

(12)

Patentschrift

(21) Anmeldenummer: A 9496/2010 (51) Int. Cl.: **C12N 15/10** (2006.01)
(86) PCT-Anmeldenummer: PCT/SI2010/000058 **C12N 15/52** (2006.01)
(22) Anmeldetag: 22.10.2010 **C12N 15/62** (2006.01)
(45) Veröffentlicht am: 15.05.2015

(56) Entgegenhaltungen:
WO 9856904 A1

(73) Patentinhaber:
KEMIJSKI INSTITUT
1000 Ljubljana (SI)
EN-FIST center odlicnosti
1000 Ljubljana (SI)

(72) Erfinder:
Lebar Tina
1295 Ivancna Gorica (SI)
Fortsetzung siehe Beiblatt!

(74) Vertreter:
SONN & PARTNER PATENTANWÄLTE
WIEN

(54) **Verbesserte Synthese eines biosynthetischen Produkts mit einer angeordneten Zusammensetzung der biosynthetischen Enzyme auf einem Nukleotidsequenzmotiv**

(57) Die Erfindung bezieht sich auf eine Methode zur Synthese einer Verbindung mit Hilfe von biosynthetischen Wegen mit der Kultivierung von genetisch veränderten Wirtszellen, wobei die Zellen mit einer oder mehreren Nukleinsäuren verändert werden, wobei die Nukleinsäure die Nukleotidsequenzen für chimäre Proteine beinhalten, die aus Enzymen des biosynthetischen Wegs oder anderen funktionellen Polypeptiden sowie Faktoren, die sich an Nukleinsäure binden, zusammengesetzt sind, und die Programm-Nukleinsäuresequenz, die Zielelemente für Nukleinsäure enthalten, die von den genannten Faktoren des chimären Proteins erkannt werden, wobei sich die Faktoren an die Nukleinsäure binden und die genannte Programm-Nukleotidsequenz zur Anordnung der Enzyme des biosynthetischen Wegs oder anderen funktionellen Polypeptiden entlang der Programm-Nukleinsäuresequenz dient und die genannte Kultivierung zur Synthese von genannten chimären Proteinen in genetisch veränderten Wirtszellen dient, was in der Produktion der genannten Verbindung resultiert. Die Erfindung bezieht sich auf Wirtszellen mit der Programm-Nukleinsäuresequenz, mit der Nukleinsäuresequenz, die die chimären Proteine exprimiert. Die Erfindung beschreibt die Methoden zu einer besseren Synthese der Verbindung.

(72) Erfinder: (Fortsetzung)

Turnsek Jernej
3310 Zalec (SI)
Ilc Tina
5270 Ajdovscina (SI)
Tomsic Nejc
1210 Ljubljana-Sentvid (SI)
Jerala Roman
1000 Ljubljana (SI)
Avbelj Monika
1000 Ljubljana (SI)
Bencina Mojca
1000 Ljubljana (SI)
Mori Jerneja
1000 Ljubljana (SI)
Stosicki Tjasa
1410 Zagorje ob Savi (SI)
Znidaric Matej
1000 Ljubljana (SI)
Bordon Jure
1000 Ljubljana (SI)
Petroni Mattia
34077 Ronchi dei Legionari (Gorizia) (IT)
Gaber Rok
1000 Ljubljana (SI)
Koprivnjak Tomaz
1133 Ljubljana (SI)
Anderluh Gregor
1000 Ljubljana (SI)
Vovk Irena
1000 Ljubljana (SI)
Glavnik Vesna
1000 Ljubljana (SI)

Beschreibung

GEBIET DER ERFINDUNG:

[0001] Die vorliegende Erfindung betrifft das Gebiet einer verbesserten Produktsynthese, die durch die gerichtete Anordnung von Chimären Proteinen erreicht wird, wobei sich die chimären Proteine aus Biosyntheseenzymen oder weiteren funktionellen Polypeptiden zusammensetzen, die an Nukleinsäure bindende Faktoren gebunden sind, wobei die gerichtete Anordnung der chimären Proteine auf der Bindung an Nukleinsäure-Erkennungsmotive auf der Programm-Nukleinsäuresequenz basiert. Bei der vorliegenden Erfindung handelt es sich um eine biotechnologische Erfindung.

STAND DER TECHNIK:

[0002] Für industrielle Anwendungen werden Biosynthesewege, die sich aus einer Vielzahl an Enzymen und weiteren Proteinen zusammensetzen, dahingehend manipuliert, dass hohe Ausbeuten der gewünschten Biosyntheseprodukte erzielt werden können. Die Verbesserung der Effizienz von Biosynthesewegen, um in kürzerer Zeit höhere Ausbeuten an Reaktionsprodukten zu erzielen, ist derzeit von großem Interesse. Es wurden bereits etliche Strategien bezüglich einer dahingehenden Optimierung umgesetzt. So wurde z. B. die Verbesserung der Ausbeute des Endprodukts einer Biosynthesereaktion bereits erzielt durch: (i) eine Vergrößerung des Pools an verfügbarem Substrat und/oder die Überexpression der Enzyme der limitierenden Biosyntheseschritte; (ii) das Einbringen von heterologen Enzymen mit bevorzugten kinetischen Eigenschaften; (iii) das Blockieren von Verzweigungen des Biosynthesewegs; (iv) die Kompartimentalisierung von Biosynthesewegen, indem Enzyme eines bestimmten Biosynthesewegs zu spezifischen Zellkompartimenten oder künstlich geschaffenen Kompartimenten (z. B. Metabolomen) dirigiert werden, oder; (v) die räumliche Annäherung von Enzymen durch die Anordnung von metabolischen Synthesewegen auf einem Proteingerüst (WO 2009/108774). Zwar weisen proteinbasierte Gerüste zahlreiche Vorteile auf, jedoch ist dieser Ansatz durch die Anzahl an verfügbaren Kombinationen von Docking-Peptiden beschränkt und bietet keine Möglichkeit zur Kontrolle der räumlichen Verteilung und Orientierung von Biosyntheseenzymen, um so den Biosyntheseweg in optimaler Weise zu unterstützen.

[0003] Keine der vorstehend genannten Lösungen kann eine optimale Anordnung der Enzyme des Biosynthesewegs gewährleisten, um so eine gewünschte Reihenfolge der Biosynthesereaktionen zur Verfügung zu stellen. In lebenden Zellen sind die Enzyme des Biosynthesewegs oder weitere funktionelle Polypeptide oftmals durch Multiproteinkomplexe oder Verankerungsmechanismen auf eine Stelle einer Größenordnung im Mikrometerbereich festgelegt, auf die im Folgenden als "Mikrobereich" Bezug genommen wird. Durch diese Art der Organisation kommt es zu einer Erhöhung der lokalen Konzentration von Enzymen sowie der Effizienz von Biosynthesewegen. Zum Erreichen einer optimalen Konzentration an individuellen Enzymen sowie zur Förderung der Ausbildung von vollständigen Multiproteinkomplexen wurden bisher zwei Ansätze angewandt: (i) das Dirigieren von Enzymen in Mikrobereiche, z. B. zu Zielorganellen; und (ii) das Ausbilden von Multiproteinkomplexen als Polypeptidgerüst (WO 2009/108774).

[0004] Das Proteingerüst weist jedoch zahlreichen Nachteile und Einschränkungen auf.

[0005] Zwar wäre es möglich, die Anzahl und Verteilung der Enzyme in einem Multiproteinkomplex anhand der Sequenz eines Polypeptidrückgrats zu programmieren, jedoch ist es aufgrund der Flexibilität der Dimerisierungsdomänen an sich sowie im Verhältnis zueinander nicht möglich, die dreidimensionale Anordnung der Polypeptide zu präzisieren (Figuren 1 und 2). Des Weiteren kann sich die Konstruktion des Polypeptidrückgrats mit den an dem Gerüst orientierten Proteindomänen aufgrund der begrenzten Anzahl an verfügbaren Proteindimerisierungsdomänen als schwierig erweisen. Darüber hinaus gibt es für jede einzelne Proteindimerisierungsdomäne unterschiedliche Bedingungen, unter denen sich diese faltet und die funktionelle Interaktion ausbildet.

ZUSAMMENFASSUNG DER ERFINDUNG:

[0006] Das Problem der Organisation des Biosynthesewegs in eine geordnete Reihenfolge von Biosyntheseenzymen, das durch die Proteindimerisierungsdomäne nicht gelöst werden kann, wird durch ein Nukleotidsequenzmotiv-Template gelöst, das hierin auch als Programm- Nukleinsäuresequenz bezeichnet wird und das die Reihenfolge der chimären Proteine, Biosyntheseenzyme und weiteren funktionellen Polypeptide bestimmt, die an Nukleinsäure bindende Faktoren gebunden sind.

[0007] Im Rahmen der vorliegenden Erfindung wurde nun entdeckt, dass es möglich ist, die dreidimensionale Anordnung von Biosyntheseenzymen oder weiteren funktionellen Polypeptiden durch eine so genannten Programm-Nukleinsäure zu manipulieren. Im Rahmen der vorliegenden Erfindung wurden chimäre Polypeptide konstruiert, die sich aus einer Nukleinsäure bindenden Domäne eines Nukleinsäure bindenden Faktors sowie aus einem Biosyntheseenzym oder einem weiteren funktionellen Polypeptid zusammensetzen. Die Bindung dieser chimären Proteine an das Zielnukleinsäuremotiv erfolgt über einen Nukleinsäure bindenden Faktor. Die Programm-Nukleinsäuresequenz legt die Anordnung der Zielnukleinsäuremotive und somit auch die Reihenfolge und die räumliche Anordnung von Biosyntheseenzymen oder weiteren funktionellen Polypeptiden fest. Durch einen auf diese Weise gerichteten Biosyntheseweg aus chimären Biosyntheseenzymen oder weiteren funktionellen Polypeptiden, gesteuert durch die Programm-Nukleinsäuresequenz, ist es möglich, eine erhöhte Ausbeute eines Biosyntheseprodukts eines solchen Synthesewegs im Vergleich zur Verwendung eines Proteingerüsts zu erzielen, in welchem die Biosyntheseenzyme lediglich ohne eine spezifische Anordnung angehäuft werden.

[0008] Durch die Verwendung einer Programm-Nukleinsäuresequenz zur Immobilisierung und Festlegung der Reihenfolge der Enzyme des Biosynthesewegs oder weiterer funktioneller Polypeptide werden ein oder mehrere Elemente der folgenden Auflistung zur Verfügung gestellt: i) eine Steigerung der Effizienz des Flusses des Synthesewegs; ii) eine Optimierung des metabolischen Flusses durch den Syntheseweg; iii) eine Verringerung der metabolischen Belastung für die Wirtszelle; iv) die Möglichkeit einer alternativen Verzweigung von metabolischen Synthesewegen, die zum Erhalt von neuen Produkten des Biosyntheseprozesses führt; und v) die Möglichkeit der Neuprogrammierung der chimären Enzyme innerhalb des Organismus durch das Einbringen von unterschiedlichen Programm-Nukleinsäuren.

