu Tyr His Ala Gly Gln		
150	155	160
TGG GGG ACC ATC TGT GAC GAC CAA TGG GAC AAT GCA GAC GCA GAC GTC	629	
Trp Gly Thr Ile Cys Asp Asp Gln Trp Asp Asn Ala Asp Ala Asp Val		
170	175	180

## DE 698 31 737 T2 2006.06.29

ATC TGT AGG CAG CTG GGG CTC AGT GGC ATT GCC AAA GCA TGG CAT CAG Ile Cys Arg Gln Leu Gly Leu Ser Gly Ile Ala Lys Ala Trp His Gln 185 190 195	677
GCA CAT TTT GGG GAA GGA TCT GGC CCA ATA TTG TTG GAT GAA GTA CGC Ala His Phe Gly Glu Gly Ser Gly Pro Ile Leu Leu Asp Glu Val Arg 200 205 210	725
TGC ACC GGA AAC GAG CTG TCA ATT GAG CAA TGT CCA AAG AGT TCC TGG Cys Thr Gly Asn Glu Leu Ser Ile Glu Gln Cys Pro Lys Ser Ser Trp 215 220 225	773
GGC GAA CAT AAC TGT GGC CAT AAA GAA GAT GCT GGA GTG TCT TGT GTT Gly Glu His Asn Cys Gly His Lys Glu Asp Ala Gly Val Ser Cys Val 230 235 240 245	821
CCT CTA ACA GAT GGT GTC ATC AGA CTG GCA GGA GGA AAA AGT ACC CAT Pro Leu Thr Asp Gly Val Ile Arg Leu Ala Gly Gly Lys Ser Thr His 250 255 260	869
GAA GGT CGC CTG GAG GTC TAC TAC AAG GGG CAG TGG GGG ACA GTC TGT Glu Gly Arg Leu Glu Val Tyr Tyr Lys Gly Gln Trp Gly Thr Val Cys 265 270 275	917
GAT GAT GGC TGG ACT GAG ATG AAC ACA TAC GTG GCT TGT CGA CTG CTG Asp Asp Gly Trp Thr Glu Met Asn Thr Tyr Val Ala Cys Arg Leu Leu 280 285 290	965
GGA TTT AAA TAC GGC AAA CAG TCC TCT GTG AAC CAT TTT GAT GGC AGC Gly Phe Lys Tyr Gly Lys Gln Ser Ser Val Asn His Phe Asp Gly Ser 295 300 305	1013
AAC AGG CCC ATA TGG CTG GAT GAC GTC AGC TGC TCA GGA AAA GAA GTC Asn Arg Pro Ile Trp Leu Asp Asp Val Ser Cys Ser Gly Lys Glu Val 310 315 320 325	1061
AGC TTC ATT CAG TGT TCC AGG AGA CAG TGG GGA AGG CAT GAC TGC AGC Ser Phe Ile Gln Cys Ser Arg Arg Gln Trp Gly Arg His Asp Cys Ser 330 335 340	1109
CAT AGA GAA GAT GTG GGC CTC ACC TGC TAT CCT GAC AGC GAT GGA CAT His Arg Glu Asp Val Gly Leu Thr Cys Tyr Pro Asp Ser Asp Gly His 345 350 355	1157
AGG CTT TCT CCA GGT TTT CCC ATC AGA CTA GTG GAT GGA GAG AAT AAG Arg Leu Ser Pro Gly Phe Pro Ile Arg Leu Val Asp Gly Glu Asn Lys 360 365 370	1205
AAG GAA GGA CGA GTG GAG GTT TTT GTC AAT GGC CAA TGG GGA ACA ATC Lys Glu Gly Arg Val Glu Val Phe Val Asn Gly Gln Trp Gly Thr Ile 375 380 385	1253
TGC GAT GAC GGA TGG ACC GAT AAG CAT GCA GCT GTG ATC TGC CGG CAA Cys Asp Asp Gly Trp Thr Asp Lys His Ala Ala Val Ile Cys Arg Gln 390 395 400 405	1301

## DE 698 31 737 T2 2006.06.29

CTT GGC TAT AAG GGT CCT GCC AGA GCA AGG ACT ATG GCT TAT TTT GGG Leu Gly Tyr Lys Gly Pro Ala Arg Ala Arg Thr Met Ala Tyr Phe Gly 410 415 420	1349
GAA GGA AAA GGC CCC ATC CAC ATG GAT AAT GTG AAG TGC ACA GGA AAT Glu Gly Lys Gly Pro Ile His Met Asp Asn Val Lys Cys Thr Gly Asn 425 430 435	1397
GAG AAG GCC CTG GCT GAC TGT GTC AAA CAA GAC ATT GGA AGG CAC AAC Glu Lys Ala Leu Ala Asp Cys Val Lys Gln Asp Ile Gly Arg His Asn 440 445 450	1445
TGC CGC CAC AGT GAG GAT GCA GGA GTC ATC TGT GAC TAT TTA GAG AAG Cys Arg His Ser Glu Asp Ala Gly Val Ile Cys Asp Tyr Leu Glu Lys 455 460 465	1493
AAA GCA TCA AGT AGT GGT AAT AAA GAG ATG CTC TCA TCT GGA TGT GGA Lys Ala Ser Ser Ser Gly Asn Lys Glu Met Leu Ser Ser Gly Cys Gly 470 475 480 485	1541
CTG AGG TTA CTG CAC CGT CGG CAG AAA CGG ATC ATT GGT GGG AAC AAT Leu Arg Leu Leu His Arg Arg Gln Lys Arg Ile Ile Gly Gly Asn Asn 490 495 500	1589
TCT TTA AGG GGT GCC TGG CCT TGG CAG GCT TCC CTC AGG CTG AGG TCG Ser Leu Arg Gly Ala Trp Pro Trp Gln Ala Ser Leu Arg Leu Arg Ser 505 510 515	1637
GCC CAT GGA GAC GGC AGG CTG CTT TGT GGA GCT ACC CTT CTG AGT AGC Ala His Gly Asp Gly Arg Leu Leu Cys Gly Ala Thr Leu Leu Ser Ser 520 525 530	1685
TGC TGG GTC CTG ACA GCT GCA CAC TGC TTC AAA AGG TAC GGA AAC AAC Cys Trp Val Leu Thr Ala Ala His Cys Phe Lys Arg Tyr Gly Asn Asn 535 540 545	1733
TCG AGG AGC TAT GCA GTT CGA GTT GGG GAT TAT CAT ACT CTG GTC CCA Ser Arg Ser Tyr Ala Val Arg Val Gly Asp Tyr His Thr Leu Val Pro 550 555 560 565	1781
GAG GAG TTT GAA CAA GAA ATA GGG GTT CAA CAG ATT GTG ATT CAC AGG Glu Glu Phe Glu Gln Glu Ile Gly Val Gln Gln Ile Val Ile His Arg 570 575 580	1829
AAC TAC AGG CCA GAC AGA AGC GAC TAT GAC ATT GCC CTG GTT AGA TTG Asn Tyr Arg Pro Asp Arg Ser Asp Tyr Asp Ile Ala Leu Val Arg Leu 585 590 595	1877
CAA GGA CCA GGG GAG CAA TGT GCC AGA CTA AGC ACC CAC GTT TTG CCA Gln Gly Pro Gly Glu Gln Cys Ala Arg Leu Ser Thr His Val Leu Pro 600 605 610	1925
GCC TGT TTA CCT CTA TGG AGA GAG AGG CCA CAG AAA ACA GCC TCC AAC Ala Cys Leu Pro Leu Trp Arg Glu Arg Pro Gln Lys Thr Ala Ser Asn 615 620 625	1973

