

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

(19) Weltorganisation für geistiges Eigentum  
Internationales Büro



(43) Internationales Veröffentlichungsdatum  
28. Februar 2002 (28.02.2002)

PCT

(10) Internationale Veröffentlichungsnummer  
WO 02/16424 A2

- (51) Internationale Patentklassifikation<sup>7</sup>: C07K 14/435 (74) **Anwalt:** ACKERMANN, Joachim; Postfach 11 13 26, 60046 Frankfurt am Main (DE).
- (21) Internationales Aktenzeichen: PCT/EP01/09565 (81) **Bestimmungsstaaten (national):** AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, 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, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.
- (22) Internationales Anmeldedatum:  
20. August 2001 (20.08.2001)
- (25) Einreichungssprache: Deutsch
- (26) Veröffentlichungssprache: Deutsch
- (30) Angaben zur Priorität:  
100 41 126.6 22. August 2000 (22.08.2000) DE
- (71) **Anmelder (für alle Bestimmungsstaaten mit Ausnahme von US):** XZILLION GMBH & CO KG [DE/DE]; 65926 Frankfurt (DE).
- (72) **Erfinder; und**
- (75) **Erfinder/Anmelder (nur für US):** KAPPEL, Andreas [DE/DE]; Wiesbadener Strasse 25, 61462 Königstein (DE). WINDHAB, Norbert [DE/DE]; Hahneckstrasse 36, 65719 Hofheim (DE). WAGNER, Thomas [DE/DE]; Sonnenbühlstrasse 73, 78464 Konstanz (DE). KIENLE, Stefan [DE/DE]; Am Buchwald 5, 60385 Frankfurt (DE). KUHN, Karsten [DE/DE]; Querstrasse 29, 44139 Dortmund (DE).
- (84) **Bestimmungsstaaten (regional):** ARIPO-Patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), eurasisches Patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), europäisches Patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI-Patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).
- Veröffentlicht:**  
— ohne internationalen Recherchenbericht und erneut zu veröffentlichen nach Erhalt des Berichts
- Zur Erklärung der Zweibuchstaben-Codes und der anderen Abkürzungen wird auf die Erklärungen ("Guidance Notes on Codes and Abbreviations") am Anfang jeder regulären Ausgabe der PCT-Gazette verwiesen.



WO 02/16424 A2

(54) **Title:** DNA BINDING PEPTIDE DOMAINS AND A METHOD FOR PROVIDING DOMAINS OF THIS TYPE

(54) **Bezeichnung:** DNA-BINDEnde PEPTIDDOMÄNEN UND EIN VERFAHREN ZUR BEREITSTELLUNG SOLCHER DOMÄNEN

(57) **Abstract:** The invention relates to peptidic domains, which are made available in a slightly synthetic manner, and which can specifically identify and bind nucleic acid sequences. The invention also relates to a method for discovering and providing specifically DNA-binding peptide domains, and to biomorph factors derived therefrom, in particular, transcription factors and repressors.

(57) **Zusammenfassung:** Die vorliegende Erfindung betrifft leicht synthetisch zugängliche peptidische Domänen, die Nukleinsäuresequenzen spezifisch erkennen und binden können, sowie ein Verfahren zur Auffindung und Bereitstellung spezifisch DNA-binder Peptidomänen, und davon abgeleiteter biomorpher Faktoren, insbesondere Transkriptionsfaktoren und Repressoren.

## **DNA-bindende Peptiddomänen und ein Verfahren zur Bereitstellung solcher Domänen.**

### Beschreibung

Die vorliegende Erfindung betrifft leicht synthetisch zugängliche peptidische Domänen, die Nukleinsäuresequenzen spezifisch erkennen und binden können, sowie ein Verfahren zur Auffindung und Bereitstellung spezifisch DNA-bindender Peptiddomänen, und davon abgeleiteter biomorpher Faktoren, insbesondere Transkriptionsfaktoren und Repressoren.

Die Transkription von DNA in RNA ist der erste Schritt der Genexpression. Die Stärke der Transkription eines Gens und damit die Menge der gebildeten RNA wird durch mit den Genen assoziierte regulatorische Promotorelemente- und Enhancerelemente determiniert, an welche Transkriptionsfaktoren binden. Diese Proteine binden mittels einer DNA-bindenden Domäne sequenzspezifisch an kurze DNA-Abschnitte in den regulatorischen Elementen der Gene. Neben der DNA-bindenden Domäne enthalten Transkriptionsfaktoren eine oder mehrere Aktivator- oder Repressordomänen. Aktivator-domänen verstärken die Transkription des dem vom Transkriptionsfaktor gebundenen regulatorischen Element zugehörigen Gens, während Repressordomänen die Transkription des Gens abschwächen. Beides geschieht durch die Rekrutierung von weiteren Proteinen durch die Aktivator- oder Repressordomänen über Protein-Protein-Interaktionen. Die DNA-Sequenz der regulatorischen Elemente eines Gens und die spezifisch an diese Sequenz bindenden Proteindomänen wirken daher als komplementäre Adressen, die im Zusammenspiel Aktivator- und Repressordomänen lokalisieren, welche letztendlich die Transkriptionsstärke eines Gens festlegen.

Da viele Krankheiten wie z. B. Krebs durch die deregulierte Expression bestimmter Gene verursacht werden, wurden verschiedene Methoden entwickelt, durch welche die Transkription dieser Gene beeinflusst und somit vom pathologischen in den

physiologischen Zustand überführt werden kann. Neben Substanzen, welche spätere Stufen der Genexpression inhibieren wie z. B. Antisense-RNA oder Ribozyme, wurden auch Methoden und Substanzen entwickelt, die mit den frühen Stufen der Genexpression, d. h. der Bindung von Transkriptionsfaktoren und der Transkription interferieren. So wurden z. B. Substanzen aus der Stoffklasse der Polyamide entwickelt, welche an regulatorische Elemente ausgewählter Gene binden können. Polyamide binden an die kleine Furche der DNA und verhindern somit die Bindung der meist aktivierenden Transkriptionsfaktoren an die regulatorischen Elemente (J.M. Gottesfeld et al., Nature, 1997, 387, 202-205) Damit können Polyamide indirekt die Transkription des jeweiligen Gens vermindern. Die DNA-Bindung der Polyamid ist allerdings wenig spezifisch, so daß ein therapeutischer Einsatz dieser Substanzen nicht in Frage kommt. Darüber hinaus kann mit Polyamiden keine Erhöhung der Transkriptionsrate erreicht werden.

Bei einer weiteren Methode werden relativ große Mengen kurzer doppelsträngiger DNA-Moleküle in Zellen eingebracht, welche eine Bindungsstelle für einen Transkriptionsfaktor enthalten, der für die deregulierte Transkription des jeweiligen Gens verantwortlich ist. Bei dieser "decoy" Strategie wird der ausgewählte Transkriptionsfaktor kompetitiv abgefangen und kann somit nicht mehr an die Bindungsstellen in den regulatorischen Elemente seiner Zielgenene binden, wodurch die Transkription dieser Gene indirekt inhibiert wird (Übersicht in Suda et al., Endocr. Rev., 1999, 20, 345-357; S. Yla-Herttua und J.F. Martin, The Lancet 355, 213-222, 2000).

In den vergangenen Jahren wurden synthetische Proteine und Peptide entwickelt, die DNA mittels anderer struktureller Motive wie z.B. dem Helix-turn-Helix, dem Basic Leucin Zipper oder dem  $\alpha$ -Helix Motiv binden. Diese Peptide und Proteine leiten sich von natürlich vorkommenden Transkriptionsfaktoren ab und behalten die DNA-Bindenspezifität des jeweiligen Faktors. Diese Peptide haben gewöhnlich eine sehr schwache Affinität zu DNA und können daher unter physiologischen Bedingungen nicht an das jeweilige DNA-Element binden. In manchen Fällen wird jedoch durch Mutagenese und chemische Modifikation eine Verstärkung der Bindung gegenüber dem unmodifizierten Protein oder Peptid erreicht (Z. Shang et al., Proc. Natl. Acad. Sci. USA, 1994, 91, 8373-8377). Solche künstlich optimierten

DNA-bindenden Proteine sind allerdings nur zur Beeinflussung der Regulation der Gene geeignet, die auch auf den Transkriptionsfaktor ansprechen aus dem sie hervorgegangen sind.

Eine relativ neue Methode zur Regulation der Transkription stellen synthetische Transkriptionsfaktoren aus der Klasse der Zinkfingerproteine vom Cys<sub>2</sub>-His<sub>2</sub> dar (D.J. Segal, B. Dreier, R.R. Beerli, C. F. Barbas III, Proc. Natl. Acad. Sci. USA, 1999, 96, 2758-2763). Diese Faktoren binden an die große Furche der DNA mittels einer fingerartigen Struktur, in welcher eine kurze  $\alpha$ -Helix und ein antiparalleles  $\beta$ -Faltblatt durch ein Zinkatom miteinander komplexiert werden (D. Rhodes, A. Klug, Sci. Am., 1993, 268, 32-39). Die DNA-bindende Aktivität der Zinkfingerproteine ist modular aufgebaut. Individuelle Finger, welche jeweils an eine unterschiedliche Trinukleotidsequenz (gewöhnlich 5'-GNN-3', wobei N jedes der 4 Nukleotide sein kann) spezifisch binden, können in einem größeren Polypeptid miteinander kombiniert werden, so daß auch längere, für das zu regulierende Zielgen einzigartige DNA-Sequenzen spezifisch gebunden werden können (Q. Liu, D.J. Segal, J.B. Ghiara, C.F. Barbas III, Proc. Natl. Acad. Sci. USA, 1997, 94, 5525-5530; R.R. Beerli, D.J. Segal, B. Dreier, C.F. Barbas III, Proc. Natl. Acad. Sci. USA, 1998, 95, 14628-14633). Diese Multifingerproteine werden gewöhnlich in einer Phagenbibliothek kodiert, weshalb die an eine spezifische DNA-Sequenz des Zielgens bindenden Mitglieder der Bibliothek mit der Phage Display Methodik isoliert werden können. Anschließend wird die cDNA des DNA-bindenden Polypeptids mit der cDNA einer Aktivator- oder Repressordomäne fusioniert; das resultierende Fusionsprotein stellt damit einen spezifischen Regulator des Zielgens dar (Übersicht der Methodik in D.J. Segal und C.F. Barbas III, Curr. Opin Chem. Biol., 2000, 4, 34-39; A. Klug, J.Mol. Biol., 1999, 293, 215-218). Kürzlich konnte erstmals gezeigt werden, daß diese Fusionsproteine prinzipiell die Transkription von chromosomal lokalisierten Genen regulieren können (R.R. Beerli, B. Dreier, C.F. Barbas, Proc. Natl. Acad. Sci. USA, 2000, 97, 1495-1500). Nachteilig ist jedoch, daß relativ große Proteine wie die Zinkfingerproteine nur schwer durch Zellmembranen transportiert werden können und dabei häufige ihre ursprüngliche Konformation verlieren. Im Gegensatz zu Polyamiden und Oligonukleotiden sind die C<sub>2</sub>H<sub>2</sub> Zinkfingerproteine wegen ihrer Größe darüber hinaus kaum synthetisch zugänglich, so daß die Proteine

nur in relativ kleinen Mengen durch Überexpression in Mikroorganismen erzeugt werden können. Weitehin ist das Auffinden spezifisch bindender künstlicher Zinkfingerproteine sehr aufwendig, da aufgrund der potentiellen Proteinstrukturvielfalt große Proteinbanken durchsucht werden müssen.

Vor diesem Hintergrund stellt sich die Aufgabe kurze peptidische biomorphe Faktoren bereitzustellen, die Nukleinsäuresequenzen spezifisch erkennen und binden können sowie ein Verfahren zur Auffindung solcher biomorpher Faktoren bereitzustellen.

Die Aufgabe wird durch Peptide enthaltend folgende Aminosäuresequenz oder einer strukturellen Variante davon gelöst:

**DPAALKRARXTEXXRRXRARKLQ (Seq. ID No. 1)**

Bevorzugt sind jeweils zwei beliebige Peptide ausgewählt aus dieser Gruppe, die über einen geeigneten Linker ein homo- oder heterodimeres Molekül oder eine Mischung aus homo- und heterodimeren Molekülen bilden.