[0009] Die Programm-Nukleinsäuresequenz trägt mehr als zwei Zielnukleinsäureelemente, bei denen es sich jeweils um spezifische Liganden für Nukleinsäure bindende Faktoren handelt. Liegt die Programm-Nukleinsäuresequenz in einer Lösung mit chimären Proteinen vor, so binden die chimären Proteine an Zielelemente innerhalb der Programmsequenz, bilden einen Nukleinsäure-Protein-Komplex und unterstützen die Bildung eines Multiproteinkomplexes. Die Reihenfolge von chimären Proteinen in einem derartigen Multiproteinkomplex kann auf Grundlage der folgenden Umstände prädiziert werden: i) die Positionierung der Zielnukleinsäuremotive ist aufgrund der bekannten dreidimensionalen Struktur der Nukleinsäuren auf Grundlage der Nukleinsäuresequenz definiert; ii) durch den Umstand, dass die Reihenfolge der Nukleinsäuremotive prädiziert werden kann, wird die gerichtete Positionierung von chimären Proteinen unterstützt; iii) durch eine derart gerichtete Positionierung von chimären Proteinen wird ermöglicht, dass die Reihenfolge der Biosynthesereaktionen durch die Programm-Nukleinsäuresequenz definiert werden kann und es kommt somit zu einer Verkürzung des Zeitintervalls, innerhalb dessen es zwischen den verschiedenen Schritten der Biosynthesereaktion zu einer Diffusion von Zwischenprodukten der Reaktion kommen kann. Im Rahmen der vorliegenden Erfindung wurde daher zu dem Schluss gelangt, dass es von Vorteil ist, eine große Anzahl an unterschiedlichen Nukleotid bindenden Domänen zur Verfügung zu haben, um komplexe Reaktionswege aufzubauen. So gibt es z. B. 262.144 unterschiedliche Kombinationen von Nukleotidmotiven, die aus neun Nukleotiden bestehen und durch unterschiedliche Nukleotid bindende Domänen erkannt werden können, was eine extrem hohe Variabilität bei der Assemblierung unterschiedlicher Biosynthesewege gestattet.

[0010] Die vorliegende Erfindung weist im Vergleich zu der Verwendung von Proteingerüsten eine Reihe von Vorteilen auf. Bei der Programm-Nukleinsäuresequenz kommt es zu keinen Schwierigkeiten hinsichtlich der Reifung und das gerichtete Nukleotidbindungsmotiv kann beliebig ausgewählt werden. Der Vorgang des so genannten Dockings des Anker-Nukleinsäure-Bindungsfaktors an das Ziehnukleinsäureelement wurde bereits ausführlich beschrieben. Aufgrund der großen räumlichen Nähe zwischen den chimären Proteinen, die an die Programm-Nukleinsäuresequenz gebunden sind, werden weitere Polypeptide, die die Synthese unter Umständen umlenken könnten, in räumlicher Hinsicht von dem Multiproteinkomplex ausgeschlossen. Die Bindung aller Komponenten des Biosynthesewegs erfolgt jeweils unter gleichartigen Reaktionsbedingungen, da alle für die Konstruktion von chimären Proteinen verwendeten Nukleinsäure-Bindungsdomänen auf der gleichen Proteinfaltung beruhen und auf die gleiche Weise mit den Nukleotiden interagieren können. Im Gegensatz hierzu ist bei der Verwendung eines Proteingerüsts der Einsatz von unterschiedlichen Arten von Proteindomänen-Dimeren erforderlich, die unter jeweils unterschiedlichen Bedingungen zu einer korrekten Faltung und einer optimalen Interaktion in der Lage sind. Die vorliegende Erfindung ist dahingehend vorteilhaft, dass sie dazu in der Lage ist, die Produktion unterschiedlicher Reaktionsprodukte in Abhängigkeit von der Sequenz der Programm-Nukleinsäure zu steuern, die jeweils in die produzierenden Zellen oder Reaktionsmedien eingebracht wird, wobei jede der unterschiedlichen Programm-DNAs die Enzyme des Biosynthesewegs jeweils unterschiedlich anordnet.

[0011] Die vorliegende Erfindung stellt ein Verfahren zur Herstellung einer Verbindung über einen Biosyntheseweg bereit, wobei das Verfahren das Kultivieren einer genetisch modifizierten Wirtszelle, die dahingehend verändert ist, um chimäre Proteine, Enzyme des Biosynthesewegs oder weitere funktionelle Polypeptide zu exprimieren, die an Nukleinsäure bindende Faktoren gebunden sind, sowie eine Programm-Nukleinsäuresequenz einschließt, die die gerichtete Anordnung von chimären Proteinen zur Ausbildung eines biosynthetischen Multiproteinkomplexes steuert.

[0012] Die vorliegende Erfindung betrifft Vorgänge, die von mindestens drei bis zu etwa 100 biosynthetische Reaktionsschritte einschließen. Drei entspricht der Mindestanzahl an Schritten, wobei die Reihenfolge der Reaktionsvorgänge variieren kann und die im Rahmen der vorliegenden Erfindung beschriebene gerichtete Anordnung zur Wirkung kommt. Die drei Schritte können z. B. der Reihenfolge der Reaktionsschritte 1, 2, 3 entsprechen, im Vergleich zu anderen Reihenfolgen wie 1, 3, 2 oder 3, 1, 2.

[0013] In einer weiteren Ausführungsform der vorliegenden Erfindung schließt das Verfahren ein: das Mischen der produzierten chimären Proteine, Enzyme des Biosynthesewegs oder weiteren funktionellen Polypeptide, die an Nukleinsäure bindende Faktoren gebunden sind, die in den Wirtszellen produziert und aus diesen isoliert werden, sowie eine Programm-Nukleinsäuresequenz, die die Reihenfolge von chimären Proteinen zur Ausbildung eines biosynthetischen Multiproteinkomplexes steuert. Auf diese Weise wird der Biosyntheseweg in vitro assembliert, und zwar entweder in Lösung oder mit einer an einem weiteren Molekül oder der festen Phase immobilisierten Programm-DNA. Zur Durchführung der Biosynthesereaktion muss der in vitro assemblierte Komplex unter Zugabe der für die Biosynthesereaktion erforderlichen Substrate und Cofaktoren oder des Enzymsystems, das die erforderlichen Cofaktoren generiert, unter Bedingungen inkubiert werden, die die Biosynthese unterstützen.

[0014] Die vorliegende Erfindung stellt eine Beschreibung zur Konstruktion von Programm-Nukleinsäuren zur Verfügung, die sowohl in vivo als auch in vitro angewandt werden kann. Bereitgestellt werden genetisch modifizierte Wirtszellen, umfassend eine Programm-Nukleinsäuresequenz sowie eine Nukleinsäuresequenz, die für chimäre Proteine, Biosynthesenzyme oder weitere funktionelle Polypeptide kodiert, die kovalent an einen Anker-Nukleinsäure-Bindungsfaktor gebunden sind und die an spezifische Ziehnukleinsäureelemente binden, die in der Programm-Nukleinsäuresequenz kodiert sind.

[0015] Die vorliegende Erfindung stellt Nukleinsäuren, die eine Programm-Nukleinsäuresequenz umfassen, zur Verwendung in einem Verfahren zur Verbesserung der Ausbeute eines

Biosyntheseprodukts zur Verfügung.

[0016] Durch die vorliegende Erfindung wird ermöglicht, biologische Prozesse auf spezifische Endprodukte auszurichten, indem aufgrund der großen räumlichen Nähe zwischen funktionellen Gruppen der Chimären Proteine, die an die Programm-Nukleinsäure gebunden sind, weitere Polypeptide ausgeschlossen werden, welche die Synthese in Richtung von nicht erwünschten Produkten umlenken könnten.

[0017] Die vorliegende Erfindung betrifft genetisch modifizierte Wirtszellen, die chimäre Proteine exprimieren und/oder die Programm-Nukleinsäuresequenz replizieren.

[0018] Die vorliegende Erfindung betrifft des Weiteren ein Verfahren, das zur Herstellung einer Verbindung eines beliebigen Biosynthesewegs eingesetzt wird, bei dem zur Synthese der betreffenden Verbindung eine große räumliche Nähe zwischen Biosyntheseenzymen erforderlich ist.

[0019] Die vorliegende Erfindung betrifft des Weiteren Prozesse wie z. B. Informations- bzw. Datenverarbeitung, chemischen Abbau, Signalgebung und Biosensorik.

KURZBESCHREIBUNG DER ZEICHNUNGEN:

[0020] Figur 1. Schematische Übersicht der Vorteile der Verwendung einer Programm-Nukleinsäuresequenz.

[0021] Figur 2. Auswirkung der Anordnung einer Spacer-Sequenz zwischen Zielnukleinsäureelementen bzw. -motiven.

[0022] Figur 3. Bindung von chimären Proteinen, an DNA bindende Faktoren gebundenen Biosyntheseenzymen, an das Zielnukleinsäuremotiv. Spezifische Bindung eines jeden der getesteten DNA bindende Faktoren Znf_Zif268, Znf_Blues, Znf_PBSII, Znf_HivC und NicTAL (Tabelle 1) an das Zielnukleinsäuremotiv sowie Unterdrückung der Expression eines Reporterproteins, der β -Galactosidase. [A] Test von Znf_Zif268, Znf_Blues, Znf_PBSII, Znf_HivC und NicTAL (Tabelle 1). [B] Spezifische Erkennung der Zielnukleinsäuremotive durch die Nukleinsäure bindende Faktoren. Das spezifische Zielnukleinsäuremotiv wurde durch ein Zielnukleinsäuremotiv ersetzt, das spezifisch für einen unterschiedlichen Nukleinsäure bindenden Faktor ist. Das Reporterprotein β -Galactosidase wurde nicht unterdrückt, wenn ein nicht geeignetes Zielnukleinsäuremotiv verwendet wurde.

[0023] Figur 4. Bindung von chimären Proteinen an die Programm-Nukleinsäure in vitro. [A] Die Hybridisierung der Programm-Nukleinsäure erfolgte durch die Injektion von zwischen 0,5 und 2 μ M für bis zu 300 s, um die abschließende Reaktion von 300 RU zu erhalten. Die Überwachung der Bindung von chimären Proteinen erfolgte mittels der Injektion unterschiedlicher Konzentrationen von chimären Proteinen für 1 min im Anschluss an einen 5-minütigen Dissoziations-schritt. Die Regeneration erfolgte mittels zweier Injektionen von 50 mM NaOH über 30 s sowie einer Injektion von 0,5 % SDS über 24 s. [B] Die chimären Proteine, von oben nach unten: Znf_GliI, Znf_Zif268, Znf_Blues, Znf_HivC, Znf_PBSII, Znf_Jazz. [C] Zu Beginn des Zyklus wurde DNA eingefangen und die chimären Proteine wurden für 1 Minute nacheinander in 2 unterschiedlichen Sequenzen injiziert: durchgezogene Linie: Znf_Jazz, Znf_Blues, Znf_Zif268, Znf_PBSII, Znf_HivC und Znf_GliI; gestrichelte Linie: Znf_GliI, Znf_PBSII, Znf_HivC, Znf_Jazz, Znf_Blues und Znf_Zif268. Die Bindung aller chimären Proteine in den Sequenzen zeigte, dass die Bindung der chimären Proteinen schwächer war, wenn Znf_GliI als erstes injiziert wurde.