## DE 698 31 737 T2 2006.06.29

TGT CAC ATA ACA GGA TGG GGA GAC ACA GGT CGT GCC TAC TCA AGA ACT Cys His Ile Thr Gly Trp Gly Asp Thr Gly Arg Ala Tyr Ser Arg Thr 630 635 640 645	2021
CTA CAA CAA GCT GCT GTG CCT CTG TTA CCC AAG AGG TTT TGT AAA GAG Leu Gln Gln Ala Ala Val Pro Leu Leu Pro Lys Arg Phe Cys Lys Glu 650 655 660	2069
AGG TAC AAG GGA CTA TTT ACT GGG AGA ATG CTC TGT GCT GGG AAC CTC Arg Tyr Lys Gly Leu Phe Thr Gly Arg Met Leu Cys Ala Gly Asn Leu 665 670 675	2117
CAA GAA GAC AAC CGT GTG GAC AGC TGC CAG GGA GAC AGT GGA GGA CCA Gln Glu Asp Asn Arg Val Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro 680 685 690	2165
CTC ATG TGT GAA AAG CCT GAT GAG TCC TGG GTT GTG TAT GGG GTG ACT Leu Met Cys Glu Lys Pro Asp Glu Ser Trp Val Val Tyr Gly Val Thr 695 700 705	2213
TCC TGG GGG TAT GGA TGT GGA GTC AAA GAC ACT CCT GGA GTT TAT ACC Ser Trp Gly Tyr Gly Cys Gly Val Lys Asp Thr Pro Gly Val Tyr Thr 710 715 720 725	2261
AGA GTC CCC GCT TTT GTA CCT TGG ATA AAA AGT GTC ACC AGT CTG Arg Val Pro Ala Phe Val Pro Trp Ile Lys Ser Val Thr Ser Leu 730 735 740	2306
TAACTTATGG AAAGCTCAAG AAATAGTAAA ACAGTAACTA TTCAGTCTTC AAAAAAAA AAAAAAAAAA	2366 2376

## SEQUENZPROTOKOLL

## (1) ALLGEMEINE ANGABEN:

- (i) ANMELDER:
  - (A) NAME: Peter Sonderegger
  - (B) STRASSE: Winterthurerstr. 190
  - (C) ORT: Zuerich
  - (E) LAND: Schweiz
  - (F) POSTLEITZAHL: CH-8057
  - (G) TELEFON: +41 1 635 55 41
  - (H) TELEFAX: +41 1 635 68 31
- (ii) BEZEICHNUNG DER ERFINDUNG: Neurotrypsin
- (iii) ANZAHL DER SEQUENZEN: 4
- (iv) COMPUTERLESBARE FASSUNG:
  - (A) DATENTRÄGER: Diskette
  - (B) COMPUTER: IBM-PC-kompatibel
  - (C) BETRIEBSSYSTEM: PC-DOS/MS-DOS
  - (D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPA)
- (vi) DATEN DER VORANMELDUNG:
  - (A) ANMELDENUMMER: CH 1997 0966/97
  - (B) ANMELDETAG: 26-APR-1997

## (2) ANGABEN ZU SEQ ID NO: 1:

- (i) SEQUENZKENNZEICHEN:
  - (A) LÄNGE: 3350 Basenpaare
  - (B) ART: Nucleinsäure
  - (C) STRANGFORM: Einzelstrang
  - (D) TOPOLOGIE: linear
- (ii) ART DES MOLEKÜLS: cDNA zu mRNA
- (iii) HYPOTHETISCH: Nein
- (iv) ANTISENSE: Nein
- (vi) URSPRUNGLICHE HERKUNFT:
  - (A) ORGANISMUS: Homo sapiens
  - (D) ENTWICKLUNGSSTADIUM: fötal
  - (F) GEWEBETYP: Gehirn
- (vii) UNMITTELBARE HERKUNFT:
  - (A) BIBLIOTHEK: cDNA library; cat.no. HL 3003a, Clontech,  
Palo Alto, CA, USA
  - (B) KLON: 3-1, 3-2, 3-6, 3-7, 3-8, 3-10, 3-11, 3-12
- (ix) MERKMAL:
  - (A) NAME/SCHLÜSSEL: Signalpeptid
  - (B) LAGE: 44..103

## (ix) MERKMAL:

- (A) NAME/SCHLÜSSEL: reifes Peptid
- (B) LAGE:104..2668

## (ix) MERKMAL:

- (A) NAME/SCHLÜSSEL: CDS (codierende Sequenz)
- (B) LAGE:44..2668

## (ix) MERKMAL:

- (A) NAME/SCHLÜSSEL: 5'UTR (5'-untranslierte Region)
- (B) LAGE:1..43

## (ix) MERKMAL:

- (A) NAME/SCHLÜSSEL: 3'UTR (3'-untranslierte Region)
- (B) LAGE:2669..3350

## (xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 1:

CGGAAGCTGG GGAGCATGGA CCAGACCCCG CAGCGCTGGC ACC ATG ACG CTC GCC	55
Met Thr Leu Ala	
-20	
CGC TTC GTG CTA GCC CTG ATG TTA GGG GCG CTC CCC GAA GTG GTC GGC	103
Arg Phe Val Leu Ala Leu Met Leu Gly Ala Leu Pro Glu Val Val Gly	
-15	
-10	
-5	
TTT GAT TCT GTC CTC AAT GAT TCC CTC CAC CAC AGC CAC CGC CAT TCG	151
Phe Asp Ser Val Leu Asn Asp Ser Leu His His Ser His Arg His Ser	
1	
5	
10	
15	
CCC CCT GCG GGT CCG CAC TAC CCC TAT TAC CTT CCC ACC CAG CAG CGG	199
Pro Pro Ala Gly Pro His Tyr Pro Tyr Tyr Leu Pro Thr Gln Gln Arg	
20	
25	
30	
CCC CCG ACG ACG CGT CCG CCG CCT CTC CCG CGC TTC CCG CGC CCC	247
Pro Pro Thr Thr Arg Pro Pro Pro Leu Pro Arg Phe Pro Arg Pro	
35	
40	
45	
CCG CGG GCG CTC CCT GCC CAG CGC CCG CAC GCC CTC CAG GCC GGG CAC	295
Pro Arg Ala Leu Pro Ala Gln Arg Pro His Ala Leu Gln Ala Gly His	
50	
55	
60	
ACG CCC CGG CCG CAC CCC TGG GGC TGC CCC GCC GGC GAG CCA TGG GTC	343
Thr Pro Arg Pro His Pro Trp Gly Cys Pro Ala Gly Glu Pro Trp Val	
65	
70	
75	
80	
AGC GTG ACG GAC TTC GGC GCC CCG TGT CTG CGG TGG GCG GAG GTG CCA	391
Ser Val Thr Asp Phe Gly Ala Pro Cys Leu Arg Trp Ala Glu Val Pro	
85	
90	
95	
CCC TTC CTG GAG CGG TCG CCC CCA GCG AGC TGG GCT CAG CTG CGA GGA	439
Pro Phe Leu Glu Arg Ser Pro Pro Ala Ser Trp Ala Gln Leu Arg Gly	
100	
105	
110	