Als strukturelle Varianten werden Aminosäuresequenzen verstanden die mindestens eine Homologie von 75%, bevorzugt eine Homologie von über 90% zur **Seq. ID No. 1** aufweisen. Bevorzugt sind funktionelle Varianten die eine sequenzspezifische DNA-Bindung eingehen können.

Als Linker kommen alle Verbindungen in Betracht die eine Dimerisierung zweier Peptide aus der Gruppe mit der **Seq. ID No. 1** oder deren strukturelle Varianten in Lösung gewährleisten. Die Linker können freie organische oder anorganische Substanzen sein, die den erfindungsgemäßen Peptiden zugesetzt werden können. Beispiele dafür sind z. B. Komplexbildner, die mit den freien Amino- oder Carboxytermini von Peptiden koordinative Bindungen eingehen können. Die Linker und Peptide können darüber hinaus durch chemische Modifikation mindestens eines der Peptide kovalent verknüpft sein. Peptidische Linker können auch direkt mit mindestens einem der Peptide fusioniert sein.

Bevorzugt sind Fusionsproteine die neben einer DNA-bindenden Domäne gemäß **Seq ID No. 1** oder einer strukturellen Variante eine Leucin-Zipper-Aminosäuredomäne besitzen. **Seq. ID No. 2** gibt die Aminosäuresequenz eines solchen Fusionsproteins beispielhaft wieder.

DPAALKRARXTEXXRRXRARKLQLEDKVEELLSKNYHLENEVARLKKLVGGER

### **Seq. ID No. 2**

Die erfindungsgemäßen Peptide können über dem Fachmann bekannte Verfahren synthetisch hergestellt werden (z. B. nach der Merryfield-Synthese), die Peptide sind auch mit molekularbiologischen Verfahren, wie der Expression in Zellkultur oder in Fermentern erhältlich.

Es hat sich überraschend gezeigt, daß die kurzen Dimere die Peptide mit **Seq. ID No. 1** oder deren Varianten enthalten spezifisch DNA-Sequenzen erkennen und somit die Aktivität von Genen in lebenden Zellen regulieren können. Durch die Kürze der Aminosäuresequenz sind die Peptide leicht synthetisch zugänglich, darüber hinaus ist aufgrund der Kürze der Peptide der Transport durch Zellmembranen erleichtert. Durch die hohe Stabilität der Konformation der erfindungsgemäßen Peptide ist gewährleistet, daß deren Funktion, also die Fähigkeit sequenzspezifisch DNA zu binden beim Transport durch Zellmembranen erhalten bleibt. Durch die unterschiedliche Struktur der Peptide gegenüber anderen künstlichen Transkriptionsfaktoren können zudem bestimmte DNA-Sequenzmotive besser gebunden werden.

Aus diesen Gründen sind die erfindungsgemäßen Dimere als Modul zur Konstruktion von künstlichen aktivierenden oder reprimierenden Transkriptionsfaktoren geeignet. Dazu werden die DNA- Bindungsdomänen mit bekannten Kerntransportsignaldomänen, Transkriptionsaktivierungsdomänen oder Transkriptionseinhibierungsdomänen wie z.B. dem Kernlokalisierungssignal des SV40 T-Antigens, der Transkriptionsaktivator-domäne des viralen VP16 Proteins, oder der KRAB-Repressordomäne fusioniert. Die so erzeugten künstlichen biomorphen Faktoren können zur Transkriptionskontrolle und damit auch zur Regulierung der Expression von Genen verwendet werden. Somit ermöglichen die erfindungsgemäßen künstlichen biomorphen Faktoren die Untersuchung von

phänotypischen Effekten der Expressionsänderung einzelner Gene in vivo. Vor allem die Verwendung solcher biomorpher Faktoren zur Bekämpfung von Krankheiten, die auf einer fehlregulierten Transkription von Genen beruhen, wie z. B. Krebs, Entzündungsreaktionen oder Suchtkrankheiten ist von großem Interesse. Bei der Verwendung der biomorphen Faktoren als Biopharmaka ist weiterhin deren leichte physiologische Abbaubarkeit vorteilhaft.

Ein weiterer Gegenstand der vorliegenden Erfindung ist daher auch ein Arzneimittel, das die erfindungsgemäßen Peptidmonomere oder -dimere und gegebenenfalls geeignete Zusatz- oder Hilfsstoffe enthält, sowie ein Verfahren zur Herstellung eines Arzneimittels zur Behandlung von Krankheiten die auf fehlregulierte oder auf die Expression von mutierten Genen zurückgehen, bei dem ein erfindungsgemäßes Peptidmonomer oder -dimer mit pharmazeutisch annehmbaren Zusatz- und/oder Hilfsstoffen formuliert wird.

Ein weiterer Gegenstand der Erfindung sind die die erfindungsgemäßen Peptide gemäß **Seq ID No. 1** oder deren strukturelle Varianten codierenden DNA-Sequenzen (**Seq. ID No. 3**) bzw. Nukleinsäuren die diese Sequenzen umfassen.

GATCCTGCTGCTCTAAAACGTGCTAGANNCACTGAANNNNNNAGGCGTNNNCG  
TGCGAGAAAGTTGCAA

(**Seq. ID No. 3**)

Als strukturelle Variante werden Nukleinsäuren mit abweichender Sequenz verstanden, die Proteine gemäß **Seq ID No. 1** oder deren strukturelle Varianten codieren. Bevorzugt sind Nukleinsäuren, die funktionelle Varianten dieser Proteine codieren.

Bevorzugt sind DNA-Sequenzen, die für Fusionsproteine aus den erfindungsgemäßen DNA-bindenden Domänen und einer Leucin-Zipper-Domäne codieren (**Seq. ID. No. 4**).

GATCCTGCTGCTCTAAAACGTGCTAGANNCAGTGAANNNNNNAGGCGTNNNCG  
TGCGAGAAAGTTGCAACTTGAAGACAAGGTTGAAGAATTGCTTTCGAAAAATTA  
TCACTTGGAAAATGAGGTTGCCAGATTAAAGAAATTAGTTGGCGAACGC

**(Seq. ID No. 4)**

Darüber hinaus können die erfindungsgemäßen Nukleinsäuren weitere codierende Bereiche wie z. B. für Kerntansprotsignaldomänen (Kernlokalisierungssignale), Transkriptionsaktivierungsdomänen (z. B. aus Transkriptionsfaktoren) oder Transkriptionseinhibierungsdomänen enthalten.

Die erfindungsgemäßen Nukleinsäuren können beispielsweise chemisch anhand der in **SEQ ID No 3** oder **4** offenbarten Sequenzen oder anhand der in **SEQ ID No 1** oder **2** oder deren strukturelle Varianten offenbarten Peptidsequenzen unter Heranziehen des genetischen Codes z. B. nach der Phosphotriester-Methode synthetisiert werden (siehe z.B. Uhlman, E. & Peyman, A. (1990) Chemical Reviews, 90, 543, No. 4).

Die erfindungsgemäßen Desoxyribonukleinsäuren können zu gentherapeutischen Zwecken aber auch zur Untersuchung von phänotypischen Effekten durch Änderung der Expression einzelner Gene in vivo eingesetzt werden. Dazu müssen DNA-Fragmente in geeignete Expressionsvektoren eingebracht sein.

Als Vektoren können je nach Expressionsorganismus alle geeigneten pro- oder eukaryontischen Expressionsvektoren verwendet werden. Bevorzugt sind kommerziell erhältliche Expressionsvektoren, die für die Wirtszelle geeignete regulatorische Sequenzen, wie z. B. den trp-Promotor für die Expression in e. coli oder den ADH-2-Promotor für die Expression in Hefen, den Baculovirus-Polyhedrin-Promotor für die Expression in Insektenzellen.

Beispiele von gentherapeutisch wirksamen Vektoren sind Virusvektoren, vorzugsweise Adenovirusvektoren, insbesondere replikationsdefiziente Adenovirusvektoren, oder Adenoassoziierte Virusvektoren, z.B. ein Adenoassoziiierter Virusvektor, der ausschließlich aus zwei insertierten terminalen Wiederholungssequenzen (ITR) besteht.

Geeignete Adenovirusvektoren sind beispielsweise in McGrory, W.J. et al. (1988) Virol. 163, 614; Gluzman, Y. et al. (1982) in "Eukaryotic Viral Vectors" (Gluzman, Y. ed.) 187, Cold Spring Harbor Press, Cold Spring Harbor, New York; Chroboczek, J. et al. (1992) Virol. 186, 280; Karlsson, S. et al. (1986) EMBO J. 5, 2377 oder WO95/00655 beschrieben.

Geeignete Adeno-assoziierte Virusvektoren sind beispielsweise in Muzyczka, N. (1992) Curr. Top. Microbiol. Immunol. 158, 97; WO95/23867; Samulski, R.J. (1989) J. Virol, 63, 3822; WO95/23867; Chiorini, J.A. et al. (1995) Human Gene Therapy 6, 1531 oder Kotin, R.M. (1994) Human Gene Therapy 5, 793 beschrieben.

Gentherapeutisch wirksame Vektoren lassen sich auch dadurch erhalten, daß man die erfindungsgemäße Nukleinsäure mit Liposomen komplexiert. Hierzu eignen sich Lipidmischungen wie bei Felgner, P.L. et al. (1987) Proc. Natl. Acad. Sci. USA 84, 7413; Behr, J.P. et al. (1989) Proc. Natl. Acad. Sci. USA 86, 6982; Felgner, J.H. et al. (1994) J. Biol. Chem. 269, 2550 oder Gao, X. & Huang, L. (1991) Biochim. Biophys. Acta 1189, 195 beschrieben. Bei der Herstellung der Liposomen wird die DNA ionisch auf der Oberfläche der Liposomen gebunden, und zwar in einem solchen Verhältnis, daß eine positive Nettoladung verbleibt und die DNA vollständig von den Liposomen komplexiert wird.

In einer weiteren Ausführungsform sind die erfindungsgemäßen Nukleinsäuren daher in einem Expressionsvektor, vorzugsweise in einem Expressionsvektor der zur gentherapeutischen Verwendung oder zur Herstellung von transgenen Mikroorganismen, Pflanzen oder Tieren geeignet ist.

Ein weiterer Gegenstand der vorliegenden Erfindung ist daher auch ein Arzneimittel, das unter Verwendung der erfindungsgemäßen Nukleinsäuren gegebenenfalls unter Verwendung weiterer Zusatz- oder Hilfsstoffe zur gentherapeutische Anwendung bei Krankheiten, die auf fehlregulierte oder auf die Expression von mutierten Genen zurückgehen, geeignet ist. Vor allem ist ein Arzneimittel geeignet, das die erfindungsgemäßen Nukleinsäuren in nackter Form oder in Form eines der oben

beschriebenen gentherapeutisch wirksamen Vektoren oder in mit Liposomen komplexierter Form enthält.

Als geeignete Zusatz- und/oder Hilfsstoffe eignen sich z.B. eine physiologische Kochsalzlösung, Stabilisatoren, Proteinaseinhibitoren, Nukleaseinhibitoren etc. Ein weiterer Erfindungsgegenstand ist eine Bibliothek die Desoxyribonukleinsäuren umfassen die Sequenzen gemäß **Seq. ID No. 3** oder eine strukturelle Variante davon besitzen und deren codierenden Bereiche eine Transkriptionsaktivierungsdomäne (AD) und gegebenenfalls eine Kerntransportsequenzdomäne (NLS) im offenen Leseraster enthält. Bevorzugt ist darüber hinaus, wenn die DNAs zusätzlich einen Linker, besonders bevorzugt einen Leucin-Zipper codierenden Bereich enthalten (Fig. 1).

Die Peptid-Monomere codierenden Bestandteile der Bibliothek bilden in Lösung , bevorzugt unter physiologischen Bedingungen, homo- und/oder heterodimere biomorphe Transkriptionsfaktoren.

Beispielhaft ist der strukturelle Aufbau der Mitglieder einer Desoxyribonukleinsäuren, umfassend NLS und AD codierende Bereiche, enthaltende Bibliothek in **Seq. ID No 5** und Fig. 1 dargestellt.

In einer bevorzugten Ausführungsform enthält die Bibliothek DNA-Sequenzen, die für die einzelnen Monomere der biomorphen Transkriptionsfaktoren codieren und deren Peptidmonomere nach der Expression in einem transformierten Expressionssystem, wie z. B. in *e. coli* oder in Hefen, über geeignete Linker Dimerisieren.