[0024] Figur 5. Bindung von Nukleinsäure bindenden Faktoren auf einer Programm-Nukleinsäuresequenz in vitro. Die Bindung von einzelnen Nukleinsäure bindenden Faktoren, Znf_Jazz, Znf_Blues, Znf_Zif268, Znf_PBSII, Znf_HivC und Znf_GliI

wurde mittels eines Mobilitäts-Split-Assays auf Agarosegel bestimmt. Die Position des Protein-DNA-Komplexes ist jeweils durch die Pfeile angezeigt.

[0025] Figur 6. In vivo Rekonstitution zweier nicht funktioneller gespaltenen (Split) GFP-Proteine, die an Nukleinsäure bindende Faktoren gebunden sind, in der Gegenwart einer Programm-Nukleinsäuresequenz. [A] Schematische Darstellung der Rekonstitution der gespaltenen GFP. [B] Es erfolgte eine Kotransfektion von jeweils 75 ng an Plasmiden, die His_tag-Znf_Gli1-Peptidlinker-nCFP und cCFP-Peptidlinker-Znf_HIVC-His_tag trugen, mit 350 ng der Programm-Nukleinsäure (Programm-SPR, Split, FRET; Tabelle 1). Die Detektion der Emission von CFP bei 475 nm erfolgte im Anschluss an eine Anregung bei 433 nm. [C] Es erfolgte eine Kotransfektion von jeweils 75 ng an Plasmiden, die chimäre gesplattene YFP (His_tag- Znf_PBSII_Peptidlinker-nYFP und cYFP-Peptidlinker-Znf_Zif268-Peptidlinker-His_tag) trugen, mit 350 ng an Plasmid, das für die Programm-Nukleinsäure kodierte (Programm-SPR, Split, FRET; Tabelle 1). Das Emissionssignal des YFP bei 529 nm wurde im Anschluss an eine Anregung bei 513 nm unter einem konfokalen Mikroskop gemessen.

[0026] Figur 7. In vitro Rekonstitution zweier nicht funktioneller gespaltenen (Split) GFP, die an Nukleinsäure bindende Faktoren gebunden sind, in der Gegenwart einer Programm-Nukleinsäuresequenz. Lysate von HEK293-Zellen, die mit Plasmiden transfiziert waren, die entweder für His_tag-Znf_Gli1-Peptidlinker-nCFP und cCFP-Peptidlinker-Znf_HIVC-His_tag oder für His_tag-Znf_PBSII-Peptidlinker-nYFP und cYFP-Peptidlinker-Znf_Zif268-Peptidlinker-His_tag kodierten, wurden mit 50 µg der Programm-Nukleinsäure (Programm-SPR, Split, FRET; Tabelle 1) vermischt. Im Anschluss an eine Inkubation für 18 h bei 4°C wurden die Fluoreszenzspektren für CFP und YFP erhalten. Die Spitzenwerte (Peaks) der Emission sind durch die Pfeile angezeigt.

[0027] Figur 8. In vivo Rekonstitution von zwei Paaren an gespaltenen fluoreszierenden Proteinen (Split-CFP und Split-YFP) wurden mittels einer FRET-Analyse bestimmt. Es erfolgte die Kotransfektion von HEK293-Zellen mit Plasmiden, die jeweils für His_tag-Znf_Gli1-Peptidlinker-nCFP, cCFP-Peptidlinker-Znf_HIVC-His_tag, His_tag-Znf_PBSII_Peptidlinker-nYFP und cYFP-Peptidlinker-Znf_Zif268-Peptidlinker-His_tag kodieren, sowie mit einem Plasmid, das für die Programm-Nukleinsäure kodiert (Programm-SPR, Split, FRET; Tabelle 1). [A. oben links] Darstellung eines aus gespaltenen CFP rekonstituierten CFP. Anregung bei 453 nm, Emission zwischen 470 und 510. [A. oben rechts] Darstellung eines aus gespaltenen YFP rekonstituierten YFP. Anregung bei 415 nm, Emission zwischen 525 und 560. [A. unten links] Überlagerung der obigen Darstellungen von CFP und YFP. [A. unten rechts] Die gleichen Zellen, gefärbt mit dem Farbstoff Hoechst 34580. Die Pfeile zeigen die FRET-positiven Zellen an, was die Kolo-kalisierung aller vier Fusionen von gespaltenen GFP innerhalb des Zellkerns belegt. [B. oben links] Darstellung des Donors, rekonstituiertes CFP vor dem Photobleichen. [B. oben rechts] Darstellung des Akzeptors, rekonstituiertes mCitrin nach dem Photobleichen. Die gebleichte Region ist durch einen Pfeil kenntlich gemacht. [B. unten links] Darstellung des Donors, rekonstituiertes CFP nach dem Photobleichen. Eine erhöhte Fluoreszenz des Donors nach dem Photobleichen des Akzeptors ist durch einen Pfeil kenntlich gemacht.

[0028] Figur 9. Biosynthese von Violacein in der Gegenwart der Programm-Nukleinsäure. Über Nacht angesetzte Kulturen von E. coli, enthaltend Plasmide, die für die chimären Proteine Znf_Blues-Peptidlinker-vioA, Znf_Zif268-Peptidlinker-vioB, Znf_PBSII-Peptidlinker-vioE, Znf_HivC-Peptidlinker-vioD, Znf_Gli1-Peptidlinker-vioC kodieren (siehe Tabelle 1), entweder mit dem oder ohne das

Plasmid, das für die Programm-Nukleinsäure kodiert (siehe Tabelle 1, Programm-Biosynthese (123456) oder Programm-Biosynthese (341256)). Die Quantifizierung der Menge an Violacein erfolgte 17,5 h nach der Extraktion und Analyse mittels TLC. Die in der Gegenwart der Programm-Nukleinsäure (Programm-Biosynthese 123456) produzierte Menge an Violacein war wesentlich höher als in den Fällen, in denen die Programm-Nukleinsäure (Programm-Biosynthese 2341256) oder keine Programm-Nukleinsäure in *E. coli* vorlag.

AUSFÜHRLICHE BESCHREIBUNG DER ERFINDUNG:

DEFINITIONEN:

[0029] Der Begriff "Programm/Nukleinsäure/Sequenz", so wie hierin verwendet, bezieht sich auf eine Nukleinsäure, die eine beliebige Anzahl an Zielnukleinsäuresequenzen einer beliebigen Länge enthält, die durch Nukleinsäure bindende Faktoren erkannt werden. Diese Zielnukleinsäuresequenzen werden durch Spacer-Nukleinsäuresequenzen einer beliebigen Länge voneinander getrennt. Die Programm-Nukleinsäuresequenz kann entweder als solche verwendet werden oder in einen geeigneten Vektor eingebracht werden, der dann in eine Wirtszelle eingebracht und durch diese amplifiziert werden soll.

[0030] Der Begriff "Ziel/Nukleinsäure/Sequenz" bzw. "Ziel/Nukleinsäure/Element/Motiv", so wie hierin verwendet, bezieht sich auf eine Nukleinsäuresequenz einer beliebigen Länge, die als Substrat für einen Bindungsfaktor dient.

[0031] Der Begriff "Nukleinsäure/bindender Faktor/Bindungsfaktor", so wie hierin verwendet, bezieht sich auf ein beliebiges Molekül, das zur Bindung eines Nukleinsäuremoleküls in der Lage ist. Der Bindungsfaktor kann natürlichen Ursprungs sein oder es kann sich dabei um ein künstlich konstruiertes vollständiges Protein oder lediglich um ein Segment eines solchen handeln, das zur sequenzspezifischen Bindung an eine Nukleinsäure in der Lage ist. Im Rahmen der vorliegenden Erfindung wird ein Nukleinsäure bindendes Element als Träger für ein Enzym, das an dem Biosyntheseweg beteiligt ist, oder für ein beliebiges weiteres funktionelles Polypeptid verwendet. Ein Nukleinsäure bindendes Molekül ist über eine chemische Bindung mit einem Enzymmolekül verbunden. Diese Verbindung wird mithilfe eines Peptidlinkers erreicht. Ein Nukleinsäure bindendes Element bindet an eine oder mehrere spezifische Stellen der Programm-Nukleinsäuresequenz und gewährleistet so eine hohe lokale Konzentration sowie eine geeignete räumliche Positionierung von Enzymen, die an dem Biosyntheseweg beteiligt sind.

[0032] Der Begriff "Spacer/Nukleinsäure/Sequenz", so wie hierin verwendet, bezieht sich auf eine Nukleinsäuresequenz einer beliebigen Länge, die Zielnukleinsäuresequenzen voneinander trennt.

[0033] Der Begriff "Nukleinsäuren", so wie hierin verwendet, bezieht sich auf eine polymere Form von Nukleotiden einer beliebigen Länge, und zwar entweder auf Ribonukleotide oder Desoxynukleotide, einschließlich einzel-, doppel-, oder mehrsträngige DNA oder RNA, genomische DNA, cDNA, DNA-RNA-Hybriden oder ein Polymer mit einem Phosphorothioat-Rückgrat, umfassend die Basen Purin und Pyrimidin oder weitere natürlich vorkommende, chemisch oder biochemisch modifizierte, nicht natürlich vorkommende oder derivatisierte Nukleotidbasen, jedoch nicht auf diese beschränkt.

[0034] Die Begriffe "Polypeptid", "Protein" und "Peptid", so wie hierin verwendet, beziehen sich auf eine polymere Form von Aminosäuren einer beliebigen Länge, die kodierte und nicht kodierte Aminosäuren, chemisch oder biochemisch modifizierte oder derivatisierte Aminosäuren und Polypeptide mit einem modifizierten Peptid- Rückgrat einschließen können. Der Begriff "funktionelles Polypeptid", so wie hierin verwendet, bezieht sich auf eine polymere Form von Aminosäuren einer beliebigen Länge, die eine beliebige Funktion exprimiert, so wie z. B. die Ausbildung der Struktur, das Dirigieren an einen spezifischen Ort, die Zielausrichtung zu Organellen, das Begünstigen oder Ausführen einer chemischen Reaktion oder die Bindung an weitere funktionelle Polypeptide. Der Begriff "Enzym des Biosynthesewegs", so wie hierin verwendet, be-

zieht sich auf eine polymere Form von Aminosäuren einer beliebigen Länge, die eine beliebige Funktion exprimiert, so wie z. B. die Ausbildung neuer chemischer Bindungen.