## DE 698 31 737 T2 2006.06.29

CAG CGC CAC AAC TTT TGT CGG AGC CCC GAC GGC GCG GGC AGA CCC TGG Gln Arg His Asn Phe Cys Arg Ser Pro Asp Gly Ala Gly Arg Pro Trp 115 120 125	487
TGT TTC TAC GGA GAC GCC CGT GGC AAG GTG GAC TGG GGC TAC TGC GAC Cys Phe Tyr Gly Asp Ala Arg Gly Lys Val Asp Trp Gly Tyr Cys Asp 130 135 140	535
TGC AGA CAC GGA TCA GTA CGA CTT CGT GGC GGC AAA AAT GAG TTT GAA Cys Arg His Gly Ser Val Arg Leu Arg Gly Gly Lys Asn Glu Phe Glu 145 150 155 160	583
GGC ACA GTG GAA GTA TAT GCA AGT GGA GTT TGG GGC ACT GTC TGT AGC Gly Thr Val Glu Val Tyr Ala Ser Gly Val Trp Gly Thr Val Cys Ser 165 170 175	631
AGC CAC TGG GAT GAT TCT GAT GCA TCA GTC ATT TGT CAC CAG CTG CAG Ser His Trp Asp Asp Ser Asp Ala Ser Val Ile Cys His Gln Leu Gln 180 185 190	679
CTG GGA GGA AAA GGA ATA GCA AAA CAA ACC CCG TTT TCT GGA CTG GGC Leu Gly Gly Lys Gly Ile Ala Lys Gln Thr Pro Phe Ser Gly Leu Gly 195 200 205	727
CTT ATT CCC ATT TAT TGG AGC AAT GTC CGT TGC CGA GGA GAT GAA GAA Leu Ile Pro Ile Tyr Trp Ser Asn Val Arg Cys Arg Gly Asp Glu Glu 210 215 220	775
AAT ATA CTG CTT TGT GAA AAA GAC ATC TGG CAG GGT GGG GTG TGT CCT Asn Ile Leu Leu Cys Glu Lys Asp Ile Trp Gln Gly Gly Val Cys Pro 225 230 235 240	823
CAG AAG ATG GCA GCT GTC ACG TGT AGC TTT TCC CAT GGC CCA ACG Gln Lys Met Ala Ala Val Thr Cys Ser Phe Ser His Gly Pro Thr 245 250 255	871
TTC CCC ATC ATT CGC CTT GCT GGA GGC AGC AGT GTG CAT GAA GGC CGG Phe Pro Ile Ile Arg Leu Ala Gly Gly Ser Ser Val His Glu Gly Arg 260 265 270	919
GTG GAG CTC TAC CAT GCT GGC CAG TGG GGA ACC GTT TGT GAT GAC CAA Val Glu Leu Tyr His Ala Gly Gln Trp Gly Thr Val Cys Asp Asp Gln 275 280 285	967
TGG GAT GAT GCC GAT GCA GAA GTG ATC TGC AGG CAG CTG GGC CTC AGT Trp Asp Asp Ala Asp Ala Glu Val Ile Cys Arg Gln Leu Gly Leu Ser 290 295 300	1015
GGC ATT GCC AAA GCA TGG CAT CAG GCA TAT TTT GGG GAA GGG TCT GGC Gly Ile Ala Lys Ala Trp His Gln Ala Tyr Phe Gly Glu Gly Ser Gly 305 310 315 320	1063
CCA GTT ATG TTG GAT GAA GTA CGC TGC ACT GGG AAT GAG CTT TCA ATT Pro Val Met Leu Asp Glu Val Arg Cys Thr Gly Asn Glu Leu Ser Ile 325 330 335	1111

## DE 698 31 737 T2 2006.06.29

GAG CAG TGT CCA AAG AGC TCC TGG GGA GAG CAT AAC TGT GGC CAT AAA Glu Gln Cys Pro Lys Ser Ser Trp Gly Glu His Asn Cys Gly His Lys 340 345 350	1159
GAA GAT GCT GGA GTG TCC TGT ACC CCT CTA ACA GAT GGG GTC ATC AGA Glu Asp Ala Gly Val Ser Cys Thr Pro Leu Thr Asp Gly Val Ile Arg 355 360 365	1207
CTT GCA GGT GGG AAA GGC AGC CAT GAG GGT CGC TTG GAG GTA TAT TAC Leu Ala Gly Gly Lys Ser His Glu Gly Arg Leu Glu Val Tyr Tyr 370 375 380	1255
AGA GGC CAG TGG GGA ACT GTC TGT GAT GAT GGC TGG ACT GAG CTG AAT Arg Gly Gln Trp Gly Thr Val Cys Asp Asp Gly Trp Thr Glu Leu Asn 385 390 395 400	1303
ACA TAC GTG GTT TGT CGA CAG TTG GGA TTT AAA TAT GGT AAA CAA GCA Thr Tyr Val Val Cys Arg Gln Leu Gly Phe Lys Tyr Gly Lys Gln Ala 405 410 415	1351
TCT GCC AAC CAT TTT GAA GAA AGC ACA GGG CCC ATA TGG TTG GAT GAC Ser Ala Asn His Phe Glu Ser Thr Gly Pro Ile Trp Leu Asp Asp 420 425 430	1399
GTC AGC TGC TCA GGA AAG GAA ACC AGA TTT CTT CAG TGT TCC AGG CGA Val Ser Cys Ser Gly Lys Glu Thr Arg Phe Leu Gin Cys Ser Arg Arg 435 440 445	1447
CAG TGG GGA AGG CAT GAC TGC AGC CAC CGC GAA GAT GTT AGC ATT GCC Gln Trp Gly Arg His Asp Cys Ser His Arg Glu Asp Val Ser Ile Ala 450 455 460	1495
TGC TAC CCT GGC GGC GAG GGA CAC AGG CTC TCT CTG GGT TTT CCT GTC Cys Tyr Pro Gly Gly Glu Gly His Arg Leu Ser Leu Gly Phe Pro Val 465 470 475 480	1543
AGA CTG ATG GAT GGA GAA AAT AAG AAA GAA GGA CGA GTG GAG GTT TTT Arg Leu Met Asp Gly Glu Asn Lys Lys Glu Gly Arg Val Glu Val Phe 485 490 495	1591
ATC AAT GGC CAG TGG GGA ACA ATC TGT GAT GAT GGA TGG ACT GAT AAG Ile Asn Gly Gln Trp Gly Thr Ile Cys Asp Asp Gly Trp Thr Asp Lys 500 505 510	1639
GAT GCA GCT GTG ATC TGT CGT CAG CTT GGC TAC AAG GGT CCT GCC AGA Asp Ala Ala Val Ile Cys Arg Gln Leu Gly Tyr Lys Gly Pro Ala Arg 515 520 525	1687
GCA AGA ACC ATG GCT TAC TTT GGA GAA GGA AAA GGA CCC ATC CAT GTG Ala Arg Thr Met Ala Tyr Phe Gly Glu Gly Lys Gly Pro Ile His Val 530 535 540	1735
GAT AAT GTG AAG TGC ACA GGA AAT GAG AGG TCC TTG GCT GAC TGT ATC Asp Asn Val Lys Cys Thr Gly Asn Glu Arg Ser Leu Ala Asp Cys Ile 545 550 555 560	1783