Als Vektoren können je nach Expressionssystem alle geeigneten pro- und/oder eukaryontischen Expressionsvektoren verwendet werden. Bevorzugt sind kommerziell erhältliche Expressionsvektoren, die einen konstitutiven oder induzierbaren Promotor besitzen, wie z. B. den pGADD424-Vektor von Clontech Laboratories GmbH, Tullastrasse 4, 69126 Heidelberg . Im allgemeinen enthalten die Expressionsvektoren auch für die Wirtszelle geeignete regulatorische Sequenzen, wie z. B. den *trp*-Promotor für die Expression in *e. coli* oder den ADH-2-Promotor für

die Expression in Hefen, den Baculovirus-Polyhedrin-Promotor für die Expression in Insektenzellen.

Bevorzugt sind Bibliotheken, die alle möglichen cDNA-Sequenzen gemäß **Seq. ID No 3** oder einer strukturellen Variante ein oder mehrfach enthalten.

Die DNA-Sequenzen der Bibliothek können in sequenzieller Form, in Form eines geeigneten Vektors oder in Form eines zellularen, mit diesen Vektoren transformierten Expressionssystems vorliegen und gelagert werden.

Die Erfindung umfaßt weiterhin die mit Hilfe dieser Bibliotheken exprimierbaren biomorphen Transkriptionsfaktoren. Auf diese Weise können auch reine Peptidbibliotheken aus biomorphen Transkriptionsfaktoren erhalten werden. Auch die künstliche Synthese von biomorphen Transkriptionsfaktoren zum Erstellen einer Peptidbibliothek kann durchgeführt werden. Beispielhaft ist der strukturelle Aufbau der Mitglieder einer Peptide enthaltende Bibliothek in **Seq. ID No 6** dargestellt.

Ein weiterer Gegenstand ist ein Verfahren zum Auffinden von peptidischen biomorphen Faktoren, die spezifisch Nukleinsäuresequenzen erkennen und binden können.

Dazu wird ein zelluläres Expressionssystem gewählt, bei dem ein essentielles Gen defekt ist. Das zelluläre Expressionssystem ist mit einem entsprechenden Wildtypgen transformiert, daß von einem eine basale Transkription erlaubenden Promotor kontrolliert wird (Insertionselement). Zum anderen enthält das transformierende Konstrukt eine DNA-Sequenz die auf sequenzspezifisch bindende biomorphe Faktoren getestet werden soll (Responseelement).

Als Responseelement sind vor allem regulatorische DNA-Sequenzen, wie Promotoren oder Operatoren von Interesse. Aber auch genspezifische DNA-Sequenzen des codierenden Bereichs eines Gens, insbesondere spezifische DNA-Sequenzen eines mutierten Bereichs eines Gens sind bevorzugte Untersuchungsobjekte.

Als zelluläres Expressionssystem kommen alle lethalen Defektmutanten in Frage, die über eine hemmbare basale Transkription eines eingeschleusten Wildtypgens oder über ein direkt lethal hemmbares Genprodukt kultiviert werden können. Bevorzugt werden leicht kultivierbare Mikroorganismen, wie z. B. *e. coli* oder Hefen verwendet, aber auch Zellkulturen von höheren Organismen sind als Expressionssystem verwendbar.

Beispielhaft können bekannte  $HIS^-$  - Hefezellen als Expressionssystem verwendet werden. Die Hefezellen sind mit einem Konstrukt transformiert, daß eine DNA-Sequenz mit einer Wildtypbindungsstelle, z. B. für Transkriptionsfaktoren, Enhancer oder Repressoren in ein oder mehrfacher Ausführung enthält. Die auf bindende biomorphe Faktoren zu testenden doppelsträngigen Sequenzen können in einen ein Wildtyp-HIS-Gen enthaltenden Vektor, wie dem pHISi-Vektor (Clontech, Heidelberg), eingefügt werden. Das HIS-Gen des pHISi-Vektors steht unter der Kontrolle eines Minimalpromoters.

Das erfindungsgemäße Verfahren zum Auffinden sequenzspezifisch DNA bindenden biomorphen Transkriptionsfaktoren und damit geeigneter Bindungsdomänen umfaßt folgende Verfahrensschritte:

- a) Transformierung eines Expressionssystems umfassend
  - einen Mikroorganismus mit einem lethalen Defekt in einem essentiellen Gen,
  - ein chromosomal integriertes Insertionselement mit einem Wildtypkopie dieses Gens unter der Kontrolle eines inhibierbaren basale Transkription erlaubenden Promoters und einem Responselement, mit einer Bank aus Expressionsvektoren, enthaltend die erfindungsgemäßen Desoxyribonukleinsäurefragmente und einen Aktivierungsdomäne,
- b) Ausplattieren der Expressionssysteme auf einem das essentielle Genprodukt fehlenden Medium,
- c) Inhibierung der Funktion des essentiellen Wildtypgens im Insertionselement,
- d) Isolierung der wachsenden Zellkulturen und Sequenzierung des erfindungsgemäßen DNA-Fragments bzw. der entsprechenden Peptidsequenz

Die folgenden Beispiele dienen der Erläuterung des Verfahrens, ohne dieses darauf zu beschränken:

Herstellung eines Insertionselementes mit einem Responseelement

Die Herstellung einer doppelsträngige DNA-Sequenz, die die Wildtyp-E2F-Bindungsstelle (Fig. 2A) grau unterlegt) dreifach enthält (3xE2Fwt-Sequenz, **Seq. ID No. 7**) wurde wie folgt durchgeführt:

Je 10µl der einzelsträngigen 10µM Oligonukleotide 3xE2F wtsense **Seq. ID No 7** und 3xE2F wtas **Seq. ID No 8** wurden in 100µl 10mM Tris-HCl pH7.9, 50mM NaCl, 10mM MgCl<sub>2</sub>, 1mM Dithiothreitol für 5 min bei 95<sup>0</sup>C erhitzt und langsam (innerhalb von 3h) auf Raumtemperatur abgekühlt.

AAAGCGCGCGAAACTAAAGCGCGCGAAACTAAAGCGCGCGAAACTAGCT

**Seq. ID No 7**

AGTTTCGCGCGCTTTGATTTGCGCGCGCTTTGATTTGCGCGCGCTTTGATC

**Seq. ID No 8**

Die 3xE2Fwt-Sequenz wird in die KpnI und SacI Restriktionsschnittstellen des pGL-2-Vektor (Promega, Madison, WI USA) einkloniert Dazu wurden 3µg pGL-2 mit je 20u KpnI und SacI komplett verdaut und über ein Agarosegel aufgereinigt.

Anschließend wurde die DNA unter Verwendung des QiaQuick Gel Extraction Kit (Qiagen/Hilden) nach Angaben des Herstellers von Agaroseresten befreit und in 50µl Wasser aufgenommen. 10µl des gereinigten Vektors wurden mit 1µl des doppelsträngigen Oligo mit Hilfe der T4 DNA Ligase in einem 20µl-Ansatz für 2h bei 25<sup>0</sup>C ligiert.

Zur Transformation wurde der E.coli k12-Stamm „Goldstar“ (Stratagene, La Jolla, San Diego, USA) über Nacht bei 37<sup>0</sup>C und 200 RPM in LB Medium mit 100µg/ml Ampicillin geschüttelt und angezogen. Am nächsten Morgen wurde 1ml der Bakterienkultur in 200ml frisches Medium überimpft und bis zu einer optischen Dichte von 0.565 bei 595nm bei 37<sup>0</sup>C und 200 RPM geschüttelt. Anschließend wurde die Kultur auf 4<sup>0</sup>C abgekühlt und bei 2500xg abzentrifugiert. Der Überstand wurde verworfen, und die pelletierten Bakterien wurden in 7.5ml LB-Medium mit 10% (w/v) Polyethylenglycol-6000, 5% Dimethylsulfoxid, 10mM MgSO<sub>4</sub>, 10mM (Promega,

Madison, USA), pH 6.8 aufgenommen, für eine Stunde auf Eis inkubiert, in flüssigem Stickstoff schockgefroren und bei  $-80^{\circ}\text{C}$  gelagert. Zur Transformation wurden  $10\mu\text{l}$  des Ligationsansatzes in  $100\mu\text{l}$   $100\text{mM}$  KCl,  $30\text{mM}$   $\text{CaCl}_2$ ,  $50\text{mM}$   $\text{MgCl}_2$  aufgenommen und mit  $100\mu\text{l}$  der aufgetauten Bakterien für 20 min auf Eis inkubiert. Nach einer 10-minütigen Inkubation bei Raumtemperatur wurden die Bakterien mit  $1\text{ml}$  LB-Medium versetzt und unter Schütteln für eine Stunde bei  $37^{\circ}\text{C}$  inkubiert. Anschließend wurde der Ansatz auf LB-Agarplatten mit  $100\mu\text{g/ml}$  Ampicillin ausgestrichen und über Nacht bei  $37^{\circ}\text{C}$  inkubiert. Einzelne Klone wurden isoliert und in  $3\text{ml}$  LB Medium mit  $100\mu\text{g/ml}$  Ampicillin über Nacht bei  $37^{\circ}\text{C}$  angezogen. Aus den Bakterien wurde die Plasmid-DNA unter Verwendung des QIAprep Spin Miniprep-Kits (Qiagen/Hilden) nach Angaben des Herstellers isoliert und gereinigt. Positive Klone wurden durch PCR wie folgt identifiziert:  $1\mu\text{l}$  DNA,  $0.5\mu\text{l}$  entspr. 5U Taq-Polymerase (Promega, Madison, USA),  $5\mu\text{l}$   $10\times$  Reaktionspuffer (Promega, Madison, USA),  $0.4\mu\text{l}$  Desoxynukleotidtriphosphat (je  $10\mu\text{M}$ ),  $4\mu\text{l}$   $12.5\text{mM}$   $\text{MgCl}_2$ , sowie je  $0.3\mu\text{l}$  der  $100\mu\text{M}$  Oligonukleotide und destilliertes Wasser ad  $50\mu\text{l}$  wurden für 3min auf  $94^{\circ}\text{C}$  erhitzt. Es schlossen sich 30 Zyklen mit  $94^{\circ}\text{C}$  für 20 sec.,  $50^{\circ}\text{C}$  für 20sec., und  $72^{\circ}\text{C}$  für 45 sec an, gefolgt von einer Inkubation bei  $72^{\circ}\text{C}$  für 5min. Die PCR-Produkte wurden durch Agarosegelelektrophorese aufgetrennt und damit ihre Größe bestimmt. Bakterienkolonien, deren Plasmid-DNA ein PCR-Produkt der erwarteten Größe aufwies, wurden in  $50\text{ml}$  LB Medium mit  $100\mu\text{g/ml}$  Ampicillin für 13h bei  $37^{\circ}\text{C}$  und 250 RPM geschüttelt. Anschließend wurde die Plasmid-DNA mit Hilfe des Qiagen Midi Präp Kits (Qiagen/Hilden) aus den Bakterien isoliert und gereinigt. Die DNA wurde in destilliertem Wasser resuspendiert. Anschließend wurde die somit amplifizierte 3xE2Fwt Sequenz aus dem Vektor isoliert. Dazu wurden  $3\mu\text{g}$  des Plasmids mit je 20U XmaI und MluI für 3 Stunden komplett verdaut. Der Ansatz wurde anschließend über ein Agarosegel aufgetrennt, und die bei 68bp migrierende 3xE2Fwt Insertionssequenz wurde mit einem Skalpell ausgeschnitten. Parallel dazu wurden  $3\mu\text{g}$  pHISi (Clontech, Heidelberg) mit je 20U XmaI und MluI für 3 Stunden komplett verdaut und über ein Agarosegel gereinigt. Anschließend wurden die DNA-Proben unter Verwendung des QiaQuick Gel Extraction Kit (Qiagen/Hilden) nach Angaben des Herstellers von Agarose befreit und in je  $50\mu\text{l}$  Wasser aufgenommen. Zur Ligation wurden  $3\mu\text{l}$  des mit XmaI und MluI verdauten pHISi Vektors mit  $10\mu\text{l}$  der mit XmaI und MluI isolierten 3xE2Fwt Insertionssequenz in einem  $20\mu\text{l}$  Ansatz mit  $1,5\mu\text{l}$  T4 DNA Ligase für 3 Stunden bei Raumtemperatur inkubiert. Zur