[0035] Der Begriff "chimäres Protein", so wie hierin verwendet, ist von allgemeiner Bedeutung und bezieht sich im Rahmen der vorliegenden Beschreibung auf eine polymere Form von Aminosäuren einer beliebigen Länge, die sich aus mehr als einem Protein/Segment bzw. aus mehr als einer Domäne zusammensetzen, die gegebenenfalls durch Linker einer beliebigen Länge miteinander verknüpft sind, die bevorzugt zwischen einer und 40 Aminosäuren enthalten, wobei es sich jeweils bei mindestens einem Protein/Segment bzw. einer Domäne um einen Nukleotid bindenden Faktor, eine vollständige oder eine Bindungsdomäne und bei dem bzw. der anderen um ein Enzym des Biosynthesewegs oder ein weiteres funktionelles Polypeptid, eine vollständige oder eine Aktivitätsdomäne handelt.

[0036] Der Begriff "heterolog", so wie hierin verwendet, bezieht sich im Zusammenhang mit einer genetisch modifizierten Wirtszelle auf ein Polypeptid, auf das mindestens eine der folgenden Aussagen zutrifft: (a) das Polypeptid ist ein Fremdpolypeptid ("exogen") für die Wirtszelle (d. h. es kommt in dieser nicht natürlich vor); (b) das Polypeptid kommt zwar natürlich in dem betreffenden Wirts-Mikroorganismus oder in der betreffenden Wirtszelle vor (ist z. B. "endogen"), wird jedoch in der Zelle entweder in einer nicht natürlichen Menge produziert (z. B. in einer Menge, die größer ist als erwartet oder größer als die natürlich produzierte Menge) oder unterscheidet sich hinsichtlich der Nukleotidsequenz derart von der endogenen Nukleotidsequenz, dass das gleiche kodierte Protein (mit der gleichen oder im Wesentlichen gleichen Aminosäuresequenz), das endogen vorliegt, in der Zelle in einer nicht natürlichen Menge produziert wird (z. B. in einer Menge, die größer ist als erwartet oder größer als die natürlich produzierte Menge).

[0037] Der Begriff "homolog", so wie hierin verwendet, bezieht sich auf Proteine oder Nukleinsäuren mit gut konservierten Aminosäure- oder Nukleotidsequenzen, die bevorzugt zu mindestens 50%, mindestens jedoch zu 20% konserviert sind, die mittels einer Technik des Protein- oder Nukleinsäurevergleichs bestimmt werden, so wie es im Stand der Technik bekannt ist. Homologe Proteine sind dadurch gekennzeichnet, dass sie innerhalb einer Zelle identische Funktionen erfüllen. Homologe Nukleinsäuren kodieren für homologe Proteine.

[0038] Der Begriff "rekombinant", so wie hierin verwendet, bedeutet, dass eine bestimmte Nukleinsäure (DNA oder RNA) das Produkt unterschiedlicher Kombinationen aus Klonierung, Restriktion und/oder Ligationsschritten ist, wodurch ein Konstrukt erhalten wird, das eine strukturell kodierende oder nicht kodierende Sequenz aufweist, die von den endogenen Nukleinsäuren unterschieden werden kann, die in natürlichen Systemen zu finden sind. Im Allgemeinen können DNA-Sequenzen, die für die strukturell kodierende Sequenz kodieren, aus cDNA-Fragmenten und kurzen Oligonukleotidlinkern oder aus einer Reihe von synthetischen Oligonukleotiden zusammengesetzt sein, um so eine synthetische Nukleinsäure zur Verfügung zu stellen, die dazu in der Lage ist, aus einer in einer Zelle enthaltenen rekombinanten Transkriptionseinheit oder in einem zellfreien Transkriptions- und Translationssystem exprimiert zu werden. Solche Sequenzen können in Form eines offenen Leserahmens bereitgestellt werden, der nicht durch interne, nicht translatierte Sequenzen oder Introns unterbrochen ist, die typischerweise in eukaryotischen Genen vorliegen. Zur Ausbildung eines rekombinanten Gens oder einer Transkriptionseinheit kann ebenfalls genomische DNA verwendet werden, die die relevanten Sequenzen umfasst. Es können ebenfalls Sequenzen nicht translaterter DNA in 5' - oder 3'-Richtung des offenen Leserahmens vorliegen, sofern diese Sequenzen die Manipulation oder Expression der kodierenden Regionen nicht beeinträchtigen, und die Sequenzen können dann tatsächlich dahingehend wirken, dass sie die Produktion eines gewünschten Produkts über verschiedene Mechanismen modulieren (siehe "regulatorische DNA-Sequenzen", infra).

[0039] Der Begriff "Wirtszelle", so wie hierin verwendet, bezieht sich auf eine in vivo oder in vitro vorliegende eukaryotische Zelle, eine prokaryotische Zelle oder eine Zelle aus einem mehrzelligen Organismus (z. B. eine Zelllinie), die als eine einzellige Einheit kultiviert wird, wobei die eukaryotischen oder prokaryotischen Zellen als Rezipienten für eine Nukleinsäure

verwendet werden können oder bereits verwendet wurden (z. B. ein Expressionsvektor, der eine Nukleotidsequenz umfasst, die für ein oder mehrere Genprodukte des Biosynthesewegs kodiert, wie z. B. für Genprodukte des Mevalonat-Synthesewegs), und schließt die Nachkommenschaft der ursprünglichen Zelle ein, die durch die Nukleinsäure genetisch modifiziert wurde. Es wird davon ausgegangen, dass die Nachkommenschaft einer einzelnen Zelle hinsichtlich ihrer Morphologie oder des Komplements von genomischer oder Gesamt-DNA nicht notwendigerweise zu 100% identisch mit der ursprünglichen Zelle sein muss, und zwar aufgrund von natürlicher, zufälliger oder gezielter Mutation. Bei einer "genetisch modifizierten Wirtszelle" (ebenfalls bezeichnet als eine "rekombinante Wirtszelle") handelt es sich um eine Wirtszelle, in die eine heterologe Nukleinsäure, z. B. ein Expressionsvektor, eingebracht wurde. So wird z. B. eine prokaryotische Wirtszelle (z. B. ein Bakterium) dadurch zu einer genetisch modifizierten prokaryotischen Wirtszelle, dass in eine geeignete prokaryotische Wirtszelle eine heterologe Nukleinsäure, z. B. eine exogene Nukleinsäure, die gegenüber der prokaryotischen Wirtszelle fremd ist (d. h. in dieser nicht natürlich vorkommt), oder eine rekombinante Nukleinsäure, die normalerweise nicht in der prokaryotischen Wirtszelle zu finden ist, eingebracht wird. Eine eukaryotische Wirtszelle wird dadurch zu einer genetisch modifizierten eukaryotischen Wirtszelle, dass in eine geeignete eukaryotische Wirtszelle eine heterologe Nukleinsäure, z. B. eine exogene Nukleinsäure, die gegenüber der eukaryotischen Wirtszelle fremd ist, oder eine rekombinante Nukleinsäure, die normalerweise nicht in der eukaryotischen Wirtszelle zu finden ist, eingebracht wird.

[0040] Der Begriff "Biosyntheseweg", so wie hierin verwendet, bezieht sich auf eine Abfolge von enzymatischen oder anderen Reaktionen, durch die eine Verbindung in eine andere Verbindung umgesetzt wird, und zwar durch die Ausbildung neuer kovalenter Bindungen in lebenden Organismen oder in vitro.

[0041] Der Begriff "in vitro", so wie hierin verwendet, bezieht sich auf einen Prozess, der nicht in einem lebenden Organismus oder einer lebenden Zelle, sondern in einer kontrollierten Umgebung stattfindet.

[0042] Der Begriff "Peptidlinker" bezieht sich auf kürzere Aminosäuresequenzen, deren Rolle unter Umständen lediglich darin besteht, die einzelnen Domänen des Fusionsproteins voneinander zu trennen. Die Rolle des Peptidlinkers in dem Fusionsprotein, in das er gegebenenfalls eingeschlossen sein kann, kann ebenfalls darin bestehen, die Stelle für eine Spaltung oder für posttranslationale Modifikationen einzubringen, einschließend die Einbringung von Stellen zur verbesserten Verarbeitung von Antigenen. Die Länge des Peptidlinkers ist nicht beschränkt, beträgt jedoch üblicherweise bis zu 30 Aminosäuren.

[0043] Im Allgemeinen wird die heterologe Nukleinsäure in einen Expressionsvektor eingebracht. Geeignete Vektoren schließen Plasmide, virale Vektoren und dergleichen ein, sind jedoch nicht auf diese beschränkt. Expressionsvektoren, die kompatibel mit den Zellen eines Wirtsorganismus sind, sind im Stand der Technik hinreichend bekannt und enthalten geeignete Kontrollelemente für die Transkription und Translation von Nukleinsäuren. Typischerweise schließt der Expressionsvektor eine Antibiotika-Resistenz-Kassette, eine Sequenz für ein chimares Protein unter der Kontrolle eines geeigneten Promotors zur Regulierung der Expression in Wirtszellen, ein Polyadenylierungssignal sowie eine Transkriptionsterminationssequenz ein.

VERFAHREN ZUR HERSTELLUNG EINES BIOSYNTHESEPRODUKTS BZW. EINES PRODUKTVORLÄUFERS.

[0044] Die vorliegende Erfindung stellt ein Verfahren zur Herstellung eines Produkts oder eines Produktvorläufers eines Biosynthesewegs in einer genetisch modifizierten Wirtszelle oder in vitro zur Verfügung. Das Verfahren umfasst allgemein das Kultivieren der genetisch modifizierten Wirtszelle unter geeigneten Bedingungen, unter denen der genetisch modifizierte Wirt umfasst: a) Nukleinsäuren, die Nukleotidsequenzen umfassen, die für mindestens drei oder mehr Bindungsfaktoren kodieren, die an Enzyme des Biosynthesewegs oder weitere funktionelle Polypeptide gebunden sind, und b) eine Nukleinsäure, die eine oder mehrere Programm-

Nukleinsäuresequenzen umfasst.

[0045] In einer weiteren Ausführungsform der vorliegenden Erfindung umfasst das Verfahren allgemein die Bildung eines Produkts *in vitro*. Das *in vitro* Verfahren umfasst: a) mindestens drei oder mehr Bindungsfaktoren, die an chimäre Enzyme des Biosynthesewegs oder weitere funktionelle Polypeptide gebunden sind, b) eine Nukleinsäure, die eine oder mehrere Programm-Nukleinsäuresequenzen umfasst, und c) das Bereitstellen eines Substrats für das erste Enzym sowie von Kofaktoren für die Enzyme in dem Gemisch.