## DE 698 31 737 T2 2006.06.29

AAG CAA GAT ATT GGA AGA CAC AAC TGC CGC CAC AGT GAA GAT GCA GGA Lys Gln Asp Ile Gly Arg His Asn Cys Arg His Ser Glu Asp Ala Gly 565 570 575	1831
GTT ATT TGT GAT TAT TTT GGC AAG AAG GCC TCA GGT AAC AGT AAT AAA Val Ile Cys Asp Tyr Phe Gly Lys Ala Ser Gly Asn Ser Asn Lys 580 585 590	1879
GAG TCC CTC TCA TCT GTT TGT GGC TTG AGA TTA CTG CAC CGT CGG CAG Glu Ser Leu Ser Ser Val Cys Gly Leu Arg Leu Leu His Arg Arg Gln 595 600 605	1927
AAG CGG ATC ATT GGT GGG AAA AAT TCT TTA AGG GGT GGT TGG CCT TGG Lys Arg Ile Ile Gly Gly Lys Asn Ser Leu Arg Gly Gly Trp Pro Trp 610 615 620	1975
CAG GTT TCC CTC CGG CTG AAG TCA TCC CAT GGA GAT GGC AGG CTC CTC Gln Val Ser Leu Arg Leu Lys Ser Ser His Gly Asp Gly Arg Leu Leu 625 630 635 640	2023
TGC GGG GCT ACG CTC CTG AGT AGC TGC TGG GTC CTC ACA GCA GCA CAC Cys Gly Ala Thr Leu Leu Ser Ser Cys Trp Val Leu Thr Ala Ala His 645 650 655	2071
TGT TTC AAG AGG TAT GGC AAC AGC ACT AGG AGC TAT GCT GTT AGG GTT Cys Phe Lys Arg Tyr Gly Asn Ser Thr Arg Ser Tyr Ala Val Arg Val 660 665 670	2119
GGA GAT TAT CAT ACT CTG GTA CCA GAG GAG TTT GAG GAA GAA ATT GGA Gly Asp Tyr His Thr Leu Val Pro Glu Glu Phe Glu Glu Ile Gly 675 680 685	2167
GTT CAA CAG ATT GTG ATT CAT CGG GAG TAT CGA CCC GAC CGC AGT GAT Val Gln Gln Ile Val Ile His Arg Glu Tyr Arg Pro Asp Arg Ser Asp 690 695 700	2215
TAT GAC ATA GCC CTG GTT AGA TTA CAA GGA CCA GAA GAG CAA TGT GCC Tyr Asp Ile Ala Leu Val Arg Leu Gln Gly Pro Glu Glu Gln Cys Ala 705 710 715 720	2263
AGA TTC AGC AGC CAT GTT TTG CCA GCC TGT TTA CCA CTC TGG AGA GAG Arg Phe Ser Ser His Val Leu Pro Ala Cys Leu Pro Leu Trp Arg Glu 725 730 735	2311
AGG CCA CAG AAA ACA GCA TCC AAC TGT TAC ATA ACA GGA TGG GGT GAC Arg Pro Gln Lys Thr Ala Ser Asn Cys Tyr Ile Thr Gly Trp Gly Asp 740 745 750	2359
ACA GGA CGA GCC TAT TCA AGA ACA CTA CAA CAA GCA GCC ATT CCC TTA Thr Gly Arg Ala Tyr Ser Arg Thr Leu Gln Gln Ala Ala Ile Pro Leu 755 760 765	2407
CTT CCT AAA AGG TTT TGT GAA GAA CGT TAT AAG GGT CGG TTT ACA GGG Leu Pro Lys Arg Phe Cys Glu Glu Arg Tyr Lys Gly Arg Phe Thr Gly 770 775 780	2455

**DE 698 31 737 T2 2006.06.29**

AGA ATG CTT TGT GCT GGA AAC CTC CAT GAA CAC AAA CGC GTG GAC AGC Arg Met Leu Cys Ala Gly Asn Leu His Glu His Lys Arg Val Asp Ser 785 790 795 800	2503
TGC CAG GGA GAC AGC GGA CCA CTC ATG TGT GAA CGG CCC GGA GAG Cys Gln Gly Asp Ser Gly Gly Pro Leu Met Cys Glu Arg Pro Gly Glu 805 810 815	2551
AGC TGG GTG GTG TAT GGG GTG ACC TCC TGG GGG TAT GGC TGT GGA GTC Ser Trp Val Val Tyr Gly Val Thr Ser Trp Gly Tyr Gly Cys Gly Val 820 825 830	2599
AAG GAT TCT CCT GGT GTT TAT ACC AAA GTC TCA GCC TTT GTA CCT TGG Lys Asp Ser Pro Gly Val Tyr Thr Lys Val Ser Ala Phe Val Pro Trp 835 840 845	2647
ATA AAA AGT GTC ACC AAA CTG TAATTCTTCA TGGAAACTTC AAAGCAGCAT Ile Lys Ser Val Thr Lys Leu 850 855	2698
TTAAACAAAT GGAAAACCTT GAACCCCCAC TATTAGCACT CAGCAGAGAT GACAACAAAT	2758
GGCAAGATCT GTTTTGCTT TGTGTTGTGG TAAAAAAATTG TGTACCCCCT GCTGCTTTG	2818
AGAAATTTGT GAACATTTTC AGAGGCCTCA GTGTAGTGG AGTGATAATC CTTAAATGAA	2878
CATTTCTAC CCTAATTCA CTGGAGTGAC TTATTCTAAG CCTCATCTAT CCCCTACCTA	2938
TTTCTCAAAA TCATTCTATG CTGATTTAC AAAAGATCAT TTTTACATTT GAACTGAGAA	2998
CCCCTTTAA TTGAATCAGT GGTGTCTGAA ATCATATTAA ATACCCACAT TTGACATAAA	3058
TGCGGTACCC TTTACTACAC TCATGAGTGG CATATTTATG CTTAGGTCTT TTCAAAAGAC	3118
TTGACAAGAA ATCTTCATAT TCTCTGTAGC CTTTGTCAAG TGAGGAAATC AGTGGTTAAA	3178
GAATTCCACT ATAAACTTT AGGCCTGAAT AGGAGTAGTA AAGCCTCAAG GACATCTGCC	3238
TGTCACAATA TATTCTCAAA GTGATCTGAT ATTTGGAAAC AAGTATCCTT GTTGAGTACC	3298
AAGTGCTACA GAAACCATAA GATAAAAATA CTTTCTACCT ACAGCGTGCC CG	3350

(2) ANGABEN ZU SEQ ID NO: 2:

- (i) SEQUENZKENNZEICHEN:
  - (A) LÄNGE: 875 Aminosäuren
  - (B) ART: Aminosäure
  - (D) TOPOLOGIE: linear

- (ii) ART DES MOLEKÜLS: Protein
- (xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 2:

Met Thr Leu Ala Arg Phe Val Leu Ala Leu Met Leu Gly Ala Leu Pro  
-20 -15 -10 -5

Glu Val Val Gly Phe Asp Ser Val Leu Asn Asp Ser Leu His His Ser  
1 5 10

## DE 698 31 737 T2 2006.06.29

His	Arg	His	Ser	Pro	Pro	Ala	Gly	Pro	His	Tyr	Pro	Tyr	Tyr	Leu	Pro
15							20							25	
Thr	Gln	Gln	Arg	Pro	Pro	Thr	Thr	Arg	Pro	Pro	Pro	Pro	Leu	Pro	Arg
30							35							40	
Phe	Pro	Arg	Pro	Pro	Arg	Ala	Leu	Pro	Ala	Gln	Arg	Pro	His	Ala	Leu
45							50							60	
Gln	Ala	Gly	His	Thr	Pro	Arg	Pro	His	Pro	Trp	Gly	Cys	Pro	Ala	Gly
							65							75	
Glu	Pro	Trp	Val	Ser	Val	Thr	Asp	Phe	Gly	Ala	Pro	Cys	Leu	Arg	Trp
							80							90	
Ala	Glu	Val	Pro	Pro	Phe	Leu	Glu	Arg	Ser	Pro	Pro	Ala	Ser	Trp	Ala
							95							105	
Gln	Leu	Arg	Gly	Gln	Arg	His	Asn	Phe	Cys	Arg	Ser	Pro	Asp	Gly	Ala
							110							120	
Gly	Arg	Pro	Trp	Cys	Phe	Tyr	Gly	Asp	Ala	Arg	Gly	Lys	Val	Asp	Trp
							125							140	
Gly	Tyr	Cys	Asp	Cys	Arg	His	Gly	Ser	Val	Arg	Leu	Arg	Gly	Gly	Lys
							145							155	
Asn	Glu	Phe	Glu	Gly	Thr	Val	Glu	Val	Tyr	Ala	Ser	Gly	Val	Trp	Gly
							160							170	
Thr	Val	Cys	Ser	Ser	His	Trp	Asp	Asp	Ser	Asp	Ala	Ser	Val	Ile	Cys
							175							185	
His	Gln	Leu	Gln	Leu	Gly	Gly	Lys	Gly	Ile	Ala	Lys	Gln	Thr	Pro	Phe
							190							200	
Ser	Gly	Leu	Gly	Leu	Ile	Pro	Ile	Tyr	Trp	Ser	Asn	Val	Arg	Cys	Arg
							205							220	
Gly	Asp	Glu	Glu	Asn	Ile	Leu	Leu	Cys	Glu	Lys	Asp	Ile	Trp	Gln	Gly
							225							235	
Gly	Val	Cys	Pro	Gln	Lys	Met	Ala	Ala	Ala	Val	Thr	Cys	Ser	Phe	Ser
							240							250	
His	Gly	Pro	Thr	Phe	Pro	Ile	Ile	Arg	Leu	Ala	Gly	Gly	Ser	Ser	Val
							255							265	
His	Glu	Gly	Arg	Val	Glu	Leu	Tyr	His	Ala	Gly	Gln	Trp	Gly	Thr	Val
							270							280	
Cys	Asp	Asp	Gln	Trp	Asp	Asp	Ala	Asp	Ala	Glu	Val	Ile	Cys	Arg	Gln
							285							300	
Leu	Gly	Leu	Ser	Gly	Ile	Ala	Lys	Ala	Trp	His	Gln	Ala	Tyr	Phe	Gly
							305							315	

## DE 698 31 737 T2 2006.06.29

Glu Gly Ser Gly Pro Val Met Leu Asp Glu Val Arg Cys Thr Gly Asn  
 320 325 330

Glu Leu Ser Ile Glu Gln Cys Pro Lys Ser Ser Trp Gly Glu His Asn  
 335 340 345

Cys Gly His Lys Glu Asp Ala Gly Val Ser Cys Thr Pro Leu Thr Asp  
 350 355 360

Gly Val Ile Arg Leu Ala Gly Gly Lys Gly Ser His Glu Gly Arg Leu  
 365 370 375 380

Glu Val Tyr Tyr Arg Gly Gln Trp Gly Thr Val Cys Asp Asp Gly Trp  
 385 390 395

Thr Glu Leu Asn Thr Tyr Val Val Cys Arg Gln Leu Gly Phe Lys Tyr  
 400 405 410

Gly Lys Gln Ala Ser Ala Asn His Phe Glu Glu Ser Thr Gly Pro Ile  
 415 420 425

Trp Leu Asp Asp Val Ser Cys Ser Gly Lys Glu Thr Arg Phe Leu Gln  
 430 435 440

Cys Ser Arg Arg Gln Trp Gly Arg His Asp Cys Ser His Arg Glu Asp  
 445 450 455 460

Val Ser Ile Ala Cys Tyr Pro Gly Gly Glu Gly His Arg Leu Ser Leu  
 465 470 475

Gly Phe Pro Val Arg Leu Met Asp Gly Glu Asn Lys Lys Glu Gly Arg  
 480 485 490

Val Glu Val Phe Ile Asn Gly Gln Trp Gly Thr Ile Cys Asp Asp Gly  
 495 500 505

Trp Thr Asp Lys Asp Ala Ala Val Ile Cys Arg Gln Leu Gly Tyr Lys  
 510 515 520

Gly Pro Ala Arg Ala Arg Thr Met Ala Tyr Phe Gly Glu Gly Lys Gly  
 525 530 535 540

Pro Ile His Val Asp Asn Val Lys Cys Thr Gly Asn Glu Arg Ser Leu  
 545 550 555

Ala Asp Cys Ile Lys Gln Asp Ile Gly Arg His Asn Cys Arg His Ser  
 560 565 570

Glu Asp Ala Gly Val Ile Cys Asp Tyr Phe Gly Lys Lys Ala Ser Gly  
 575 580 585

Asn Ser Asn Lys Glu Ser Leu Ser Ser Val Cys Gly Leu Arg Leu Leu  
 590 595 600

His Arg Arg Gln Lys Arg Ile Ile Gly Gly Lys Asn Ser Leu Arg Gly  
 605 610 615 620

Gly Trp Pro Trp Gln Val Ser Leu Arg Leu Lys Ser Ser His Gly Asp  
                   625                       630                      635

Gly Arg Leu Leu Cys Gly Ala Thr Leu Leu Ser Ser Cys Trp Val Leu  
                   640                       645                      650

Thr Ala Ala His Cys Phe Lys Arg Tyr Gly Asn Ser Thr Arg Ser Tyr  
                   655                       660                      665

Ala Val Arg Val Gly Asp Tyr His Thr Leu Val Pro Glu Glu Phe Glu  
                   670                       675                      680

Glu Glu Ile Gly Val Gln Gln Ile Val Ile His Arg Glu Tyr Arg Pro  
                   685                       690                      695                      700

Asp Arg Ser Asp Tyr Asp Ile Ala Leu Val Arg Leu Gln Gly Pro Glu  
                   705                       710                      715

Glu Gln Cys Ala Arg Phe Ser Ser His Val Leu Pro Ala Cys Leu Pro  
                   720                       725                      730

Leu Trp Arg Glu Arg Pro Gln Lys Thr Ala Ser Asn Cys Tyr Ile Thr  
                   735                       740                      745

Gly Trp Gly Asp Thr Gly Arg Ala Tyr Ser Arg Thr Leu Gln Gln Ala  
                   750                       755                      760

Ala Ile Pro Leu Leu Pro Lys Arg Phe Cys Glu Glu Arg Tyr Lys Gly  
                   765                       770                      775                      780

Arg Phe Thr Gly Arg Met Leu Cys Ala Gly Asn Leu His Glu His Lys  
                   785                       790                      795

Arg Val Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro Leu Met Cys Glu  
                   800                       805                      810

Arg Pro Gly Glu Ser Trp Val Val Tyr Gly Val Thr Ser Trp Gly Tyr  
                   815                       820                      825

Gly Cys Gly Val Lys Asp Ser Pro Gly Val Tyr Thr Lys Val Ser Ala  
                   830                       835                      840

Phe Val Pro Trp Ile Lys Ser Val Thr Lys Leu  
                   845                       850                      855

## (2) ANGABEN ZU SEQ ID NO: 3:

- (i) SEQUENZKENNZEICHEN:
  - (A) LÄNGE: 2356 Basenpaare
  - (B) ART: Nucleinsäure
  - (C) STRANGFORM: Einzelstrang
  - (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: cDNA zu mRNA

(iii) HYPOTHETISCH: Nein

(iv) ANTISENSE: Nein

## (vi) URSPRUNGLICHE HERKUNKFT:

- (A) ORGANISMUS: Mus musculus
- (B) STAMM: ICR-ZUR
- (D) ENTWICKLUNGSSTADIUM: Postnatal-Tag 10
- (F) GEWEBETYP: Gehirn
- (G) ZELLTYP: Neuron

## (vii) UNMITTELBARE HERKUNFT:

- (A) BIBLIOTHEK: cDNA library; cat.no. ML 3002a, Clontech,  
Palo Alto, CA, USA
- (B) KLON: cDNA-Klon #25