Transformation wurde der E.coli k12-Stamm „Goldstar“ (Stratagene, La Jolla, San Diego, USA) über Nacht bei 37°C und 200 RPM in LB Medium mit 100µg/ml Ampicillin geschüttelt und angezogen. Am nächsten Morgen wurde 1ml der Bakterienkultur in 200ml frisches Medium überimpft und bis zu einer optischen Dichte von 0.565 bei 595nm bei 37°C und 200 RPM geschüttelt. Anschließend wurde die Kultur auf 4°C abgekühlt und bei 2500xg abzentrifugiert. Der Überstand wurde verworfen, und die pelletierten Bakterien wurden in 7.5ml LB-Medium mit 10% (w/v) Polyethylenglycol-6000, 5% Dimethylsulfoxid, 10mM MgSO<sub>4</sub>, 10mM (Promega, Madison, USA), pH 6.8 aufgenommen, für eine Stunde auf Eis inkubiert, in flüssigem Stickstoff schockgefroren und bei -80°C gelagert. Zur Transformation wurden 10µl des Ligationsansatzes in 100µl 100mM KCl, 30mM CaCl<sub>2</sub>, 50mM MgCl<sub>2</sub> aufgenommen und mit 100µl der aufgetauten Bakterien für 20 min auf Eis inkubiert. Nach einer 10-minütigen Inkubation bei Raumtemperatur wurden die Bakterien mit 1ml LB-Medium versetzt und unter Schütteln für eine Stunde bei 37°C inkubiert. Anschließend wurde der Ansatz auf LB-Agarplatten mit 100µg/ml Ampicillin ausgestrichen und über Nacht bei 37°C inkubiert. Einzelne Klone wurden isoliert und in 3ml LB Medium mit 100µg/ml Ampicillin über Nacht bei 37°C angezogen. Aus den Bakterien wurde die Plasmid-DNA unter Verwendung des QIAprep Spin Miniprep-Kits (Qiagen/Hilden) nach Angaben des Herstellers isoliert und gereinigt. Positive Klone wurden durch Verdau mit XmaI und MluI identifiziert. Die entsprechenden Bakterien wurden in 50 ml LB Medium mit 100µg/ml Ampicillin für 13h bei 37°C und 250 RPM geschüttelt. Anschließend wurde die Plasmid-DNA mit Hilfe des Qiagen Midi Präp Kits (Qiagen/Hilden) aus den Bakterien isoliert und gereinigt. Die DNA wurde in destilliertem Wasser resuspendiert. Das resultierende Plasmid wird als pHISi3xE2Fwt (**Seq. ID. No. 9**) bezeichnet.

#### Herstellung transformierter HIS<sup>-</sup>-Hefestämme

In dem Plasmid pHISi3xE2Fwt wird das HIS3-Gen des Vektors neben dem im Vektor vorhandenen basalen HIS3 Promoter unter die Kontrolle des Insertionselementes gestellt. Der Hefestamm YM 4271 (Clontech, Heidelberg) wird mit dem so hergestellten pHISi-3xE2Fwt-Vektor gemäß den Angaben des Herstellers (Yeast Protocols Handbook, Clontech, Heidelberg S. 20-21, 1999, Clontech Laboratories Inc.) mit 1 µg des mit Xho I linearisierten Plasmids wie folgt transformiert:

1 µg pHISi3xE2Fwt wurde für 1 Stunde mit 10U XhoI komplett linearisiert, unter Verwendung des QiaQuick PCR Purification Kit (Qiagen/Hilden) nach Angaben des Herstellers aufgereinigt und in 50 µl Wasser aufgenommen. Mehrere Kolonien des Hefestammes YM 4271 (Clontech, Heidelberg) wurden in 50 ml YPD Medium bis zu einer optischen Dichte bei 600nm von 1.5 bei 30°C angezogen und so in 300ml YPD Medium überimpft, daß die Kultur eine optische Dichte bei 600 nm von 0.2-0.3 hatte. Diese Kultur wurde unter Schütteln bei 30°C inkubiert, bis sie eine optische Dichte bei 600 nm von 0.4-0.6 hatte. Die Kultur wurde anschließend bei 1000xg für 5 Minuten abzentrifugiert, und die Zellen wurden in 1.5 ml 100 mM Lithiumacetat pH 7.5, 10mM Tris-HCl pH7.5, 1mM EDTA aufgenommen. 100 µl der Zellen, 50 µl des linearisierten Plasmids sowie 100 µg hitzedenaturierter Lachssperma-DNA wurden mit 600 µl einer PEG/LiAc-Lösung (100mM Lithiumacetat pH 7.5, 10 mM Tris-HCl pH7.5, 1mM EDTA, 40% (w/v) Polyethylenglycol-4000) versetzt und für 30 min bei 30°C geschüttelt. Nach Zugabe von 70 µl Dimethylsulfoxid wurde der Ansatz für 15 min bei 42°C inkubiert. Anschließend wurden die Zellen abzentrifugiert, in 1ml 10 mM Tris-HCl pH7.5, 1mM EDTA gewaschen und komplett auf eine Minimalmediumagarplatte ohne Histidinzusatz ausplattiert, auf der nach 7 Tagen bei 30°C 12 Kolonien gewachsen waren.

Die geringe basale Expression von HIS3 erlaubt die Selektion von transformierten YM 4271-Hefen die das Insertionselement in ihre chromosomale DNA eingebaut haben durch Ausplattieren der Zellkulturen auf SD-Agarplatten mit Histidin freiem Medium. Die unter diesen Bedingungen wachsenden YM 4271-Hefestämme werden isoliert.

Auswahl eines besonders geeigneten transformierter HIS<sup>-</sup>-Hefestamms

12 aus unabhängigen Rekombinationsereignissen resultierende Kolonien wurden anschließend auf Minimalmedium-Agarplatten ohne Histidin ausgestrichen, die steigende Mengen des HIS3-Antagonisten 3-Aminotriazol (3-AT) enthielten 0mM, 0.5mM, 1mM, 1.5mM, 2mM, 3mM, 6mM, 9mM, 12mM, 15mM, 18mM, 30mM, 45mM; ein Klon wurde bereits bei 0.5mM deutlich und bei 2mM vollständig supprimiert; dieser und ein weiterer Klon wuchsen nicht mehr bei 3mM, während die 10 anderen erst bei >30mM im Wachstum inhibiert wurden.

Das Wachstum einer dieser Kolonien, als yeast3xE2Fwt bezeichnet, wurde bereits bei einer Konzentration von 2mM 3-AT vollständig unterdrückt (Fig. 2D) ohne Antagonisten (Fig. 2C)). Diese Kolonie wurde isoliert und für die weiteren Untersuchungen herangezogen.

Der ausgewählte Stamm yeast3xE2Fwt besitzt ein vorteilhaftes sehr geringes basales Expressionsniveau von HIS3, was bereits durch geringe Mengen 3-AT kompetitiv inhibiert wird. Der Stamm wird als YM 4271 3xE2Fwt bezeichnet

Erzeugung einer DNA-sequenzspezifische Bindungspartner enthaltende Bibliothek

3µg pGADD424 (Clontech/Heidelberg, **Seq. ID No 11**) wurden mit je 20u EcoRI und BamHI komplett verdaut und über ein Agarosegel aufgereinigt. 1µg des doppelsträngigen Peptidomänen codierenden Oligonukleotids **Seq ID No 10** wurde synthetisiert und für 3h mit je 10u EcoRI und BamHI verdaut und über ein Agarosegel aufgereinigt. Anschließend wurde die DNA unter Verwendung des QiaQuick Gel Extraction Kit (Qiagen/Hilden) nach Angaben des Herstellers von Agaroseresten befreit und in 50µl Wasser aufgenommen. 3µl des Vektors wurden mit 0.1µl, 0.3µl und 1µl des Oligonukleotids für 3h bei Raumtemperatur in 20µl-Ansätzen mit Hilfe der T4-DNA Ligase ligiert. Anschließend wurden die Ansätze mit je 200µl n-Butanol geschüttelt und für 10 Minuten bei 13000 UpM in einer Tischzentrifuge zentrifugiert. Die Überstände wurden verworfen, und die DNA wurde durch Eindampfen im Vakuum von Butanolresten befreit sowie in je 10µl Wasser aufgenommen. Zur Transformation wurden 500ml einer exponentiell wachsenden Bakterienkultur des E.coli K12-Stammes TOP10 für 10 min bei 4000xg zentrifugiert, in 500ml kaltem Wasser resuspendiert und erneut zentrifugiert. Diese Prozedur wurde zweimal wiederholt. Anschließend wurden die Bakterien in 7.5ml 10% Glycerin (v/v) aufgenommen und als 40µl-Aliquots in flüssigen Stickstoff schockgefroren. Zur eigentlichen Transformation wurden je 5µl der oben beschriebenen gereinigten Ligationsansätze mit einem auf Eis aufgetauten Aliquot der Bakterien vermischt und in eine kalte Elektroporationsküvette überführt. Nach einem elektrischen Puls von 1,8 kV bei 200Ω und 25µF wurden die Bakterien für

eine Stunde in 1ml LB-Medium bei 37°C und 250 RPM geschüttelt und anschließend auf LB-Agarplatten mit 100µg/ml Ampicilin über Nacht bei 37°C inkubiert.

Am folgenden Tag wurden die Klone der dicht bewachsenen Platten abgeschabt und in 250 ml LB Medium mit 100µg/ml Ampicilin für 3h bei 37°C und 250 RPM geschüttelt. Anschließend wurde die Plasmid-DNA mit Hilfe des Qiagen Maxi Präp Kits (Qiagen/Hilden) aus den Bakterien isoliert und gereinigt. Die DNA wurde in destilliertem Wasser resuspendiert und hatte eine Konzentration von 1.3µg/µl.

```
GGGAATTCGATCCTGCTGCTCTAAAACGTGCTAGANNCCTGAANNNNNNAGG
CGTNNNCGTGCGAGAAAGTTGCAACTTGAAGACAAGGTTGAAGAATTGCTTTTCG
AAAATTATCACTTGGAAAATGAGGTTGCCAGATTAAGAAATTAGTTGGCGAAC
GCTGAGGATCCCC
```

#### **Seq ID No 10:**

Auffinden von sequenzspezifisch bindenden biomorphen Transkriptionsfaktoren

Die erzeugte cDNA-Bibliothek wird in den Hefestamm yeast3xE2Fwt nach Angaben des Herstellers (Yeast protocols handbook, Clontech, Heidelberg) wie folgt transformiert:

Mehrere Kolonien des Hefestammes YM 4271 3xE2Fwt wurden in 50 ml YPD Medium bis zu einer optischen Dichte bei 600nm von 1.5 bei 30°C angezogen und so in 300ml YPD Medium überimpft, daß die Kultur eine optische Dichte bei 600 nm von 0.2-0.3 hatte. Diese Kultur wurde unter Schütteln bei 30°C inkubiert, bis sie eine optische Dichte bei 600 nm von 0.4-0.6 hatte. Die Kultur wurde anschließend bei 1000xg für 5 Minuten abzentrifugiert, und die Zellen wurden in 1.5 ml 100 mM Lithiumacetat pH 7.5, 10mM Tris-HCl pH7.5, 1mM EDTA aufgenommen. 1 ml der Zellen, 50µg der Bibliothek sowie 2000µg hitzedenaturierter Lachssperma-DNA wurden mit 6ml einer PEG/LiAc-Lösung (100mM Lithiumacetat pH 7.5, 10 mM Tris-HCl pH7.5, 1mM EDTA, 40% (w/v) Polyethylenglycol-4000) versetzt und für 30 min bei 30°C geschüttelt. Nach Zugabe von 700µl Dimethylsulfoxid wurde der Ansatz für 15 min bei 42°C inkubiert. Anschließend wurden die Zellen abzentrifugiert, in 10 mM Tris-HCl pH7.5, 1mM EDTA gewaschen und komplett auf 6 Minimalmediumagarplatten ohne Histidin- und Leucinzusatz sowie 2mM 3-

Aminotriazol ausplattiert, auf der nach 7 Tagen bei 30°C 153 Kolonien gewachsen waren.

Die Selektion von Hefestämmen, die sowohl das Insertionselement als auch einen Bibliotheksbestandteil inklusive des pGAD424-Vektor-Leu2-Marker enthalten, wird einfach durch Ausplattieren auf SD-Agarplatten ohne Histidin und Leucin durchgeführt.

Zur Selektion derjenigen Hefezellen, bei denen das HIS3-Expressionsniveau durch Bindung eines biomorphen Transkriptionsfaktors aus der Bibliothek an die trimerisierte E2F-Bindungsstelle erhöht ist, wird den Agarplatten 2mM 3-Aminotriazol zugesetzt.