[0046] Ein Bindungsfaktor bindet mit einer ausreichenden Assoziation an das Zielnukleinsäureelement innerhalb der Programm-Nukleinsäuresequenz und stellt so eine Assoziation der chimären funktionellen Polypeptide des Biosynthesewegs mit der Programm-Nukleinsäuresequenz zur Verfügung. Die Assoziation der chimären funktionellen Polypeptide des Biosynthesewegs mit der Programm-Nukleinsäuresequenz ist von einer ausreichenden Affinität, um die chimären Enzyme des Biosynthesewegs oder weitere funktionelle Polypeptide auf einer Programm-Nukleinsäuresequenz zu immobilisieren. Die Programm-Nukleinsäuresequenz legt die Reihenfolge der einander benachbarten Bindungsfaktoren fest und damit auch die Abfolge von einander benachbarten chimären Enzymen oder weiteren funktionellen Polypeptiden. Die Reihenfolge der Zielnukleinsäureelemente mit daran gebundenen funktionellen Polypeptiden kann entsprechend der Reihenfolge festgelegt werden, in der diese innerhalb einer bestimmten Biosynthesereaktion zum Einsatz kommen. Alternativ kann die Reihenfolge von einander benachbarten Zielnukleinsäureelementen mit entsprechenden funktionellen Polypeptiden dadurch verändert werden, dass die Abfolge von einander benachbarten Zielsequenzen innerhalb der Programm-Nukleinsäuresequenz verändert wird, wodurch eine alternative Reihenfolge von einander benachbarten Biosyntheseezymen oder weiteren funktionellen Polypeptiden zur Verfügung gestellt wird. Die Kultivierung der Wirtszelle erfolgt so, dass ein Substrat für das erste Enzym: a) in der Zelle vorliegt, oder b) der Zelle aus der extrazellulären Umgebung bereitgestellt wird. Die Enzyme des Biosynthesewegs werden in der Zelle synthetisiert und setzen das Substrat in ein Produkt um, oder b) die Biosyntheseezyme werden in der Zelle synthetisiert, aus der Zelle freigesetzt und die Umwandlung des Substrats in ein Produkt erfolgt in der extrazellulären Umgebung.

[0047] Durch die Verwendung der Programm-Nukleinsäuresequenz zur Immobilisierung der funktionellen Polypeptide des Biosynthesewegs werden eines oder mehrere der folgenden Kriterien erfüllt: eine Steigerung der Effizienz des Synthesewegs, eine Optimierung des metabolischen Flusses, eine Reduktion der metabolischen Belastung und der Konzentration von potentiell toxischen freien Zwischenprodukten im Zytosol sowie die mögliche Ausbildung von alternativen Verzweigungen der metabolischen Synthesewege, wodurch in einem Biosyntheseprozess neue Produkte hergestellt werden können. Das Gleichgewicht zwischen den Aktivitätsniveaus der Enzyme wird durch die Verwendung der Programm-Nukleinsäuresequenz erreicht. So kann z. B. die Sequenz des Nukleinsäureprogramms auf eine Weise verändert werden, dass mehrere Kopien von Zielnukleinsäureelementen für chimäre Polypeptide mit einer geringeren Aktivität in Richtung der Umsetzung des Substrats eingeschlossen sind, was zu einer Erhöhung der Anzahl an Kopien der funktionellen Polypeptide mit einer geringeren Aktivität gegenüber der Anzahl der Kopien von funktionellen Polypeptiden mit einer höheren Aktivität führt. Dies kann von Vorteil sein, wenn das funktionelle Polypeptid mit der geringeren Aktivität einen den Durchsatz limitierenden Schritt in dem betreffenden Biosyntheseweg katalysiert. Mit der Verwendung einer Programm-Nukleinsäuresequenz wird durch die größere räumliche Nähe der an dem Biosyntheseweg beteiligten Enzyme zum einen die Effizienz gesteigert und zum anderen der Fluss des Synthesewegs optimiert. Aufgrund der gesteigerten Effizienz und des optimierten Flusses können mit geringeren Mengen an Enzymen gleiche oder höhere Ausbeuten des Produkts erzielt werden. Ein niedriges Niveau der Enzymproduktion in einer Wirtszelle ist vorteilhaft, da hierdurch die metabolische Belastung für die Wirtszelle reduziert wird. Da die Umsetzung von Zwischenprodukten des Biosynthesewegs mit der Verwendung einer Programm-Nukleinsäuresequenz auf eine effizientere Weise abläuft, kommt es zu einer Reduktion der Menge bzw. Konzentration an Zwischenprodukten des Synthesewegs, die in freier Form im

Zytosol oder Zytoplasma vorliegen. Geringere Mengen an freien Zwischenprodukten des Synthesewegs sind dann vorteilhaft, wenn diese Zwischenprodukte toxisch auf die Wirtszelle wirken.

[0048] In einigen Ausführungsformen werden mindestens drei chimäre Enzyme des Biosynthesewegs oder weitere funktionelle Polypeptide auf dem Nukleinsäureprogramm immobilisiert. Das erste chimäre funktionelle Enzym oder Polypeptid produziert ein erstes Produkt, das ein Substrat für das zweite chimäre Enzym oder das weitere funktionelle Polypeptid darstellt. Die chimären Proteine, Enzyme des Biosynthesewegs oder weiteren funktionellen Polypeptide sind in großer räumlicher Nähe zueinander angeordnet. Auf diese Weise ist die effektive Konzentration des ersten Produkts hoch und das zweite Enzym des Biosynthesewegs kann in effizienter Weise auf das erste Produkt einwirken.

[0049] Es können drei oder mehr (z. B. drei, vier, fünf, sechs, sieben oder mehr) funktionelle Polypeptide auf der Programm-Nukleinsäuresequenz immobilisiert werden. So schließt z.B. in einigen Ausführungsformen die Programm-Nukleinsäuresequenz (vom 5'- zum 3'-Ende) ein: a) eine Kopie eines Zielnukleinsäureelements für das erste chimäre Polypeptid des Enzyms des Biosynthesewegs oder das weitere funktionelle Polypeptid, b) eine Kopie eines Zielnukleinsäureelements für das zweite chimäre funktionelle Polypeptid des Biosynthesewegs und c) eine Kopie eines Zielnukleinsäureelements für das dritte chimäre funktionelle Polypeptid des Biosynthesewegs. In weiteren Ausführungsformen schließt eine Programm-Nukleinsäuresequenz (vom 5'- zum 3'-Ende) ein: a) eine Kopie einer Zielsequenz für das erste chimäre funktionelle Polypeptid des Biosynthesewegs und b) zwei oder mehr (z. B. zwei, drei, vier, oder mehr) Kopien eines Zielnukleinsäureelements für das zweite chimäre Enzym des Biosynthesewegs oder das weitere funktionelle Polypeptid. Auf diese Weise kann das Verhältnis eines beliebigen chimären funktionellen Polypeptids in einem Biosyntheseweg variiert werden. So kann z. B. das Verhältnis eines ersten chimären funktionellen Polypeptids des Biosynthesewegs zu einem zweiten chimären funktionellen Polypeptid des Biosynthesewegs zwischen etwa 0,1:10 und etwa 10:0,1 und das Verhältnis von dem ersten zu dem dritten chimären Protein zwischen etwa 0,1:10 und etwa 10:0,1 etc. variieren.

[0050] In einigen Ausführungsformen sind mindestens drei oder mehr (z. B. vier, fünf, sechs, sieben oder mehr) chimäre Enzyme des Biosynthesewegs auf einem Nukleinsäureprogramm immobilisiert. Das erste chimäre Enzym des Biosynthesewegs produziert ein erstes Produkt, das als Substrat für das zweite Produkt dient, das wiederum als Substrat für das dritte Produkt des Enzyms des Biosynthesewegs dient: In diesen Ausführungsformen werden die Reihenfolge und die Anzahl an Kopien der Enzyme des Biosynthesewegs durch die Sequenz der Zielnukleinsäuresequenz für ein bestimmtes chimäres Enzym des Biosynthesewegs in einem Nukleinsäureprogramm vorgegeben.

KONSTRUKTION EINER PROGRAMM-NUKLEINSÄURESEQUENZ.

[0051] Die Konstruktion einer Programm-Nukleinsäuresequenz erfolgt, um Enzyme eines Biosynthesewegs in einem funktionellen Komplex zu organisieren. Eine Programm-Nukleinsäuresequenz umfasst drei oder mehr Zielnukleinsäureelemente für die Bindung an Nukleinsäure bindende Faktoren. Durch die Bindung eines Nukleinsäure bindenden Faktors, der an ein chimäres Enzym des Biosynthesewegs gebunden ist, an das Zielnukleinsäureelement wird für die Immobilisierung eines Enzyms auf der Programm-Nukleinsäuresequenz gesorgt. Jedes Zielnukleinsäureelement verfügt über einen entsprechenden Nukleinsäure-Bindungspartner in einem chimären Enzym des Biosynthesewegs oder einem funktionellen Polypeptid. Ein Zielnukleinsäureelement kann entweder unmittelbar zu einem weiteren Zielnukleinsäureelement benachbart sein oder es kann durch eine Spacer-Nukleinsäuresequenz von einem benachbarten Zielnukleinsäureelement getrennt sein. In den in vivo Ausführungsformen kann die Programm-Nukleinsäuresequenz in jeweils unterschiedlichen Formen in eine Vielzahl von unterschiedlichen Arten von Wirtszellen eingebracht werden. Beispiele für unterschiedliche Formen der Programm-Nukleinsäuresequenz schließen Baculovirus-Vektoren, Bakteriophagen-Vektoren, Plasmide, Phagemide, Cosmide, Fosmide, künstliche bakterielle Chromosomen, virale Vektoren

(z. B. virale Vektoren, die auf dem Vacciniavirus, Poliovirus, Adenovirus, Adeno-assoziierten Virus, SV40, Herpes-Simplex-Virus und dergleichen basieren, jedoch nicht auf diese beschränkt), P1-basierte künstliche Chromosomen, Hefepasmide, künstliche Hefechromosomen und weitere Vektoren ein, sind jedoch nicht auf diese beschränkt. Die Programm-Nukleinsäuresequenz kann ebenfalls in Form eines linearen Nukleinsäuremoleküls in eine Vielzahl an unterschiedlichen Wirtszellen eingebracht werden. Alle diese Formen der Programm-Nukleinsäuresequenz können ebenfalls in den in vitro Ausführungsformen der vorliegenden Erfindung verwendet werden. Setzt sich die Programm-Nukleinsäuresequenz zumindest zum Teil aus RNA zusammen, so kann sie als solche verwendet werden; die Verwendung von DNA, die für die Programm-Nukleinsäuresequenz kodiert, ist allerdings ebenfalls möglich. DNA, die für die Programm-Nukleinsäuresequenz kodiert, die in den in vitro oder in vivo Ausführungsformen der vorliegenden Erfindung verwendet wird, kommt in unterschiedlichen Formen vor, wie z. B. in einer beliebigen der vorstehend genannten Formen, jedoch nicht auf diese beschränkt.