## (ix) MERKMAL:

- (A) NAME/SCHLÜSSEL: Signalpeptid
- (B) LAGE:24..86

## (ix) MERKMAL:

- (A) NAME/SCHLÜSSEL: reifes Peptid
- (B) LAGE:87..2306

## (ix) MERKMAL:

- (A) NAME/SCHLÜSSEL: CDS (codierende Sequenz)
- (B) LAGE:24..2306

## (ix) MERKMAL:

- (A) NAME/SCHLÜSSEL: PolyA-Bereich
- (B) LAGE: Einer von (2341, 2356)

## (ix) MERKMAL:

- (A) NAME/SCHLÜSSEL: 5'UTR (5'- untranslierte Region)
- (B) LAGE:1..23

## (ix) MERKMAL:

- (A) NAME/SCHLÜSSEL: 3'UTR (3'- untranslierte Region)
- (B) LAGE:2307..Eine von(2341, 2356)

## (xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 3:

GGACCACACT CGGCGCCGCA GCC ATG GCG CTC GCC CGC TGC GTG CTG GCT Met Ala Leu Ala Arg Cys Val Leu Ala -21 -20 -15	50
GTG ATT TTA GGG GCA CTG TCT GTA GTG GCC CGC GCT GAT CCG GTC TCG Val Ile Leu Gly Ala Leu Ser Val Val Ala Arg Ala Asp Pro Val Ser -10 -5 1	98
CGC TCT CCC CTT CAC CGC CCG CAT CCG TCC CCA CCG CGT TCC CAA CAC Arg Ser Pro Leu His Arg Pro His Pro Ser Pro Pro Arg Ser Gln His 5 10 15 20	146
GCG CAC TAC CTT CCC AGC TCG CGG CGG CCA CCC AGG ACC CCG CGC TTC Ala His Tyr Leu Pro Ser Ser Arg Arg Pro Pro Arg Thr Pro Arg Phe 25 30 35	194

## DE 698 31 737 T2 2006.06.29

CCG CTC CCG CTG CGG ATC CCC GCT GCC CAG CGC CCG CAG GTC CTC AGC Pro Leu Pro Leu Arg Ile Pro Ala Ala Gln Arg Pro Gln Val Leu Ser 40 45 50	242
ACC GGG CAC ACG CCC CCG ACG ATT CCA CGC CGC TGC GGG GCA GGA GAG Thr Gly His Thr Pro Pro Thr Ile Pro Arg Arg Cys Gly Ala Gly Glu 55 60 65	290
TCG TGG GGC AAT GCC ACC AAC CTC GGC GTC CCG TGT CTA CAC TGG GAC Ser Trp Gly Asn Ala Thr Asn Leu Gly Val Pro Cys Leu His Trp Asp 70 75 80	338
GAG GTG CCG CCC TTC CTG GAG CGG TCG CCC CCG GCC AGT TGG GCT GAG Glu Val Pro Pro Phe Leu Glu Arg Ser Pro Pro Ala Ser Trp Ala Glu 85 90 95 100	386
CTG CGA GGG CAG CCG CAC AAC TTC TGC CGG AGC CCG GAT GGC TCG GGC Leu Arg Gly Gln Pro His Asn Phe Cys Arg Ser Pro Asp Gly Ser Gly 105 110 115	434
AGA CCT TGG TGC TTC TAT CGG AAT GCC CAG GGC AAA GTA GAC TGG GGC Arg Pro Trp Cys Phe Tyr Arg Asn Ala Gln Gly Lys Val Asp Trp Gly 120 125 130	482
TAC TGC GAT TGT GGT CAA GGC CCG GCG TTG CCC GTC ATT CGC CTT GTT Tyr Cys Asp Cys Gly Gln Gly Pro Ala Leu Pro Val Ile Arg Leu Val 135 140 145	530
GGT GGG AAC AGT GGG CAT GAA GGT CGA GTG GAG CTG TAC CAC GCT GGC Gly Gly Asn Ser Gly His Glu Gly Arg Val Glu Leu Tyr His Ala Gly 150 155 160	578
CAG TGG GGG ACC ATC TGT GAC GAC CAA TGG GAC AAT GCA GAC GCA GAC Gln Trp Gly Thr Ile Cys Asp Asp Gln Trp Asp Asn Ala Asp Ala Asp 165 170 175 180	626
GTC ATC TGT AGG CAG CTG GGG CTC AGT GGC ATT GCC AAA GCA TGG CAT Val Ile Cys Arg Gln Leu Gly Leu Ser Gly Ile Ala Lys Ala Trp His 185 190 195	674
CAG GCA CAT TTT GGG GAA GGA TCT GGC CCA ATA TTG TTG GAT GAA GTA Gln Ala His Phe Gly Glu Gly Ser Gly Pro Ile Leu Leu Asp Glu Val 200 205 210	722
CGC TGC ACC GGA AAC GAG CTG TCA ATT GAG CAA TGT CCA AAG AGT TCC Arg Cys Thr Gly Asn Glu Leu Ser Ile Glu Gln Cys Pro Lys Ser Ser 215 220 225	770
TGG GGC GAA CAT AAC TGT GGC CAT AAA GAA GAT GCT GGA GTG TCT TGT Trp Gly Glu His Asn Cys Gly His Lys Glu Asp Ala Gly Val Ser Cys 230 235 240	818
GTT CCT CTA ACA GAT GGT GTC ATC AGA CTG GCA GGA GGA AAA AGT ACC Val Pro Leu Thr Asp Gly Val Ile Arg Leu Ala Gly Gly Lys Ser Thr 245 250 255 260	866

## DE 698 31 737 T2 2006.06.29

CAT GAA GGT CGC CTG GAG GTC TAC TAC AAG GGG CAG TGG GGG ACA GTC His Glu Gly Arg Leu Glu Val Tyr Tyr Lys Gly Gln Trp Gly Thr Val 265 270 275	914
TGT GAT GAT GGC TGG ACT GAG ATG AAC ACA TAC GTG GCT TGT CGA CTG Cys Asp Asp Gly Trp Thr Glu Met Asn Thr Tyr Val Ala Cys Arg Leu 280 285 290	962
CTG GGA TTT AAA TAC GGC AAA CAG TCC TCT GTG AAC CAT TTT GAT GGC Leu Gly Phe Lys Tyr Gly Lys Gln Ser Ser Val Asn His Phe Asp Gly 295 300 305	1010
AGC AAC AGG CCC ATA TGG CTG GAT GAC GTC AGC TGC TCA GGA AAA GAA Ser Asn Arg Pro Ile Trp Leu Asp Asp Val Ser Cys Ser Gly Lys Glu 310 315 320	1058
GTC AGC TTC ATT CAG TGT TCC AGG AGA CAG TGG GGA AGG CAT GAC TGC Val Ser Phe Ile Gln Cys Ser Arg Arg Gln Trp Gly Arg His Asp Cys 325 330 335 340	1106
AGC CAT AGA GAA GAT GTG GGC CTC ACC TGC TAT CCT GAC AGC GAT GGA Ser His Arg Glu Asp Val Gly Leu Thr Cys Tyr Pro Asp Ser Asp Gly 345 350 355	1154
CAT AGG CTT TCT CCA GGT TTT CCC ATC AGA CTA GTG GAT GGA GAG AAT His Arg Leu Ser Pro Gly Phe Pro Ile Arg Leu Val Asp Gly Glu Asn 360 365 370	1202
AAG AAG GAA GGA CGA GTG GAG GTT TTT GTC AAT GGC CAA TGG GGA ACA Lys Lys Glu Gly Arg Val Glu Val Phe Val Asn Gly Gln Trp Gly Thr 375 380 385	1250
ATC TGC GAT GAC GGA TGG ACC GAT AAG CAT GCA GCT GTG ATC TGC CGG Ile Cys Asp Asp Gly Trp Thr Asp Lys His Ala Ala Val Ile Cys Arg 390 395 400	1298
CAA CTT GGC TAT AAG GGT CCT GCC AGA GCA AGG ACT ATG GCT TAT TTT Gln Leu Gly Tyr Lys Gly Pro Ala Arg Ala Arg Thr Met Ala Tyr Phe 405 410 415 420	1346
GGG GAA GGA AAA GGC CCC ATC CAC ATG GAT AAT GTG AAG TGC ACA GGA Gly Glu Gly Lys Gly Pro Ile His Met Asp Asn Val Lys Cys Thr Gly 425 430 435	1394
AAT GAG AAG GCC CTG GCT GAC TGT GTC AAA CAA GAC ATT GGA AGG CAC Asn Glu Lys Ala Leu Ala Asp Cys Val Lys Gln Asp Ile Gly Arg His 440 445 450	1442
AAC TGC CGC CAC AGT GAG GAT GCA GGA GTC ATC TGT GAC TAT TTA GAG Asn Cys Arg His Ser Glu Asp Ala Gly Val Ile Cys Asp Tyr Leu Glu 455 460 465	1490
AAG AAA GCA TCA AGT AGT GGT AAT AAA GAG ATG CTC TCA TCT GGA TGT Lys Lys Ala Ser Ser Ser Gly Asn Lys Glu Met Leu Ser Ser Gly Cys 470 475 480	1538