Während bei der Kontrolltransformation von yeast3xE2Fwt mit 50 µg pGAD424 alleine nur eine Hefekolonie wuchs (Fig. 3A)), vermutlich durch unspezifische Rückmutation, wurden bei der Transformation mit 50µg der Bibliothek 153 Kolonien (Ausschnitt Fig. 3B)) erhalten.

Die unter den gegebenen Selektionsbedingungen wachsenden Hefekolonien sind mit Bibliotheksbestandteilen transformiert die für expremierbare biomorphe Transkriptionsfaktoren codieren deren Wirkung spezifisch auf das HIS3-Gen gerichtet ist und die die E2F-Bindungsstelle, die einzige mehrfach vorkommende Sequenz des HIS3-Promotors (die HIS3-Expression lässt sich nur durch an multiple Sequenzwiederholungen bindende Faktoren regulieren, da nur die Bindung mehrerer Transkriptionsfaktoren zu einem kooperativen und somit steilen Anstieg der Transkriptionsinitiation führt, enthalten. Diese Kolonien besitzen demnach Bibliotheksmitglieder, deren Expressionsprodukte spezifisch an die trimerisierte E2F-Bindungsstelle des HIS3-Genkonstrukts binden.

Um die Sequenz dieser Bibliotheksmitglieder zu bestimmen, wurden die Plasmide aus 32 der Hefezellkulturen isoliert, in E.coli transformiert und vermehrt (Yeast protocols handbook, Clontech, Heidelberg).

Dazu wurden einzelne Hefekolonien in 0.5 ml SD-Minimalmedium ohne Histidin- und Leucinzusatz über Nacht bei 30°C geschüttelt. Am nächsten Tag wurden die Zellen

durch Zentrifugation pelletiert und soviel von dem Medium entfernt, daß etwa 50µl übrig blieben. In diesem Mediumrest wurden die Zellen durch vortexen resuspendiert und mit einer Spatelspitze des zellwallzerstörenden Enzyms Lyticase versetzt. Unter Schütteln wurden die Zellen für eine Stunde bei 37<sup>0</sup>C inkubiert und anschließend mit 10µl einer 20% Sodiumdodecylsulfat-Lösung versetzt. Zum Lysieren wurden die Zellen für 1 min gevortext, kurz bei -20<sup>0</sup>C eingefroren und schnell aufgetaut. Die Zelltrümmer wurden durch Zentrifugation bei 14,000 UpM abgetrennt, und die im Überstand vorhandene Plasmid-DNA wurde unter Verwendung des QiaQuick PCR Purification Kit von Verunreinigungen befreit sowie in 50µl Wasser aufgenommen. Anschließend wurde die Plasmid-DNA in Bakterien transformiert und vermehrt, um größere Mengen der Plasmide zu erhalten. Zur Transformation wurde der E.coli k12-Stamm „Goldstar“ (Stratagene, La Jolla, San Diego, USA) über Nacht bei 37<sup>0</sup>C und 200 RPM in LB Medium mit 100µg/ml Ampicillin geschüttelt und angezogen. Am nächsten Morgen wurde 1ml der Bakterienkultur in 200ml frisches Medium überimpft und bis zu einer optischen Dichte von etwa 0.5 bei 595nm bei 37<sup>0</sup>C und 200 RPM geschüttelt. Anschließend wurde die Kultur auf 4<sup>0</sup>C abgekühlt und bei 2500xg abzentrifugiert. Der Überstand wurde verworfen, und die pelletierten Bakterien wurden in 7.5ml LB-Medium mit 10% (w/v) Polyethylenglycol-6000, 5% Dimethylsulfoxid, 10mM MgSO<sub>4</sub>, 10mM (Promega, Madison, USA), pH 6.8 aufgenommen, für eine Stunde auf Eis inkubiert, in flüssigem Stickstoff schockgefroren und bei -80<sup>0</sup>C gelagert. Zur Transformation wurden 20µl der isolierten Plasmid-DNA in 100µl 100mM KCl, 30mM CaCl<sub>2</sub>, 50mM MgCl<sub>2</sub> aufgenommen und mit 100µl der aufgetauten Bakterien für 20 min auf Eis inkubiert. Nach einer 10-minütigen Inkubation bei Raumtemperatur wurden die Bakterien mit 1ml LB-Medium versetzt und unter Schütteln für eine Stunde bei 37<sup>0</sup>C inkubiert. Anschließend wurde der Ansatz auf LB-Agarplatten mit 100µg/ml Ampicillin ausgestrichen und über Nacht bei 37<sup>0</sup>C inkubiert. Von jeder Transformation wurde ein einzelner Klon in 3 ml LB-Medium über Nacht bei 37<sup>0</sup>C angezogen. Anschließend wurden die Plasmide unter Verwendung des Qiagen Tip20 Kits nach Angaben des Herstellers (Qiagen, Hilden ) präpariert und aufgereinigt.

Die Sequenzierung der Plasmide mit dem Oligonukleotid 5'-GATGTATATAACTATCTATTTCG-3' (**Seq. ID No 12**) wird mit dem Fachmann

bekannten Standardmethoden (z. B. nach Sanger, apparativ mit Sequenzer) durchgeführt.

Die folgenden Peptidsequenzen (Tabelle 1, dargestellt Aminosäuren 1 bis 20) konnten als die 3xE2Fwt-Sequenz spezifisch erkennende Domänen ermittelt werden:

Tabelle 1:

Seq. ID No.										10										20
13	D	P	A	A	L	K	R	A	R	G	T	E	V	V	R	R	G	R	A	R
14	D	P	A	A	L	K	R	A	R	C	T	E	V	M	R	R	Q	R	A	R
15	D	P	A	A	L	K	R	A	R	P	T	E	N	V	R	R	G	R	A	R
16	D	P	A	A	L	K	R	A	R	N	T	E	V	T	R	R	S	R	A	R
17	D	P	A	A	L	K	R	A	R	V	T	E	N	S	R	R	D	R	A	R
18	D	P	A	A	L	K	R	A	R	H	T	E	T	S	R	R	I	R	A	R
19	D	P	A	A	L	K	R	A	R	V	T	E	V	I	R	R	G	R	A	R
20	D	P	A	A	L	K	R	A	R	I	T	E	G	I	R	R	L	R	A	R
21	D	P	A	A	L	K	R	A	R	S	T	E	L	N	R	R	G	R	A	R
22	D	P	A	A	L	K	R	A	R	G	T	E	R	L	R	R	G	R	A	R
23	D	P	A	A	L	K	R	A	R	G	T	E	A	T	R	R	V	R	A	R
24	D	P	A	A	L	K	R	A	R	C	T	E	E	V	R	R	W	R	A	R
25	D	P	A	A	L	K	R	A	R	C	T	E	V	Q	R	R	G	R	A	R
26	D	P	A	A	L	K	R	A	R	D	T	E	M	L	R	R	C	R	A	R
27	D	P	A	A	L	K	R	A	R	D	T	E	M	V	R	R	A	R	A	R
28	D	P	A	A	L	K	R	A	R	G	T	E	V	V	R	R	C	R	A	R
29	D	P	A	A	L	K	R	A	R	G	T	E	V	V	R	R	C	R	A	R
30	D	P	A	A	L	K	R	A	R	I	T	E	N	A	R	R	G	R	A	R
31	D	P	A	A	L	K	R	A	R	C	T	E	E	M	R	R	G	R	A	R
32	D	P	A	A	L	K	R	A	R	C	T	E	P	S	R	R	G	R	A	R
33	D	P	A	A	L	K	R	A	R	N	T	E	S	G	R	R	T	R	A	R
34	D	P	A	A	L	K	R	A	R	D	T	E	G	D	R	R	R	R	A	R
35	D	P	A	A	L	K	R	A	R	G	T	E	A	E	R	R	G	R	A	R

36	D	P	A	A	L	K	R	A	R	R	T	E	L	L	R	R	H	R	A	R
37	D	P	A	A	L	K	R	A	R	I	T	E	N	A	R	R	G	R	A	R
38	D	P	A	A	L	K	R	A	R	I	T	E	M	G	R	R	K	R	A	R
39	D	P	A	A	L	K	R	A	R	C	T	E	Y	C	R	R	I	R	A	R
40	D	P	A	A	L	K	R	A	R	G	T	E	E	Y	R	R	H	R	A	R
41	D	P	A	A	L	K	R	A	R	S	T	E	L	T	R	R	I	R	A	R
42	D	P	A	A	L	K	R	A	R	G	T	E	P	V	R	R	S	R	A	R
43	D	P	A	A	L	K	R	A	R	G	T	E	A	E	R	R	G	R	A	R
44	D	P	A	A	L	K	R	A	R	D	T	E	A	S	R	R	M	R	A	R
45	D	P	A	A	L	K	R	A	R	N	T	E	S	G	R	R	T	R	A	R

Die Domänen weisen an den für die sequenzspezifische Bindung entscheidenden Positionen Aminosäuren mit sehr ähnlichen Eigenschaften auf, z. B. Valin und Isoleucin (schattiert dargestellt). Aus den Einzelsequenzen kann eine Konsensussequenz (**Seq. ID No. 13**) abgeleitet werden.

Der Vergleich der aus den DNA-Sequenzen dieser Bibliotheksmitglieder abgeleiteten Aminosäuresequenz ergibt, daß in allen 4 Positionen, die in **Seq. ID No 1** mit X gekennzeichnet sind, Aminosäurereste mit kurzen Seitenketten (G,V,C,S) dominierten (Tabelle 2). Besonders deutlich ist dabei das häufige Auftreten der Aminosäure Glycin in Position X-1 und X-4.

Tabelle 2

	G	M	P	H	N	C	Q	S	T	Y	A	I	L	F	W	V	D	E	K	R
<b>Position X-1</b>	8	0	1	1	3	6	0	2	0	0	0	4	0	0	0	2	4	0	0	1
<b>Position X-2</b>	2	3	2	0	4	0	0	2	1	1	4	0	3	0	0	6	0	3	0	1
<b>Position X-3</b>	3	2	0	0	1	1	1	4	3	1	2	2	3	0	0	6	1	2	0	0
<b>Position X-4</b>	11	1	0	2	0	3	1	2	2	0	1	3	1	0	1	1	1	0	1	1

Aus dem Vergleich der Aminosäuresequenz der Binder kann eine Konsensussequenz der Erkennungsaminosäuren (Pos. X-1: G oder C, Pos. X-2: V, Pos X-3: V, Pos. X-4: G) der neuartigen Peptide abgeleitet werden, die jedoch nicht als eigenständige Sequenz unter den analysierten Bibliotheksmitgliedern vorkam. Dies kann jedoch dadurch begründet werden, daß bei einer geschätzten Effizienz der Bibliothekstransformation von ca. 200,000 Transformationsereignissen und einem Verhältnis Vektor (mit Insert)/Vektor (ohne Insert) von 1:5 nur etwa 40,000 verschiedene Bibliotheksmitglieder transformiert wurden. Bei einer Komplexität der Bibliothek von 104,000 kodierenden Mitgliedern wurde daher jedes Mitglied mit einer Wahrscheinlichkeit von 0.38 in die Hefe transformiert.

## Patentansprüche:

1. Biomorphe Peptide enthaltend eine Aminosäuresequenz gemäß **Seq. ID No. 1** oder einer strukturellen Variante davon.
2. Biomorphe Peptide gemäß Anspruch 1, dadurch gekennzeichnet, daß die Peptide in homo- und/oder heterodimerer Form vorliegen.
3. Biomorphe Peptide gemäß einem der vorherigen Ansprüche, dadurch gekennzeichnet, daß mindestens zwei Peptide ausgewählt aus der Gruppe gemäß **Seq. ID No 1** oder einer strukturellen Variante davon über einen Linker verknüpft sind.
4. Biomorphe Peptide gemäß Anspruch 3 dadurch gekennzeichnet, daß als Linker Komplexbildner oder Leucin-Zipper enthalten sind.
5. Biomorphe Peptide gemäß einem der vorherigen Ansprüche enthaltend eine Aminosäuresequenz gemäß **Seq. ID No. 2**.
6. Biomorphe Peptide gemäß einem der vorherigen Ansprüche enthaltend eine Aminosäuresequenz gemäß **Seq. ID No. 1**, einer strukturellen Variante davon oder **Seq. ID No 2** und Kerntransportsignaldomänen, Transkriptionsaktivierungsdomänen oder Transkriptionshemmungsdomänen.
7. Verwendung von biomorphen Peptiden enthaltend eine Aminosäuresequenz gemäß **Seq. ID No. 1**, einer strukturellen Variante davon oder **Seq. ID No 2** zur Herstellung von Arzneimitteln zur Behandlung von genregulatorischen Störungen.
8. Nukleinsäuren umfassend biomorphe Peptide nach den Ansprüchen 1 bis 6 codierende Nukleinsäuresequenzen gemäß **Seq. ID No 3** oder einer strukturellen Variante davon.