[0052] In einem Produktionssystem kann die Programm-Nukleinsäuresequenz in Form von einer oder mehrerer Kopien vorliegen. Durch das Vorliegen von mehreren Kopien einer einzelnen Programm-Nukleinsäuresequenz kann eine größere Anzahl an funktionellen Biosynthesekomplexen zur Verfügung gestellt werden, die aus immobilisierten chimären Enzymen des Biosynthesewegs konstruiert sind. Einzelne Kopien der Programm-Nukleinsäuresequenzen können in einem oder mehreren Nukleinsäuremolekülen vorliegen. Folglich wird durch die Amplifikation der Anzahl an einzelnen Nukleinsäuremolekülen in dem Produktionssystem mit einer oder mehreren Kopien der Programm-Nukleinsäuresequenz ebenfalls eine größere Gesamtanzahl an Kopien der Programm-Nukleinsäuresequenzen in dem produzierenden System zur Verfügung gestellt.

[0053] Zum Beispiel, jedoch nicht hierauf beschränkt, kann eine bakterielle Wirtszelle mit einem selbstreplizierenden Plasmid oder einer beliebigen weiteren Art von Nukleinsäuremolekül, das eine Programm-Nukleinsäuresequenz in einer oder mehreren Kopien trägt, transformiert sein, wodurch eine höhere Gesamtanzahl an Programm-Nukleinsäuresequenzen in dem Produktionssystem zur Verfügung gestellt wird. Ein Plasmid, das eine oder mehrere Kopien der Programm-Nukleinsäuresequenzen trägt, kann in dem System ebenfalls in einer oder mehreren Kopien vorliegen (z. B., jedoch nicht hierauf beschränkt, kann ein Plasmid mit einer hohen Kopienzahl zur Transformation einer bakteriellen Wirtszelle verwendet werden). Mit einer höheren Anzahl an Plasmiden mit jeweils einer oder mehreren Kopien der Programm-Nukleinsäuresequenzen wird in dem Produktionssystem ebenfalls eine höhere Gesamtanzahl an Programm-Nukleinsäuresequenzen zur Verfügung gestellt.

[0054] Zum Beispiel, jedoch nicht hierauf beschränkt, kann eine bakterielle Wirtszelle mit einer Nukleinsäuresequenz transformiert sein, die für ein einzelsträngiges, doppelsträngiges, teilweise doppelsträngiges, mehrsträngiges oder teilweise mehrsträngiges RNA-Molekül kodiert. Die Programm-Nukleinsäuresequenz kann in einer oder mehreren Kopien auf einem einzelnen RNA-Molekül vorliegen. Durch das Vorliegen von mehreren Kopien von Programm-Nukleinsäuresequenzen auf einem einzelnen RNA-Molekül wird eine höhere Gesamtanzahl an Programm-Nukleinsäuresequenzen in dem Produktionssystem zur Verfügung gestellt. Eine für ein RNA-Molekül kodierende Nukleinsäuresequenz kann ebenfalls in einer oder mehreren Kopien in dem Produktionssystem vorliegen. Mit einer höheren Anzahl an Kopien von für ein RNA-Molekül kodierenden Nukleinsäuresequenzen wird in dem Produktionssystem eine höhere Anzahl von RNA-Molekülen und damit ebenfalls eine höhere Gesamtanzahl an Programm-Nukleinsäuresequenzen zur Verfügung gestellt. Die Gesamtanzahl an einzelnen RNA-Molekülen in dem Produktionssystem kann auch mithilfe von Promotoren reguliert werden, die zur Anregung der Transkription von kodierenden Nukleinsäuresequenzen in RNA-Moleküle verwendet werden, die aus einer oder mehreren Kopien der Programm-Nukleinsäuresequenzen bestehen. Durch stärkere Promotoren wird in dem Produktionssystem eine höhere Anzahl an RNA-Molekülen und damit auch eine höhere Gesamtanzahl an einzelnen Programm-Nukleinsäuresequenzen zur Verfügung gestellt. Mit der höheren Gesamtanzahl an Programm-Nuklein-

säuremolekülen wird eine höhere Anzahl an funktionellen Biosynthesekomplexen zur Verfügung gestellt, die aus immobilisierten chimären Enzymen des Biosynthesewegs konstruiert sind.

[0055] Eine Programm-Nukleinsäuresequenz umfasst mindestens drei Zielnukleinsäuresequenzen und folglich binden mindestens drei entsprechende Nukleinsäure bindende Elemente in drei chimären Enzymen des Biosynthesewegs an die Programm-Nukleinsäuresequenz. Eine Programm-Nukleinsäuresequenz weist eine, zwei oder mehr Kopien eines jeden Zielnukleinsäureelements auf. Dem entsprechend kann jedes Zielnukleinsäureelement in einer oder mehreren Kopien vorliegen. Die Kopien können aneinander angrenzen oder durch eine Spacer-Nukleinsäuresequenz voneinander getrennt sein. Die Kopien des Zielnukleinsäureelements können ebenfalls durch ein oder mehrere Zielnukleinsäureelemente voneinander getrennt sein, die wiederum ebenfalls in einer oder mehreren Kopien vorliegen können. Liegen auf einer Programm-Nukleinsäuresequenz nur drei Zielnukleinsäureelemente vor, so können diese verschieden voneinander sein, wodurch eine geeignete räumliche Anordnung oder eine hohe lokale Konzentration für die zwei unterschiedlichen chimären Enzyme des Biosynthesewegs oder die weiteren funktionellen Polypeptide zur Verfügung gestellt werden kann. Liegen auf einer Programm-Nukleinsäuresequenz nur drei Zielnukleinsäureelemente vor, so können diese auch von der gleichen Art sein, wodurch eine geeignete räumliche Anordnung oder eine hohe lokale Konzentration für die drei Kopien der gleichen chimären Enzyme des Biosynthesewegs zur Verfügung gestellt werden kann.

[0056] Ein spezifisches Zielnukleinsäureelement ist durch die Nukleotidsequenz vom 5'- zum 3'-Ende eines Nukleinsäuremoleküls definiert. Auf einer doppelsträngigen Programm-Nukleinsäuresequenz oder einem doppelsträngigen Segment einer Programm-Nukleinsäuresequenz kann es sich bei der Zielnukleinsäuresequenz auch um die revers komplementäre Sequenz der ursprünglichen Zielnukleinsäuresequenz handeln. An das Zielnukleinsäureelement sowie an die revers komplementäre Sequenz der Zielnukleinsäuresequenz bindet der gleiche Nukleinsäure bindende Faktor, jedoch in einer unterschiedlichen Orientierung und an einer unterschiedlichen Stelle der Programm-Nukleinsäuresequenz. Dem entsprechend kann die Verwendung einer revers komplementären Sequenz des Zielnukleinsäureelements anstelle des Zielnukleinsäureelements zweckmäßig sein, wenn zwei chimäre Enzyme des Biosynthesewegs oder weitere funktionelle Polypeptide zu groß sind, um direkt aneinander anzugrenzen, oder zum Erreichen einer unterschiedlichen räumlichen Anordnung von zwei Kopien des gleichen chimären Enzyms des Biosynthesewegs oder weiterer funktioneller Polypeptide auf einer Programm-Nukleinsäuresequenz.

[0057] Ein Spacer-Nukleinsäureelement kann auf einer Programm-Nukleinsäuresequenz zwischen zwei Zielnukleinsäureelementen angeordnet werden, um ausreichend Platz für große chimäre Enzyme des Biosynthesewegs oder weitere funktionelle Polypeptide zu schaffen, die ihre Wirkung in dem Biosyntheseweg der Reihe nach ausüben. In einer doppelsträngigen Programm-Nukleinsäuresequenz oder einem doppelsträngigen Segment einer Programm-Nukleinsäuresequenz kann durch die Spacer-Nukleinsäuresequenz auch eine geeignete räumliche Anordnung von zwei aneinander angrenzenden oder nahe beieinander liegenden chimären Enzymen des Biosynthesewegs oder weiteren funktionellen Polypeptiden zur Verfügung gestellt werden. So kann es sich z. B. in einigen Ausführungsformen bei der Programm-Nukleinsäuresequenz um eine doppelsträngige DNA-Sequenz handeln, die eine gedrehte Helixstruktur ausbildet. Dem entsprechend führt eine Variation der Länge einer Spacer-Nukleinsäuresequenz zu einer unterschiedlichen räumlichen Position der immobilisierten chimären Enzyme des Biosynthesewegs sowie zu einem größeren Abstand zwischen zwei einander benachbarten chimären Enzymen des Biosynthesewegs (Figur 2).

[0058] Eine einzelne Programm-Nukleinsäuresequenz weist die allgemeine Formel $[(X \text{ oder } X')^n S^m]$ auf, wobei X^n für eine Zielnukleinsäuresequenz steht, X'^n für eine zu der Zielnukleinsäuresequenz revers komplementäre Sequenz steht, S für eine optionale Spacer-Nukleotidsequenz einer beliebigen Länge steht, m einer ganzen Zahl von 3 aufwärts entspricht und für eine Reihe an Zielnukleinsäuresequenzen mit einer optionalen Spacer-Nukleinsäuresequenz in der Programm-Nukleinsäure steht, n einem Indikator für die Art des Zielnukleinsäureelements

(z. B. X1, X2 und so weiter) entspricht und für unterschiedliche einzelne Zielnukleinsäureelemente steht, p einem Indikator für die Art des Spacer-Elements (z. B. S1, S2 und so weiter) entspricht und für unterschiedliche einzelne Spacer-Elemente steht. Eine einzelne Zielnukleinsäuresequenz in der Programm-Nukleinsäuresequenz kann eine beliebige Länge aufweisen und ist abhängig von dem entsprechenden Nukleinsäure bindenden Element. So können z. B. in einigen Ausführungsformen der vorliegenden Erfindung dreigliedrige DNA bindende Zinkfingerproteine verwendet werden. In diesen Fällen liegt die Länge der Zielnukleinsäuresequenzen üblicherweise, jedoch nicht immer, bei 9 oder 10 Nukleotiden.

[0059] So weist z. B. in einigen Ausführungsformen die Programm-Nukleinsäuresequenz die Formel (X1)(S1)(X2)(S2)(X3)(S3)(X4) auf, wobei X1 für die erste Zielnukleinsäuresequenz zur Bindung des ersten chimären Enzyms des Biosynthesewegs steht; wobei X2 für die zweite Zielnukleinsäuresequenz zur Bindung des zweiten chimären Enzyms des Biosynthesewegs steht; wobei X3 für die dritte Zielnukleinsäuresequenz zur Bindung des dritten chimären Enzyms des Biosynthesewegs steht; wobei X4 für die vierte Zielnukleinsäuresequenz zur Bindung des vierten chimären Enzyms des Biosynthesewegs steht; wobei S1, sofern vorhanden, für die erste Spacer-Nukleotidsequenz steht, die eine geeignete räumliche Anordnung von einander benachbarten chimären Enzymen des Biosynthesewegs gewährleistet; wobei S2, sofern vorhanden, für die zweite Spacer-Nukleotidsequenz steht, die eine geeignete räumliche Anordnung von einander benachbarten chimären Enzymen des Biosynthesewegs gewährleistet; wobei S3, sofern vorhanden, für die dritte Spacer-Nukleotidsequenz steht, die eine geeignete räumliche Anordnung von einander benachbarten chimären Enzymen des Biosynthesewegs gewährleistet.