## DE 698 31 737 T2 2006.06.29

GGA CTG AGG TTA CTG CAC CGT CGG CAG AAA CGG ATC ATT GGT GGG AAC Gly Leu Arg Leu Leu His Arg Arg Gln Lys Arg Ile Ile Gly Gly Asn 485 490 495 500	1586
AAT TCT TTA AGG GGT GCC TGG CCT TGG CAG GCT TCC CTC AGG CTG AGG Asn Ser Leu Arg Gly Ala Trp Pro Trp Gln Ala Ser Leu Arg Leu Arg 505 510 515	1634
TCG GCC CAT GGA GAC GGC AGG CTG CTT TGT GGA GCT ACC CTT CTG AGT Ser Ala His Gly Asp Gly Arg Leu Leu Cys Gly Ala Thr Leu Leu Ser 520 525 530	1682
AGC TGC TGG GTC CTG ACA GCT GCA CAC TGC TTC AAA AGG TAC GGA AAC Ser Cys Trp Val Leu Thr Ala Ala His Cys Phe Lys Arg Tyr Gly Asn 535 540 545	1730
AAC TCG AGG AGC TAT GCA GTT CGA GTT GGG GAT TAT CAT ACT CTG GTC Asn Ser Arg Ser Tyr Ala Val Arg Val Gly Asp Tyr His Thr Leu Val 550 555 560	1778
CCA GAG GAG TTT GAA CAA GAA ATA GGG GTT CAA CAG ATT GTG ATT CAC Pro Glu Glu Phe Glu Gln Glu Ile Gly Val Gln Gln Ile Val Ile His 565 570 575 580	1826
AGG AAC TAC AGG CCA GAC AGA AGC GAC TAT GAC ATT GCC CTG GTT AGA Arg Asn Tyr Arg Pro Asp Arg Ser Asp Tyr Asp Ile Ala Leu Val Arg 585 590 595	1874
TTG CAA GGA CCA GGG GAG CAA TGT GCC AGA CTA AGC ACC CAC GTT TTG Leu Gln Gly Pro Gly Glu Gln Cys Ala Arg Leu Ser Thr His Val Leu 600 605 610	1922
CCA GCC TGT TTA CCT CTA TGG AGA GAG AGG CCA CAG AAA ACA GCC TCC Pro Ala Cys Leu Pro Leu Trp Arg Glu Arg Pro Gln Lys Thr Ala Ser 615 620 625	1970
AAC TGT CAC ATA ACA GGA TGG GGA GAC ACA GGT CGT GCC TAC TCA AGA Asn Cys His Ile Thr Gly Trp Gly Asp Thr Gly Arg Ala Tyr Ser Arg 630 635 640	2018
ACT CTA CAA CAA GCT GCT GTG CCT CTG TTA CCC AAG AGG TTT TGT AAA Thr Leu Gln Gln Ala Ala Val Pro Leu Leu Pro Lys Arg Phe Cys Lys 645 650 655 660	2066
GAG AGG TAC AAG GGA CTA TTT ACT GGG AGA ATG CTC TGT GCT GGG AAC Glu Arg Tyr Lys Gly Leu Phe Thr Gly Arg Met Leu Cys Ala Gly Asn 665 670 675	2114
CTC CAA GAA GAC AAC CGT GTG GAC AGC TGC CAG GGA GAC AGT GGA GGA Leu Gln Glu Asp Asn Arg Val Asp Ser Cys Gln Gly Asp Ser Gly Gly 680 685 690	2162
CCA CTC ATG TGT GAA AAG CCT GAT GAG TCC TGG GTT GTG TAT GGG GTG Pro Leu Met Cys Glu Lys Pro Asp Glu Ser Trp Val Val Tyr Gly Val 695 700 705	2210

ACT TCC TGG GGG TAT GGA TGT GGA GTC AAA GAC ACT CCT GGA GTT TAT	2258
Thr Ser Trp Gly Tyr Gly Cys Gly Val Lys Asp Thr Pro Gly Val Tyr	
710 715 720	
ACC AGA GTC CCC GCT TTT GTA CCT TGG ATA AAA AGT GTC ACC AGT CTG	2306
Thr Arg Val Pro Ala Phe Val Pro Trp Ile Lys Ser Val Thr Ser Leu	
725 730 735 740	
TAACCTTATGG AAAGCTCAAG AAATAGTAAA ACAGTAACTA TTCAGTCTTC	2356

## (2) ANGABEN ZU SEQ ID NO: 4:

- (i) SEQUENZKENNZEICHEN:  
 (A) LÄNGE: 761 Aminosäuren  
 (B) ART: Aminosäure  
 (D) TOPOLOGIE: linear

- (ii) ART DES MOLEKÜLS: Protein  
 (xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 4:

Met Ala Leu Ala Arg Cys Val Leu Ala Val Ile Leu Gly Ala Leu Ser  
 -21 -20 -15 -10

Val Val Ala Arg Ala Asp Pro Val Ser Arg Ser Pro Leu His Arg Pro  
 -5 1 5 10

His Pro Ser Pro Pro Arg Ser Gln His Ala His Tyr Leu Pro Ser Ser  
 15 20 25

Arg Arg Pro Pro Arg Thr Pro Arg Phe Pro Leu Pro Leu Arg Ile Pro  
 30 35 40

Ala Ala Gln Arg Pro Gln Val Leu Ser Thr Gly His Thr Pro Pro Thr  
 45 50 55

Ile Pro Arg Arg Cys Gly Ala Gly Glu Ser Trp Gly Asn Ala Thr Asn  
 60 65 70 75

Leu Gly Val Pro Cys Leu His Trp Asp Glu Val Pro Pro Phe Leu Glu  
 80 85 90

Arg Ser Pro Pro Ala Ser Trp Ala Glu Leu Arg Gly Gln Pro His Asn  
 95 100 105

Phe Cys Arg Ser Pro Asp Gly Ser Gly Arg Pro Trp Cys Phe Tyr Arg  
 110 115 120

Asn Ala Gln Gly Lys Val Asp Trp Gly Tyr Cys Asp Cys Gly Gln Gly  
 125 130 135

Pro Ala Leu Pro Val Ile Arg Leu Val Gly Gly Asn Ser Gly His Glu  
 140 145 150 155