9. Nukleinsäuren gemäß Anspruch 8, dadurch gekennzeichnet, daß sie eine Leucin-Zipper-Sequenz enthalten.
10. Nukleinsäuren gemäß Anspruch 8 oder 9, dadurch gekennzeichnet, daß sie Kerntansprotsignaldomänen, Transkriptionsaktivierungsdomänen oder Transkriptionseinhibierungsdomänen codierende Bereiche enthalten.
11. Vektoren enthalten Nukleinsäuren nach Anspruch 8 oder 9.
12. Nukleinsäuren gemäß den Ansprüchen 8 bis 10, dadurch gekennzeichnet, daß die Nukleinsäure eine DNA oder cDNA ist.
13. Verwendung von Nukleinsäuren oder Vektoren nach den Ansprüchen 8 bis 11 zur Herstellung von gentherapeutischen Arzneimitteln zur Behandlung von genregulatorischen Störungen.
14. Bibliotheken umfassend Desoxyribonukleinsäuren die eine Nukleinsäuresequenz gemäß **Seq. ID No 3** oder einer strukturellen Variante davon und einen Transkriptionsaktivierungsdomäne codierenden Bereich enthalten.
15. Bibliotheken gemäß Anspruch 14, dadurch gekennzeichnet, daß die Desoxyribonukleinsäuren zusätzlich einen Kerntransprotsignaldomäne und/oder einen Leucin-Zipper codierenden Bereich enthalten.
16. Peptidbibliotheken umfassend biomorphe Transkriptionsfaktoren die Aminosäuresequenzen gemäß **Seq. ID No 1** oder einer strukturellen Variante davon enthalten.
17. Verfahren zum Auffinden sequenzspezifisch DNA bindender biomorpher Transkriptionsfaktoren umfassend folgende Verfahrensschritte:
  - e) Transformierung eines Expressionssystems umfassend

- einen Mikroorganismus mit einem lethalen Defekt in einem essentiellen Gen,
  - ein chromosomal integriertes Insertionselement mit einer Wildtypkopie dieses Gens unter der Kontrolle eines inhibierbaren basale Transkription erlaubenden Promoters und einem Responseelement, mit einer Bibliothek aus Expressionsvektoren, enthaltend die erfindungsgemäßen Desoxyribonukleinsäurefragmente und einen Aktivierungsdomäne,
- f) Ausplattieren der Expressionssysteme auf einem das essentielle Genprodukt fehlenden Medium,
- g) Inhibierung der basalen Transkription des essentiellen Wildtypgens im Insertionselement,
- h) Isolierung der wachsenden Zellkulturen und Sequenzierung des erfindungsgemäßen DNA-Fragments bzw. der entsprechenden Peptidsequenz.
18. Biomorphes Peptid enthaltend eine Aminosäuresequenz gemäß **Seq. ID No. 1** oder einer strukturellen Variante, dadurch gekennzeichnet, daß die Aminosäure an Position X-1 ein G, C, I, D, N, S, P, H oder R ist, die Aminosäure an Position X-2 ein V, A, N, M, L, E, G, P, S, T, Y oder R ist, die Aminosäure an Position X-3 ein V, S, G, T, L, M, A, I, E, N, C, Q, Y oder D ist, die Aminosäure an Position X-4 ein G, C, I, H, S, T, M, Q, A, L, W, V, D, K oder R ist.
19. Biomorphes Peptid gemäß Anspruch 18, dadurch gekennzeichnet, daß die Aminosäure an Position X-1 ein G, C, I, oder D ist, die Aminosäure an Position X-2 ein V, A, oder N ist, die Aminosäure an Position X-3 ein V oder S ist, die Aminosäure an Position X-4 ein G ist.
20. Biomorphes Peptid gemäß Anspruch 18 oder 19, dadurch gekennzeichnet, daß die Aminosäure an Position X-1 ein G ist, die Aminosäure an Position X-2 ein V ist, die Aminosäure an Position X-3 ein V ist, die Aminosäure an Position X-4 ein G ist.

21. Künstliche biomorphe Faktoren enthaltend ein Peptid mit einer Sequenz gemäß Anspruch 18 bis 20.
22. Künstliche biomorphe Transkriptionsfaktoren oder Repressoren gemäß Anspruch 21 die spezifisch die E2Fwt-Sequenz binden.

**Fig. 1**



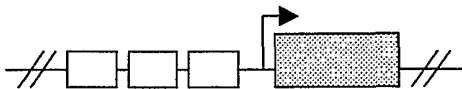
**Fig. 2**

**A)**

AAAGCGCGCGAAACTAAAGCGCGCGAAACTAAAGCGCGCGAAACTAGCT

(Seq. ID No. 7)

**B)**



**C)**

- 3-AT



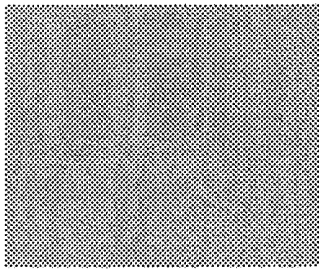
**D)**

+ 3-AT

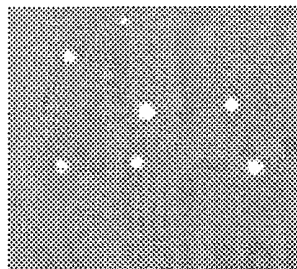


**Fig. 3**

**A)**



**B)**



## SEQUENZPROTOKOLL

<110> Xzillion GmbH & Co KG

<120> DNA-bindende Peptidomänen und ein Verfahren zur  
Bereitstellung solcher Domänen

<130> 200at10.wo

<140>

<141>

<150> 10041126.6

<151> 2000-08-22

<160> 45

<170> PatentIn Ver. 2.1

<210> 1

<211> 23

<212> PRT

<213> Künstliche Sequenz

<220>

<223> Beschreibung der künstlichen Sequenz: Peptidomäne

<400> 1

Asp Pro Ala Ala Leu Lys Arg Ala Arg Xaa Thr Glu Xaa Xaa Arg Arg

1 5 10 15

Xaa Arg Ala Arg Lys Leu Gln

20

<210> 2

<211> 52

<212> PRT

<213> Künstliche Sequenz

<220>

<223> Beschreibung der künstlichen Sequenz:

Leucin-Zipper-Fusionsprotein

<400> 2

Asp Pro Ala Ala Leu Lys Arg Ala Arg Xaa Thr Glu Xaa Xaa Arg Arg

1 5 10 15



Transkriptionsaktivierungssequenz

<400> 5

```

atggataaag cgggaattaat tcccgcgcct ccaaaaaaga agagaaaggt cgaattgggt 60
accgccgcca attttaataca aagtgggaat attgctgata gctcattgtc cttcactttc 120
actaacagta gcaacgggtcc gaacctcata acaactcaaa caaattctca agcgctttca 180
caaccaattg cctcctctaa cgttcatgat aacttcatga ataatgaaat cacggctagt 240
aaaattgatg atggtaataa ttcaaacca ctgtcacctg gttggacgga ccaaaactgcg 300
tataacgcgt ttggaatcac tacagggatg tttaatacca ctacaatgga tgatgtatat 360
aactatctat tcgatgatga agatacccca ccaaaccba aaaaagagat cgaattcgat 420
cctgctgctc taaaacgtgc tagannccac tgaannnnnn aggcgtnnnc gtgcgagaaa 480
gttgcaactt gaagacaagg ttgaagaatt gctttcgaaa aattatcact tggaaaatga 540
ggttgccaga ttaaagaaat tagttggcga acgctgagga tcccca 586
    
```

<210> 6

<211> 191

<212> PRT

<213> Künstliche Sequenz

<220>

<223> Beschreibung der künstlichen Sequenz:

Kerntransportsequenz und  
 Transkriptionsaktivierungssequenz enthaltende  
 Peptide

<400> 6

```

Met Asp Lys Ala Glu Leu Ile Pro Glu Pro Pro Lys Lys Lys Arg Lys
 1           5           10           15

Val Glu Leu Gly Thr Ala Ala Asn Phe Asn Gln Ser Gly Asn Ile Ala
          20           25           30

Asp Ser Ser Leu Ser Phe Thr Phe Thr Asn Ser Ser Asn Gly Pro Asn
          35           40           45

Leu Ile Thr Thr Gln Thr Asn Ser Gln Ala Leu Ser Gln Pro Ile Ala
          50           55           60

Ser Ser Asn Val His Asp Asn Phe Met Asn Asn Glu Ile Thr Ala Ser
          65           70           75           80

Lys Ile Asp Asp Gly Asn Asn Ser Lys Pro Leu Ser Pro Gly Trp Thr
          85           90           95

Asp Gln Thr Ala Tyr Asn Ala Phe Gly Ile Thr Thr Gly Met Phe Asn
          100          105          110
    
```

Thr Thr Thr Met Asp Asp Val Tyr Asn Tyr Leu Phe Asp Asp Glu Asp  
 115 120 125

Thr Pro Pro Asn Pro Lys Lys Glu Ile Glu Phe Asp Pro Ala Ala Leu  
 130 135 140

Lys Arg Ala Arg Xaa Thr Glu Xaa Xaa Arg Arg Xaa Arg Ala Arg Lys  
 145 150 155 160