[0060] Als weiteres Beispiel weist in einigen Ausführungsformen die Programm-Nukleinsäuresequenz die Formel (X1)(S1)(X2)(S1)(X2)(S1) (X3) auf, wobei X1 für die erste Zielnukleinsäuresequenz zur Bindung des ersten chimären Enzyms des Biosynthesewegs steht; wobei X2 für die zweite Zielnukleinsäuresequenz zur Bindung des zweiten chimären Enzyms des Biosynthesewegs steht und zwei Mal direkt aufeinander folgend wiederholt wird; wobei X3 für die dritte Zielnukleinsäuresequenz zur Bindung des dritten chimären Enzyms des Biosynthesewegs steht; und wobei S1, sofern vorhanden für die Spacer-Nukleotidsequenz der gleichen Art steht, die zwischen jedem Paar von Zielnukleinsäuresequenzen angeordnet ist. Durch in Folge wiederholte Zielnukleotidsequenzen (im vorliegenden Fall X2) kann eine höhere lokale Konzentration eines den Durchsatz limitierenden chimären Enzyms des Biosynthesewegs zur Verfügung gestellt werden, was zu einer erhöhten Effizienz des Flusses des Synthesewegs führt. Durch in Folge wiederholte Zielnukleotidsequenzen (im vorliegenden Fall X2) kann ebenfalls eine erhöhte Dimerisierung (oder Multimerisierung, wenn eine spezifische Zielnukleinsäuresequenz mehr als zwei Mal in Folge wiederholt wird) eines dimeren (oder multimeren) chimären Enzyms des Biosynthesewegs zur Verfügung gestellt werden. Durch in Folge wiederholte Zielnukleotidsequenzen (im vorliegenden Fall X2) kann ebenfalls eine erhöhte Effizienz des Flusses des Synthesewegs in Biosynthesewegen mit chimären Enzymen zur Verfügung gestellt werden, in denen ein Enzym mehr als ein Mal in Folge auf ein Substrat einwirkt.

[0061] Als weiteres Beispiel weist in einigen Ausführungsformen die Programm-Nukleinsäuresequenz die Formel (X1)(S1)(X2)(S2)(X1)(S3) (X3) auf, wobei X1 für die erste Zielnukleinsäuresequenz zur Bindung des ersten chimären Enzyms des Biosynthesewegs steht und in der Programm-Nukleinsäuresequenz erneut wiederholt wird, jedoch nicht in direkter Aufeinanderfolge; wobei X2 für die zweite Zielnukleinsäuresequenz zur Bindung des zweiten chimären Enzyms des Biosynthesewegs steht; wobei X3 für die dritte Zielnukleinsäuresequenz zur Bindung des dritten chimären Enzyms des Biosynthesewegs steht; wobei S1, sofern vorhanden, für die erste Spacer-Nukleotidsequenz steht, die eine geeignete räumliche Anordnung von einander benachbarten chimären Enzymen des Biosynthesewegs gewährleistet; wobei S2, sofern vorhanden, für die zweite Spacer-Nukleotidsequenz steht, die eine geeignete räumliche Anordnung von einander benachbarten chimären Enzymen des Biosynthesewegs gewährleistet; wobei S3, sofern vorhanden, für die dritte Spacer-Nukleotidsequenz steht, die eine geeignete räumliche Anordnung von einander benachbarten chimären Enzymen des Biosynthesewegs gewährleistet.

tet. Diese Programm-Nukleinsäuresequenz kann zur Verbesserung der Effizienz von Biosynthesewegen mit chimären Enzymen verwendet werden, in denen ein chimäres Enzym des Biosynthesewegs (im vorliegenden Fall das chimäre Enzym des Biosynthesewegs, das an die Zielnukleotidsequenz XI bindet) in einem Biosyntheseweg mit chimären Enzymen mehr als ein Mal wirkt, jedoch nicht in unmittelbarer Aufeinanderfolge.

KONSTRUKTION VON (NUKLEINSÄURE-) BINDUNGSFAKTOREN.

[0062] Die vorliegende Erfindung betrifft chimäre Polypeptide, die sich aus einem Nukleinsäure bindenden Faktor (NSBF) und einem Enzym des Biosynthesewegs oder einem weiteren funktionellen Polypeptid zusammensetzen, die über einen Aminosäurelinker miteinander verbunden sind. Ein NSBF ist entweder natürlichen Ursprungs oder künstlich hergestellt und kann im Prinzip aus einem beliebigen Organismus isoliert werden. Bei einem NSBF handelt es sich um ein beliebiges Polypeptid, Protein bzw. Proteinsegment oder eine beliebige Domäne, das bzw. die an eine Nukleinsäure bindet. Ein NSBF ist eine unabhängig gefaltete Proteindomäne, die mindestens ein Motiv enthält, das eine Nukleinsäuresequenz erkennt. Die Interaktion zwischen einem NSBF und Nukleotiden erfolgt auf eine sequenzspezifische Weise. Beispiele für Polypeptide mit einem NSBF schließen Helix-Turn-Helix, Zinkfinger, Leucin-Zipper, Winged-Helix, Winged-Helix-Turn-Helix, Helix-Loop-Helix und die HMG-Box ein, sind jedoch nicht auf diese beschränkt.

[0063] Polypeptide mit einem basischen Helix-Loop-Helix-Motiv werden in Leucin-Zipper-Faktoren, Helix-Loop-Helix-Faktoren, Helix-Loop-Helix-/Leucin-Zipper-Faktoren, NF-1, RP-X und bHSH klassifiziert und sind gekennzeichnet durch zwei α -Helices, die durch eine Schleife (Loop) miteinander verbunden sind. Transkriptionsfaktoren, die diese Domäne einschließen, binden typischerweise als Dimere an eine als E-Box (palindromische Sequenz (CACGTG)) bezeichnete Konsensussequenz. Die bHLH-Transkriptionsfaktoren binden an nicht palindromische Sequenzen, die der E-Box in vielen Fällen ähnlich sind.

[0064] Als weiteres Beispiel sind in einigen Ausführungsformen die NSBF aus der Überklasse von Zink koordinierenden DNA bindenden Domänen (Zinkfingerdomänen) ausgewählt, die wiederum in die folgenden Unterklassen unterteilt ist: Cys4-Zinkfinger des Typs nukleärer Rezeptor, diverse Cys4-Zinkfinger, Cys2His2-Zinkfingerdomänen, Cys6-Cystein-Zink-Cluster und Zinkfinger mit abwechselnder Zusammensetzung. Einzelne Zinkfingerdomänen kommen typischerweise in Form von aufeinander folgenden Wiederholungen (tandem repeats/arrays) vor, wobei zwei, drei oder mehr Finger die DNA bindende Domäne des Proteins umfassen. Diese aufeinander folgenden Wiederholungen können in der großen Furche der DNA binden und sind typischerweise um jeweils 3 bp voneinander beabstandet. Die α -Helix einer jeden Domäne (oftmals auch als die "Erkennungshelix" bezeichnet) kann die Nukleotide der Nukleinsäuren auf eine sequenzspezifische Weise kontaktieren; die Reste einer einzelnen Erkennungshelix können 4 oder mehr Nukleotide kontaktieren und so ein überlappendes Muster an Kontakten mit benachbarten Zinkfingern ausbilden.

[0065] Als weiteres Beispiel sind in einigen Ausführungsformen die NSBF ausgewählt aus der Helix-Turn-Helix-Überklasse an Nukleotid bindenden Polypeptiden, die wiederum in sechs Klassen unterteilt ist: die Homeodomäne, Forkhead/Winged-Helix, Hitzeschockfaktoren, Tryptophan-Cluster und die Transkriptionenhancer-Faktor (TEA)-Domäne. Bei Helix-Turn-Helix handelt es sich um ein wichtiges Strukturmotiv, das zur Bindung von DNA in der Lage ist, wobei die Erkennung der und Bindung an die DNA über die beiden α -Helices erfolgt, von denen sich die eine am N-terminalen Ende des Motivs und die andere am C-Terminus befindet.

[0066] Bei den Beta-Gerüst-Faktoren mit Kontakten zur kleinen Furche der DNA handelt es sich um die vierte Überklasse, die wiederum in mehrere Klassen unterteilt ist: RHR (Rel-Homologie-Domäne), STAT, p53, die MADS-Box, beta-Barrel-alpha-Helix-Transkriptionsfaktoren, TATA bindende Proteine, heteromere CCAAT-Faktoren, Grainyhead, Faktoren der Kälteschock-Domäne und Runt.

[0067] Ebenfalls von Bedeutung sind weitere Transkriptionsfaktoren, die an eine spezifische

Nukleotidsequenz binden, wie z. B. mutierte Restriktionsenzyme, die zwar über keine Restriktionsaktivität, jedoch über die Fähigkeit zur Sequenzerkennung verfügen.

[0068] Die vorliegende Erfindung betrifft, ist jedoch nicht beschränkt auf, Polypeptide mit einer Zinkfingerdomäne, die eine kleine, unabhängig gefaltete, zinkhaltige Minidomäne bilden, die eine spezifische Nukleinsäuresequenz erkennt. Die Zinkfingerdomänen treten typischerweise als aufeinander folgende Wiederholungen mit mindestens zwei, drei, oder mehr Fingern auf, welche die DNA bindende Domäne des Proteins umfassen. Jeder Finger erkennt aus 3 Basenpaaren bestehende Teilbindungsorte und bindet an diese. Die spezifische Bindung wird durch Aminosäuren der Minidomäne der Zinkfingerdomäne an den Positionen 1, 2, 3 und 6 vom Beginn der α -Helix vermittelt. In einem modularen oder kombinatorischen Ansatz ist es nun möglich, die sich wiederholenden Minidomänen zu multiplizieren und so durch Variation in bestimmten Schlüsselaminosäureresten eine chemische Unverwechselbarkeit zu erreichen. Dem entsprechend werden Finger mit einer Spezifität für unterschiedliche Triplets miteinander kombiniert, um so die spezifische Erkennung längerer Nukleinsäuresequenzen zu ermöglichen. Durch die Kombination von mindestens zwei oder mehr Zinkfingerdomänen ist es möglich, die nicht spezifische Bindung an in Wirtsorganismen vorliegende Nukleinsäuren zu minimieren und die Spezifität der Zinkfinger enthaltenden Polypeptide, die an die Programm-Nukleinsäure binden, zu erhöhen.

[0069] Zinkfinger bindende Motive sind stabile Strukturen, bei deren Bindung an ihr jeweiliges Ziel es nur selten zu Konformationsänderungen kommt. Im Rahmen der vorliegenden Erfindung wurde nun eine funktionelle Polypeptiddomäne mit mindestens einer Linkersequenz verknüpft, die kovalent an das jeweils relevante Enzym des Biosynthesewegs gebunden ist. Die Linkersequenz bestimmt die Position des Enzyms des Biosynthesewegs oder des weiteren funktionellen Polypeptids in Relation zu der Zinkfinger bindenden Domäne und dem jeweils benachbarten Enzym des Biosynthesewegs oder weiteren funktionellen Polypeptid. Die Linkersequenz kann hinsichtlich ihrer Länge und Aminosäuresequenz variieren und stellt sicher, dass das kovalent gebundene Enzym eine geeignete tertiäre Struktur ausbilden und so seine biologische Funktion beibehalten kann.