Gly Arg Val Glu Leu Tyr His Ala Gly Gln Trp Gly Thr Ile Cys Asp  
 160 165 170

## DE 698 31 737 T2 2006.06.29

Asp Gln Trp Asp Asn Ala Asp Ala Asp Val Ile Cys Arg Gln Leu Gly  
 175 180 185

Leu Ser Gly Ile Ala Lys Ala Trp His Gln Ala His Phe Gly Glu Gly  
 190 195 200

Ser Gly Pro Ile Leu Leu Asp Glu Val Arg Cys Thr Gly Asn Glu Leu  
 205 210 215

Ser Ile Glu Gln Cys Pro Lys Ser Ser Trp Gly Glu His Asn Cys Gly  
 220 225 230 235

His Lys Glu Asp Ala Gly Val Ser Cys Val Pro Leu Thr Asp Gly Val  
 240 245 250

Ile Arg Leu Ala Gly Gly Lys Ser Thr His Glu Gly Arg Leu Glu Val  
 255 260 265

Tyr Tyr Lys Gly Gln Trp Gly Thr Val Cys Asp Asp Gly Trp Thr Glu  
 270 275 280

Met Asn Thr Tyr Val Ala Cys Arg Leu Leu Gly Phe Lys Tyr Gly Lys  
 285 290 295

Gln Ser Ser Val Asn His Phe Asp Gly Ser Asn Arg Pro Ile Trp Leu  
 300 305 310 315

Asp Asp Val Ser Cys Ser Gly Lys Glu Val Ser Phe Ile Gln Cys Ser  
 320 325 330

Arg Arg Gln Trp Gly Arg His Asp Cys Ser His Arg Glu Asp Val Gly  
 335 340 345

Leu Thr Cys Tyr Pro Asp Ser Asp Gly His Arg Leu Ser Pro Gly Phe  
 350 355 360

Pro Ile Arg Leu Val Asp Gly Glu Asn Lys Lys Glu Gly Arg Val Glu  
 365 370 375

Val Phe Val Asn Gly Gln Trp Gly Thr Ile Cys Asp Asp Gly Trp Thr  
 380 385 390 395

Asp Lys His Ala Ala Val Ile Cys Arg Gln Leu Gly Tyr Lys Gly Pro  
 400 405 410

Ala Arg Ala Arg Thr Met Ala Tyr Phe Gly Glu Gly Lys Gly Pro Ile  
 415 420 425

His Met Asp Asn Val Lys Cys Thr Gly Asn Glu Lys Ala Leu Ala Asp  
 430 435 440

Cys Val Lys Gln Asp Ile Gly Arg His Asn Cys Arg His Ser Glu Asp  
 445 450 455

Ala Gly Val Ile Cys Asp Tyr Leu Glu Lys Lys Ala Ser Ser Ser Gly  
 460 465 470 475

Asn Lys Glu Met Leu Ser Ser Gly Cys Gly Leu Arg Leu Leu His Arg  
                   480                     485                  490  
  
 Arg Gln Lys Arg Ile Ile Gly Gly Asn Asn Ser Leu Arg Gly Ala Trp  
                   495                     500                  505  
  
 Pro Trp Gln Ala Ser Leu Arg Leu Arg Ser Ala His Gly Asp Gly Arg  
                   510                     515                  520  
  
 Leu Leu Cys Gly Ala Thr Leu Leu Ser Ser Cys Trp Val Leu Thr Ala  
                   525                     530                  535  
  
 Ala His Cys Phe Lys Arg Tyr Gly Asn Asn Ser Arg Ser Tyr Ala Val  
                   540                     545                  550                  555  
  
 Arg Val Gly Asp Tyr His Thr Leu Val Pro Glu Glu Phe Glu Gln Glu  
                   560                     565                  570  
  
 Ile Gly Val Gln Gln Ile Val Ile His Arg Asn Tyr Arg Pro Asp Arg  
                   575                     580                  585  
  
 Ser Asp Tyr Asp Ile Ala Leu Val Arg Leu Gln Gly Pro Gly Glu Gln  
                   590                     595                  600  
  
 Cys Ala Arg Leu Ser Thr His Val Leu Pro Ala Cys Leu Pro Leu Trp  
                   605                     610                  615  
  
 Arg Glu Arg Pro Gln Lys Thr Ala Ser Asn Cys His Ile Thr Gly Trp  
                   620                     625                  630                  635  
  
 Gly Asp Thr Gly Arg Ala Tyr Ser Arg Thr Leu Gln Gln Ala Ala Val  
                   640                     645                  650  
  
 Pro Leu Leu Pro Lys Arg Phe Cys Lys Glu Arg Tyr Lys Gly Leu Phe  
                   655                     660                  665  
  
 Thr Gly Arg Met Leu Cys Ala Gly Asn Leu Gln Glu Asp Asn Arg Val  
                   670                     675                  680  
  
 Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro Leu Met Cys Glu Lys Pro  
                   685                     690                  695  
  
 Asp Glu Ser Trp Val Val Tyr Gly Val Thr Ser Trp Gly Tyr Gly Cys  
                   700                     705                  710                  715  
  
 Gly Val Lys Asp Thr Pro Gly Val Tyr Thr Arg Val Pro Ala Phe Val  
                   720                     725                  730  
  
 Pro Trp Ile Lys Ser Val Thr Ser Leu  
                   735                     740

### Patentansprüche

1. Neurotrypsine der Formeln I oder II
- I: Neurotrypsin des Menschen  
 II: Neurotrypsin der Maus
2. Nukleotidsequenzen, welche die Neurotrypsine gemäss Anspruch 1 kodieren, dadurch gekennzeichnet, dass sie die kodierenden Nukleotidsequenzen der Verbindungen der Formeln I oder II umfassen.
  3. Verwendung der kodierenden Nukleotidsequenzen der Verbindungen der Formeln I oder II für die Herstellung von rekombinanten Proteinen.
  4. Verwendung der Proteine mit den kodierten Aminosäuresequenzen der Verbindungen der Formeln I

oder II als Ziele für die Entwicklung von pharmazeutischen Arzneimitteln, beispielsweise für die Hemmung oder die Förderung der katalytischen Aktivität der kodierten Proteine der Formeln I oder II.

5. Verwendung der Proteine mit den kodierten Aminosäuresequenzen der Verbindungen der Formeln I oder II für die Bestimmung der Raumstruktur, beispielsweise für die Bestimmung der Raumstruktur mittels Kristallographie oder mittels kernmagnetischer Resonanzspektroskopie (NMR).

6. Verwendung der kodierten Aminosäuresequenzen der Verbindungen der Formeln I oder II für die Voraussage der Proteinstruktur mittels computerisierter Proteinstruktur-Voraussagungs-Methoden.

7. Verwendung der kodierten Aminosäuresequenzen der Verbindungen der Formeln I oder II als Antigene für die Herstellung von Antikörpern, wie beispielsweise Antikörper, welche die Protease-Funktion hemmen oder fördern, oder Antikörper, welche für immunohistochimische Studien verwendet werden können.

8. Verwendung der kodierenden Nukleotidsequenzen der Verbindungen der Formeln I oder II für die Herstellung von nicht-menschlichen transgenen Tieren, wie beispielsweise transgene Mäuse.

9. Verwendung der kodierenden Nukleotidsequenzen der Verbindungen der Formeln I oder II für die Inaktivierung oder die Mutation des entsprechenden Gens mittels *in vitro* „Gene Targeting“-Techniken, wie beispielsweise die Eliminierung des Gens in der Maus durch homologe Rekombination.

Es folgt kein Blatt Zeichnungen