Leu Gln Leu Glu Asp Lys Val Glu Glu Leu Leu Ser Lys Asn Tyr His  
 165 170 175

Leu Glu Asn Glu Val Ala Arg Leu Lys Lys Leu Val Gly Glu Arg  
 180 185 190

<210> 7

<211> 49

<212> DNA

<213> Künstliche Sequenz

<220>

<223> Beschreibung der künstlichen Sequenz: 3xE2Fwt  
 Sequenz sense

<400> 7

aaagcgcgcg aaactaaagc gcgcgaaact aaagcgcgcg aaactagct 49

<210> 8

<211> 49

<212> DNA

<213> Künstliche Sequenz

<220>

<223> Beschreibung der künstlichen Sequenz: 3xE2Fwt  
 antisense Sequenz

<400> 8

agtttcgcgc gctttgattt cgcgcgcttt gatttcgcgc gctttgatc 49

<210> 9

<211> 6838

<212> DNA

<213> Künstliche Sequenz

<220>

<223> Beschreibung der künstlichen Sequenz: Plasmid  
pH/Si3xE2Fwt

<400> 9

```

gaattcccgg gaggtacaaa gcgcgcgaaa ctaaagcgcg cgaaactaaa gcgcgcgaaa 60
ctagctctta cgcggttcgcg aatcgatccg cggctctagaa attcctggca ttatcacata 120
atgaattata cattatataa agtaatgtga tttcttcgaa gaatatacta aaaaatgagc 180
aggcaagata aacgaaggca aagatgacag agcagaaagc cctagtaaag cgtattacaa 240
atgaaaccaa gattcagatt gcgatctctt taaaggggtg tcccctagcg atagagcact 300
cgatcttccc agaaaaagag gcagaagcag tagcagaaca ggccacacaa tcgcaagtga 360
ttaacgtcca cacaggtata gggtttctgg accatatgat acatgctctg gccaaagcatt 420
ccggctggtc gctaatcgtt gagtgcattg gtgacttaca catagacgac catcacacca 480
ctgaagactg cgggattgct ctccggtcaag cttttaaaga ggccctactg gcgctggtgag 540
taaaaagggt tggatcagga tttgcgctt tggatgaggc actttccaga gcggtggtag 600
atctttcгаа caggccgtac gcagttgtcg aacttggttt gcaaagggag aaagttaggag 660
atctctcttg cgagatgatc ccgcattttc ttgaaagctt tgcaaggct agcagaatta 720
ccctccacgt tgattgtctg cgaggcaaga atgatcatca ccgtagtgag agtgcgttca 780
aggctcttgc ggttgccata agagaagcca cctcgcccaa tggtagcaac gatgttccct 840
ccaccaaggg tgttcttatg tagtgacacc gattatttaa agctgcagca tacgatatat 900
atacatgtgt atatatgtat acctatgaat gtcagtaagt atgtatacga acagtatgat 960
actgaagatg acaaggtaat gcatcattct atacgtgtca ttctgaacga ggcgcgcttt 1020
ccttttttct ttttgctttt tctttttttt tctcttgaac tcgagaaaaa aatatataaa 1080
gagatggagg aacgggaaaa agttagttgt ggtgataggg ggcaagtggg attccgtaag 1140
aacaacaaga aaagcatttc atattatggc tgaactgagc gaacaagtgc aaaatttaag 1200
catcaacgac aacaacgaga atggttatgt tcctcctcac ttaagaggaa aaccaagaag 1260
tgccagaaat aacatgagca actacaataa caacaacggc ggctacaacg gtggccgtgg 1320
cgggtggcagc ttcttttagca acaaccgtcg tgggtggttac ggcaacggtg gtttcttcgg 1380
tggaacaac ggtggcagca gatctaacgg ccgttctggg gtagatgga tcgatggcaa 1440
acatgtccca gctccaagaa acgaaaaggc cgagatcgcc atatttgggtg tccccgagga 1500
tcctctacgc cggacgcacg gtggccggca tcaccggcgc cacaggtgcg gttgctggcg 1560
cctatatcgc cgacatcacc gatggggaag atcgggctcg ccacttcggg ctcatgagcg 1620
cttggttcgg cgtgggtatg gtggcaggcc ccgtggccgg gggactgttg ggcgccatct 1680
ccttgcatgc accattcctt gcggcggcgg tgctcaacgg cctcaaccta ctactgggct 1740
gcttcctaag gcaggagtcg cataagggag agcgtcgacc gatgcccttg agagccttca 1800
accagtcag ctccctccgg tgggcgcggg gcatgactat cgtcgccgca ctatgactg 1860
tcttctttat catgcaactc gtaggacagg tgccggcagc gctctgggtc attttcggcg 1920
aggaccgctt tcgctggagc gcgacgatga tcggcctgtc gcttgcggtc ttcggaatct 1980
tgcaacgcoct cgctcaagcc ttcgctactg gtcccgccac caaacgtttc ggcgagaagc 2040
aggccattat cgccggcatg gcggccgacg cgctgggcta cgtcttgctg gcgttcgcca 2100
cgcgaggctg gatggccttc cccattatga ttcttctcgc ttccggcggc atcgggatgc 2160
ccgcggttga ggccatgctg tccaggcagg tagatgacga ccatcaggga cagcttcaag 2220
gatcgtcgc ggctcttacc agcctaactt cgatcattgg accgctgatc gtcacggcga 2280
tttatgccgc ctccggcagc acatggaacg ggttggcatg gattgtaggc gccgccctat 2340
accttgtctg cctccccgcg ttgcgtcgcg gtgcatggag ccgggccacc tcgacctgaa 2400
tggaagccgg cggcacctcg ctaacggatt caccactcca agaattggag ccaatcaatt 2460
cttgcggaga actgtgaatg cgcaaaccaa cccttggcag aacatatcca tcgctgccgc 2520
catctccagc agccgcacgc ggcgcacgcg ggggggggtt tcaattcaat tcatcatttt 2580
ttttttattc ttttttttga tttcggtttc tttgaaattt ttttgattcg gtaatctccg 2640

```

aacagaagga agaacgaagg aaggagcaca gacttagatt ggtatatata cgcatatgta 2700  
gtgttgaaga aacatgaaat tgcccagtat tcttaaccca actgcacaga acaaaaacct 2760  
gcaggaaacg aagataaatc atgtcgaaaag ctacatataa ggaacgtgct gctactcatc 2820  
ctagtccctgt tgctgcccaag ctatttaata tcatgcacga aaagcaaaca aacttgtgtg 2880  
cttcattgga tgttcgtacc accaaggaat tactggagtt agttgaagca ttaggtccca 2940  
aaatltgttt actaaaaaca catgtggata tcttgactga tttttccatg gagggcacag 3000  
ttaagccgct aaaggcatta tccgcccaagt acaatltttt actcttcgaa gacagaaaat 3060  
ttgctgacat tggtaataca gtcaaattgc agtactctgc ggggtgtatac agaatagcag 3120  
aatgggcaga cattacgaat gcacacgggtg tgggtgggcc aggtattgtt agcgggttga 3180  
agcaggcggc agaagaagta acaaaggaac ctagaggcct tttgatgta gcagaattgt 3240  
catgcaaggg ctccctatct actggagaat atactaaggg tactgttgac attgcaaga 3300  
gcgacaaaga ttttgttatc ggctttattg ctcaaagaga catgggtgga agagatgaag 3360  
gttacgattg gttgattatg acaccgggtg tgggtttaga tgacaaggga gacgcattgg 3420  
gtcaacagta tagaacctgt gatgatgtgg tctctacagg atctgacatt attattgttg 3480  
gaagaggact atttgcaaag ggaagggatg ctaaggtaga ggggtaacgt tacagaaaag 3540  
caggctggga agcatatltt agaagatgcg gccagcaaaa ctaaaaaact gtattataag 3600  
taaatgcatg tataactaac tcacaaatta gagcttcaat ttaattatat cagttattac 3660  
ccctagagtc gacctgcagg catgcaagct tttgttccct ttagtgaggg ttaatlttca 3720  
gcttggcgta atcatggtca tagctgtttc ctgtgtgaaa ttgttatccg ctcaaatc 3780  
cacacaacat acgagccgga agcataaagt gtaaagcctg ggggtgcctaa tgagtgcagc 3840  
aactcacatt aattgcgttg cgtcactgc ccgctttcca gtcgggaaac ctgtcgtgcc 3900  
agctgcatta atgaatcggc caacgcgcgg ggagaggcgg tttgcgtatt gggcgctctt 3960  
ccgcttccct gctcactgac tcgctgcgct cggctgttcg gctgcggcga gcggtatcag 4020  
ctcactcaaa ggcggttaata cggttatcca cagaatcagg ggataacgca ggaaagaaca 4080  
tgtgagcaaa aggccagcaa aaggccagga accgtaaaaa ggccgcgttg ctggcgtttt 4140  
tccataggct ccgccccct gacgagcatc acaaaaatcg acgctcaagt cagaggtggc 4200  
gaaacccgac aggactataa agataaccagg cglttcccc tggaaagctcc ctctgcgct 4260  
ctcctgttcc gaccctgccg cttaccggat acctgtccgc ctttctccct tcgggaagcg 4320  
tggcgctttc tcatagctca cgctgtaggt atctcagttc ggtgtaggtc gttcgtcca 4380  
agctgggctg tgtgcacgaa cccccgttc agcccagccg ctgcgcctta tccggttaact 4440  
atcgtcttga gtccaaccg gtaagacacg acttatcgcc actggcagca gccactggta 4500  
acaggattag cagagcgagg tatgtaggcg gtgctacaga gttcttgaag tgggtggccta 4560  
actacggcta cactagaagg acagtatltt gtatctgcgc tctgctgaag ccagttacct 4620  
tcggaaaaag agttggtagc tcttgatccg gcaaaaaac caccgctggg agcgggtggt 4680  
ttttgtttg caagcagcag attacgcgca gaaaaaagg atctcaagaa gatcctttga 4740  
tcttttctac ggggtctgac gctcagtgga acgaaaactc acgtaaggg attttggta 4800  
tgagattatc aaaaaggatc ttcacctaga tcctltttaa ttaaaaatga agttttaa 4860  
caatctaaag tatatatgag taaacttggg ctgacagtta ccaatgctta atcagtgagg 4920  
cacctatctc agcgatctgt ctatltcgtt catccatagt tgcctgactc cccgtcgtgt 4980  
agataactac gatacgggag ggcttaccat ctggccccag tgctgcaatg ataccgcgag 5040  
accacgctc accggctcca gatttatcag caataaacca gccagccgga agggccgagc 5100  
gcagaagtgg tcctgcaact ttatccgcct ccatccagtc tattaattgt tgccgggaag 5160  
ctagagtaag tagttcgcca gttaatagtt tgcgcaacgt tgttgccatt gctacaggca 5220  
tcgtgggtgc acgctcgtcg tttgggtatgg cttcattcag ctccggttcc caacgatcaa 5280  
ggcgagttac atgatcccc atgttgtgca aaaaagcggg tagctccttc ggtcctccga 5340  
tcgttgtcag aagtaagttg gccgcagtggt tatcactcat gggtatggca gcaactgcata 5400  
attctcttac tgtcatgcca tccgtaagat gctlttctgt gactggtgag tactcaacca 5460  
agtcattctg agaatagtgt atgcggcgac cgagttgctc ttgcccggcg tcaatacggg 5520

```

ataataccgc gccacatagc agaactttaa aagtgctcat cattggaaaa cgttcttcgg 5580
ggcgaaaact ctcaaggatc ttaccgctgt tgagatccag ttcgatgtaa cccactcgtg 5640
cacccaactg atcttcagca tcttttactt tcaccagcgt ttctgggtga gcaaaaacag 5700
gaaggcaaaa tgccgcaaaa aagggaataa gggcgacacg gaaatggtga atactcatac 5760
tcttcctttt tcaatattat tgaagcattt atcaggggta ttgtctcatg agcggataca 5820
tatttgaatg tatttagaaa aataaaciaa taggggttcc gcgcacattt ccccgaaaag 5880
tgccacctga cgtctaagaa accattatta tcatgacatt aacctataaa aataggcgta 5940
tcacgaggcc ctttcgtctc gcgcggttctg gtgatgacgg tgaaacctc tgacacatgc 6000
agctcccgga gacggtcaca gcttgtctgt aagcggatgc cgggagcaga caagcccgtc 6060
agggcgcgtc agcgggtggt ggcgggtgtc ggggctggct taactatgcg gcatcagagc 6120
agattgtact gagagtgcac catatgcggt gtgaaatacc gcacagatgc gtaaggagaa 6180
aataccgcat caggaaattg taaacgttaa tattttgta aaattcgcgt taaattttg 6240
ttaaatacgc tcatttttta accaataggc cgaaatcggc aaaatccctt ataaatcaaa 6300
agaatagacc gagatagggg tgagtgttgt tccagtttg aacaagagtc cactattaa 6360
gaacgtggac tccaacgtca aaggcgaaa aaccgtctat caggcgatg gccactacg 6420
tgaacctca ccctaataca gttttttggg gtcgaggtgc cgtaaagcac taaatcgaa 6480
ccctaaaggg agccccgat ttagagcttg acggggaaaag ccggcgaacg tggcgagaaa 6540
ggaagggaag aaagcgaaaag gagcgggcgc tagggcgctg gcaagtgtag cggtcacgct 6600
gcgcgtaacc accacaccgc cgcgccttaa tgcgcgcta cagggcgcgt cgcccatc 6660
gccattcagg ctgcgcaact gttgggaagg gcgatcgggtg cgggcctctt cgctattacg 6720
ccagctggcg aaagggggat gtgctgcaag gcgattaagt tgggtaacgc cagggttttc 6780
ccagtcacga cgttgtaaaa cgacggccag tgaattgtaa tacgactcac tatagggc 6838

```

<210> 10

<211> 175

<212> DNA

<213> Künstliche Sequenz

<220>

<223> Beschreibung der künstlichen Sequenz: Peptidomäne  
codierende Nukleinsäure

<400> 10

```

gggaattcga tcctgctgct ctaaaacgtg ctaganncac tgaannnnn aggcgtnnc 60
gtgcgagaaa gttgcaactt gaagacaagg ttgaagaatt gctttcgaaa aattatcact 120
tggaaaatga ggttgccaga ttaaagaaat tagttggcga acgctgagga tcccc 175

```

<210> 11

<211> 6659

<212> DNA

<213> Künstliche Sequenz

<220>

<223> Beschreibung der künstlichen Sequenz: Vektor  
pGADD424

<400> 11  
 gggaattcga tcctgctgct ctaaaacgtg ctaganncac tgaannnnnn aggcgtnnnc 60  
 gtgcgagaaa gttgcaactt gaagacaagg ttgaagaatt gctttcgaaa aattatcact 120  
 tggaaaatga ggttgccaga ttaaagaaat tagttggcga acgctgagga tcccc 175

<210> 12  
 <211> 22  
 <212> DNA  
 <213> Künstliche Sequenz

<220>  
 <223> Beschreibung der künstlichen Sequenz: Primer

<400> 12  
 gatgtatata actatctatt cg 22

<210> 13  
 <211> 20  
 <212> PRT  
 <213> Künstliche Sequenz

<220>  
 <223> Beschreibung der künstlichen Sequenz:  
 Konsensussequenz

<400> 13  
 Asp Pro Ala Ala Leu Lys Arg Ala Arg Gly Thr Glu Val Val Arg Arg  
 1 5 10 15  
 Gly Arg Ala Arg  
 20

<210> 14  
 <211> 20  
 <212> PRT  
 <213> Künstliche Sequenz

<220>  
 <223> Beschreibung der künstlichen Sequenz:  
 3xE2Fwt-Sequenz erkennende Domäne

<400> 14  
 Asp Pro Ala Ala Leu Lys Arg Ala Arg Cys Thr Glu Val Met Arg Arg  
 1 5 10 15

Gln Arg Ala Arg  
20

<210> 15  
<211> 20  
<212> PRT  
<213> Künstliche Sequenz

<220>  
<223> Beschreibung der künstlichen Sequenz:  
3xE2Fwt-Sequenz erkennende Domäne

<400> 15  
Asp Pro Ala Ala Leu Lys Arg Ala Arg Pro Thr Glu Asn Val Arg Arg  
1 5 10 15

Gly Arg Ala Arg  
20

<210> 16  
<211> 20  
<212> PRT  
<213> Künstliche Sequenz

<220>  
<223> Beschreibung der künstlichen Sequenz:  
3xE2Fwt-Sequenz erkennende Domäne

<400> 16  
Asp Pro Ala Ala Leu Lys Arg Ala Arg Asn Thr Glu Val Thr Arg Arg  
1 5 10 15

Ser Arg Ala Arg  
20

<210> 17  
<211> 20  
<212> PRT  
<213> Künstliche Sequenz

<220>  
<223> Beschreibung der künstlichen Sequenz:  
3xE2Fwt-Sequenz erkennende Domäne

<400> 17

Asp Pro Ala Ala Leu Lys Arg Ala Arg Val Thr Glu Asn Ser Arg Arg  
 1                      5                      10                      15

Asp Arg Ala Arg  
                     20

<210> 18  
 <211> 20  
 <212> PRT  
 <213> Künstliche Sequenz

<220>  
 <223> Beschreibung der künstlichen Sequenz:  
           3xE2Fwt-Sequenz erkennende Domäne

<400> 18  
 Asp Pro Ala Ala Leu Lys Arg Ala Arg His Thr Glu Thr Ser Arg Arg  
 1                      5                      10                      15

Ile Arg Ala Arg  
                     20

<210> 19  
 <211> 20  
 <212> PRT  
 <213> Künstliche Sequenz

<220>  
 <223> Beschreibung der künstlichen Sequenz:  
           3xE2Fwt-Sequenz erkennende Domäne

<400> 19  
 Asp Pro Ala Ala Leu Lys Arg Ala Arg Val Thr Glu Val Ile Arg Arg  
 1                      5                      10                      15

Gly Arg Ala Arg  
                     20

<210> 20  
 <211> 20  
 <212> PRT  
 <213> Künstliche Sequenz

<220>  
 <223> Beschreibung der künstlichen Sequenz:

## 3xE2Fwt-Sequenz erkennende Domäne

&lt;400&gt; 20

Asp Pro Ala Ala Leu Lys Arg Ala Arg Ile Thr Glu Gly Ile Arg Arg  
 1                    5                    10                    15

Leu Arg Ala Arg  
                   20

&lt;210&gt; 21

&lt;211&gt; 20

&lt;212&gt; PRT

&lt;213&gt; Künstliche Sequenz

&lt;220&gt;

<223> Beschreibung der künstlichen Sequenz:  
 3xE2Fwt-Sequenz erkennende Domäne

&lt;400&gt; 21

Asp Pro Ala Ala Leu Lys Arg Ala Arg Ser Thr Glu Leu Asn Arg Arg  
 1                    5                    10                    15

Gly Arg Ala Arg  
                   20

&lt;210&gt; 22

&lt;211&gt; 20

&lt;212&gt; PRT

&lt;213&gt; Künstliche Sequenz

&lt;220&gt;

<223> Beschreibung der künstlichen Sequenz:  
 3xE2Fwt-Sequenz erkennende Domäne

&lt;400&gt; 22

Asp Pro Ala Ala Leu Lys Arg Ala Arg Gly Thr Glu Arg Leu Arg Arg  
 1                    5                    10                    15

Gly Arg Ala Arg  
                   20

&lt;210&gt; 23

&lt;211&gt; 20

&lt;212&gt; PRT

&lt;213&gt; Künstliche Sequenz

&lt;220&gt;

&lt;223&gt; Beschreibung der künstlichen Sequenz:

3xE2Fwt-Sequenz erkennende Domäne

&lt;400&gt; 23

Asp Pro Ala Ala Leu Lys Arg Ala Arg Gly Thr Glu Ala Thr Arg Arg  
 1                    5                    10                    15

Val Arg Ala Arg  
                   20

&lt;210&gt; 24

&lt;211&gt; 20

&lt;212&gt; PRT

&lt;213&gt; Künstliche Sequenz

&lt;220&gt;

&lt;223&gt; Beschreibung der künstlichen Sequenz:

3xE2Fwt-Sequenz erkennende Domäne

&lt;400&gt; 24

Asp Pro Ala Ala Leu Lys Arg Ala Arg Cys Thr Glu Glu Val Arg Arg  
 1                    5                    10                    15

Trp Arg Ala Arg  
                   20

&lt;210&gt; 25

&lt;211&gt; 20

&lt;212&gt; PRT

&lt;213&gt; Künstliche Sequenz

&lt;220&gt;

&lt;223&gt; Beschreibung der künstlichen Sequenz:

3xE2Fwt-Sequenz erkennende Domäne

&lt;400&gt; 25

Asp Pro Ala Ala Leu Lys Arg Ala Arg Cys Thr Glu Val Gln Arg Arg  
 1                    5                    10                    15

Gly Arg Ala Arg  
                   20

&lt;210&gt; 26



<210> 29

<211> 20

<212> PRT

<213> Künstliche Sequenz

<220>

<223> Beschreibung der künstlichen Sequenz:

3xE2Fwt-Sequenz erkennende Domäne

<400> 29

Asp Pro Ala Ala Leu Lys Arg Ala Arg Gly Thr Glu Val Val Arg Arg  
 1                    5                    10                    15

Cys Arg Ala Arg  
                   20

<210> 30

<211> 20

<212> PRT

<213> Künstliche Sequenz

<220>

<223> Beschreibung der künstlichen Sequenz:

3xE2Fwt-Sequenz erkennende Domäne

<400> 30

Asp Pro Ala Ala Leu Lys Arg Ala Arg Ile Thr Glu Asn Ala Arg Arg  
 1                    5                    10                    15

Gly Arg Ala Arg  
                   20

<210> 31

<211> 20

<212> PRT

<213> Künstliche Sequenz

<220>

<223> Beschreibung der künstlichen Sequenz:

3xE2Fwt-Sequenz erkennende Domäne

<400> 31

Asp Pro Ala Ala Leu Lys Arg Ala Arg Cys Thr Glu Glu Met Arg Arg  
 1                    5                    10                    15

Gly Arg Ala Arg  
20

<210> 32  
<211> 20  
<212> PRT  
<213> Künstliche Sequenz

<220>  
<223> Beschreibung der künstlichen Sequenz:  
3xE2Fwt-Sequenz erkennende Domäne

<400> 32  
Asp Pro Ala Ala Leu Lys Arg Ala Arg Cys Thr Glu Pro Ser Arg Arg  
1 5 10 15

Gly Arg Ala Arg  
20

<210> 33  
<211> 20  
<212> PRT  
<213> Künstliche Sequenz

<220>  
<223> Beschreibung der künstlichen Sequenz:  
3xE2Fwt-Sequenz erkennende  
Domäne

<400> 33  
Asp Pro Ala Ala Leu Lys Arg Ala Arg Asn Thr Glu Ser Gly Arg Arg  
1 5 10 15

Thr Arg Ala Arg  
20

<210> 34  
<211> 20  
<212> PRT  
<213> Künstliche Sequenz

<220>  
<223> Beschreibung der künstlichen Sequenz:  
3xE2Fwt-Sequenz erkennende Domäne



&lt;220&gt;

<223> Beschreibung der künstlichen Sequenz:  
3xE2Fwt-Sequenz erkennende Domäne

&lt;400&gt; 37

Asp Pro Ala Ala Leu Lys Arg Ala Arg Ile Thr Glu Asn Ala Arg Arg  
1                    5                    10                    15

Gly Arg Ala Arg  
                  20

&lt;210&gt; 38

&lt;211&gt; 20

&lt;212&gt; PRT

&lt;213&gt; Künstliche Sequenz

&lt;220&gt;

<223> Beschreibung der künstlichen Sequenz:  
3xE2Fwt-Sequenz erkennende Domäne

&lt;400&gt; 38

Asp Pro Ala Ala Leu Lys Arg Ala Arg Ile Thr Glu Met Gly Arg Arg  
1                    5                    10                    15

Lys Arg Ala Arg  
                  20

&lt;210&gt; 39

&lt;211&gt; 20

&lt;212&gt; PRT

&lt;213&gt; Künstliche Sequenz

&lt;220&gt;

<223> Beschreibung der künstlichen Sequenz:  
3xE2Fwt-Sequenz erkennende Domäne

&lt;400&gt; 39

Asp Pro Ala Ala Leu Lys Arg Ala Arg Cys Thr Glu Tyr Cys Arg Arg  
1                    5                    10                    15

Ile Arg Ala Arg  
                  20

&lt;210&gt; 40

&lt;211&gt; 20

<212> PRT

<213> Künstliche Sequenz

<220>

<223> Beschreibung der künstlichen Sequenz:

3xE2Fwt-Sequenz erkennende Domäne

<400> 40

Asp	Pro	Ala	Ala	Leu	Lys	Arg	Ala	Arg	Gly	Thr	Glu	Glu	Tyr	Arg	Arg
1				5					10					15	

His	Arg	Ala	Arg
			20

<210> 41

<211> 20

<212> PRT

<213> Künstliche Sequenz

<220>

<223> Beschreibung der künstlichen Sequenz:

3xE2Fwt-Sequenz erkennende Domäne

<400> 41

Asp	Pro	Ala	Ala	Leu	Lys	Arg	Ala	Arg	Ser	Thr	Glu	Leu	Thr	Arg	Arg
1				5					10					15	

Ile	Arg	Ala	Arg
			20

<210> 42

<211> 20

<212> PRT

<213> Künstliche Sequenz

<220>

<223> Beschreibung der künstlichen Sequenz:

3xE2Fwt-Sequenz erkennende Domäne

<400> 42

Asp	Pro	Ala	Ala	Leu	Lys	Arg	Ala	Arg	Gly	Thr	Glu	Pro	Val	Arg	Arg
1				5					10					15	

Ser	Arg	Ala	Arg
			20

<210> 43

<211> 20

<212> PRT

<213> Künstliche Sequenz

<220>

<223> Beschreibung der künstlichen Sequenz:

3xE2Fwt-Sequenz erkennende Domäne

<400> 43

Asp	Pro	Ala	Ala	Leu	Lys	Arg	Ala	Arg	Gly	Thr	Glu	Ala	Glu	Arg	Arg
1				5					10					15	

Gly	Arg	Ala	Arg
			20

<210> 44

<211> 20

<212> PRT

<213> Künstliche Sequenz

<220>

<223> Beschreibung der künstlichen Sequenz:

3xE2Fwt-Sequenz erkennende Domäne

<400> 44

Asp	Pro	Ala	Ala	Leu	Lys	Arg	Ala	Arg	Asp	Thr	Glu	Ala	Ser	Arg	Arg
1				5					10					15	

Met	Arg	Ala	Arg
			20

<210> 45

<211> 20

<212> PRT

<213> Künstliche Sequenz

<220>

<223> Beschreibung der künstlichen Sequenz:

3xE2Fwt-Sequenz erkennende Domäne

<400> 45

Asp	Pro	Ala	Ala	Leu	Lys	Arg	Ala	Arg	Asn	Thr	Glu	Ser	Gly	Arg	Arg
1				5					10					15	

Thr Arg Ala Arg

20