[0070] Die Anzahl an chimären Enzymen des Biosynthesewegs oder weiteren funktionellen Polypeptiden, die an die NSBF gebunden sind, kann drei, vier, fünf oder mehr betragen, ist jedoch nicht hierauf beschränkt, und kann sich des Weiteren von der Anzahl an einzelnen Ziel-nukleinsäureelementen auf einer Programm-Nukleinsäuresequenz unterscheiden. Die Anzahl an chimären Enzymen des Biosynthesewegs oder weiteren funktionellen Polypeptiden, die an die NSBF gebunden sind, ist abhängig von der Anzahl der Schritte in einem Biosyntheseweg. Erfordert der Biosyntheseweg drei, vier oder fünf unterschiedliche Enzyme oder weitere funktionelle Polypeptide, um das Substrat in das Produkt oder einen Produktvorläufer umzusetzen, so sind jeweils mindestens drei, vier oder fünf unterschiedliche Bindungselemente in einer festgelegten Reihenfolge in der Programm-Nukleinsäuresequenz eingeschlossen. Das Nukleinsäureprogramm, das einige Wiederholungen der einzelnen Zielelemente enthält, ist ebenfalls in den Umfang der vorliegenden Erfindung eingeschlossen.

[0071] In einigen Ausführungsformen der vorliegenden Erfindung können RNA-Moleküle als Programm-Nukleinsäuresequenzen verwendet werden. In diesem Fall sollten Nukleinsäure bindende Faktoren verwendet werden, die anstelle von spezifischen DNA-Sequenzen spezifische RNA-Sequenzen erkennen. Ohne hierauf beschränkt zu sein, kann es sich bei diesen Sequenzen um die folgenden handeln: das RNA-Erkennungsmotiv (RRM (RNA recognition motif), auch bezeichnet als das RBD- oder das RNP-Motiv), die heterogene nukleäre RNP-K-Homologie-Domäne (KH-Domäne), Zinkfingerdomänen (die am besten charakterisierten RNA bindenden Zinkfingerdomänen sind die des Typs CCHH und CCCH, es können jedoch auch andere Arten verwendet werden) und Pumilio (Puf)-Domänen. Wenngleich RNA bindende Domänen, die spezifische sekundäre Strukturen auf RNA-Molekülen erkennen, ebenfalls als Nukleinsäure bindende Faktoren verwendet werden können (z. B. S1-Domänen, jedoch nicht auf diese beschränkt), so sind sequenzspezifische RNA bindende Domänen am besten geeignet (z. B. die Zinkfinger RNA bindenden Domänen, jedoch nicht auf diese beschränkt). Im

Rahmen der vorliegenden Erfindung ist die Verwendung von RNA bindende Domänen, die spezifische doppelsträngige RNA-Sequenzen erkennen, bevorzugt, da mit diesen eine geeignete räumliche Anordnung gewährleistet werden kann. Verfahren zur Konstruktion von sequenzspezifischen doppelsträngigen RNA bindenden Zinkfingern sind in der Literatur beschrieben.

KONSTRUKTION VON SYNTHESWEGEN.

[0072] Die vorliegende Erfindung betrifft ein Verfahren zur Steigerung der Ausbeute bei der Herstellung von Biosyntheseprodukten. Das Verfahren kombiniert: (i) eine Programm-Nukleinsäuresequenz; und (ii) chimäre Proteine von Enzymen des Biosynthesewegs oder weitere funktionelle Polypeptide mit NSBF. Die erforderlichen Komponenten (i) und (ii) können (i) in vitro in einer Lösung vermischt werden, in Gegenwart von Substraten und Kofaktoren für die Enzyme des Biosynthesewegs, oder (ii) die Komponenten können in vivo in Wirtszellen eingebracht und dort exprimiert werden.

[0073] Im Rahmen der vorliegenden Erfindung bezieht sich der Begriff eines Biosynthesewegs auf eine beliebige Kaskade von Enzymen, die in der festgelegten Reihenfolge eine neue chemische Bindung an das Substrat ausbilden, z. B. natürlich vorkommend oder künstlich konstruiert. Der Begriff eines künstlichen Biosynthesewegs bezieht sich auf die Herstellung eines Produkts, das als solches in der Natur nicht vorkommt. Der Begriff des künstlichen Biosynthesewegs bezieht sich ebenfalls auf die Herstellung eines Produkts, das zwar als solches in der Natur vorkommt, wobei jedoch die Enzyme des Biosynthesewegs nicht von dem gleichen Organismus stammen oder mithilfe genetischer Modifikationen erhalten werden.

[0074] Die vorliegende Erfindung betrifft einen Biosyntheseweg, also eine beliebige gekoppelte enzymatische Reaktion, die sequentiell mit drei oder mehr Enzymen zu erfolgen hat. Der Biosyntheseweg (d. h. dessen Fluss, Effizienz) kann gemessen werden, indem die Konzentration von Produktvorläufern und Endprodukten mittels der im Stand der Technik bekannten geeigneten Verfahren bestimmt wird.

[0075] Die vorliegende Erfindung betrifft einen beliebigen Biosyntheseweg, einschließend, jedoch nicht beschränkt auf:

- [0076]** • katabolische oder anabolische Synthesewege,
- [0077]** • primäre metabolische Synthesewege, wie z. B. die Synthese von Aminosäuren, Fettsäuren, Kohlenhydraten, Pyrimidinen und Purinen, jedoch nicht auf diese beschränkt,
- [0078]** • sekundäre metabolische Synthesewege, wie z. B. die Synthese von sekundären Metaboliten mit biologischen Aktivitäten und pharmakologischen Eigenschaften (Polyketiden), seltenen Peptiden, Hormonen, Terpenoiden (Carotenoiden, ...), Pigmenten (Violacein, Melanin, . . .) und dergleichen, jedoch nicht auf diese beschränkt.

CAROTENOIDE ENZYME DES BIOSYNTHESWEGS

[0079] Carotenoide gehören zur Kategorie der Tetraterpenoide. Astaxanthin, Zeaxanthin und Canthaxanthin sind von β -Caroten abgeleitet und ihre Synthese wird durch die carotenoiden Biosyntheseenzyme crtE, crtB, crtI, crtY, crtO und crtZ vermittelt, wobei die Reihenfolge und die Beteiligung der Enzyme jeweils ausschlaggebend für die Art des erhaltenen Produkts ist. So führen z. B. Biosynthesewege, die crtE, crtB, crtI, crtY, crtO und crtZ einschließen, zur Synthese von Astaxanthin, crtE, crtB, crtI, crtY und crtZ führen zur Synthese von Zeaxanthin und crtL-E, crtB, crtI, crtY und crtO zur Synthese von Canthaxanthin.

Violacein-Biosyntheseweg

[0080] Die Gene für die Biosynthese des Pigments Violacein befinden sich in einem Operon, das sich aus vioE-, vioD-, vioC-, vioB- und vioA-Genen zusammensetzt. VioA ist eine von FAD abhängige L-Trp- Oxidase, die IPA-Imin erzeugt. Letzteres ist das Substrat für VioB, eine Hämoxygenase-Oxidase, die das IPA-Imin in eine Verbindung mit einer bisher nicht bekannten chemischen Formel umsetzt (Verbindung X). Bei VioE handelt es sich um ein erst vor kurzem

entdecktes einzigartiges Protein, für das bisher keine Homologa charakterisiert wurden und das ein Schlüsselenzym des Violacein- Biosynthesewegs darstellt - seine Wirkung besteht darin, dass es die Verbindung X in Zwischenprodukte umwandelt, die dann von VioD und VioC übernommen werden können, bei denen es sich wiederum um von FAD abhängige Monooxygenasen handelt, die diese Verbindungen zum Erhalt von Violacein hydroxylieren.

Resveratrol-Biosyntheseweg

[0081] Bei Resveratrol (3,5,4'-Transhydroxystilben) handelt es sich um ein in Pflanzen produziertes Polyphenol. Der Resveratrol-Biosyntheseweg besteht aus vier Enzymen: Phenylalaninammoniumlyase (PAL) und Zimtsäure 4-Hydroxylase (C4H), die durch ein einziges Enzym, die Tyrosinammoniumlyase (TAL), ersetzt werden können, 4- Cumarat:CoA-Ligase (4CL) und Stilbensynthase (STS). Die ersten beiden Enzyme, PAL und C4H, wandeln die Aminosäure Phenylalanin in p-Cumarsäure (4-Cumarsäure) um. Das dritte Enzym, 4CL, bindet die p-Cumarsäure an die Pantetheingruppe des Coenzyms A (CoA), was zum Erhalt von 4-Cumaroyl-CoA führt. Das abschließende Enzym in diesem Biosyntheseweg, STS, katalysiert die Kondensation von Resveratrol von einem Molekül 4-Cumaroyl-CoA und drei Molekülen Malonyl-CoA, die aus der Biosynthese von Fettsäuren stammen.

[0082] Im Rahmen der vorliegenden Erfindung werden Enzyme aus bestimmten Biosynthesewegen, such wie dem Violacein-, Resveratrol- und Carotenoid-Biosyntheseweg, an NSBF fusioniert, um diese chimären Proteine räumlich aneinander anzunähern. Durch die sequentielle Bindung von Enzymen, so wie gesteuert durch die Programm-Nukleinsäuresequenz, ist es erfindungsgemäß nun möglich, eine spezifische Reihenfolge der enzymatischen Reaktionen festzulegen. Dadurch werden schnellere und effizientere Reaktionen ermöglicht und es besteht des Weiteren die Möglichkeit zur Erzeugung eines neuen Produkts durch die Umordnung einer spezifischen Reihenfolge von Proteinreaktionen (so wird z. B. mit einer Reihenfolge 1-2-3 der Enzyme eine andere Verbindung synthetisiert als mit der Reihenfolge 1-3-2). Die vorliegende Erfindung ist dahingehend vorteilhaft, dass sie die Herstellung von künstlichen Verbindungen mit gewünschten Eigenschaften ermöglicht.

Rekombinante DNA

[0083] Im Rahmen der vorliegenden Erfindung werden molekularbiologische Standardverfahren eingesetzt, die im Stand der Technik allgemein bekannt sind.

[0084] Die erfindungsgemäßen Proteine aus dem Polypeptidmaterial können durch die Expression von für Proteine kodierender DNA in einem geeigneten Wirtsorganismus synthetisiert werden. Die für die Proteine kodierende DNA wird in einen geeigneten Expressionsvektor eingebracht. Geeignete Vektoren schließen Plasmide, virale Vektoren und dergleichen ein, sind jedoch nicht auf diese beschränkt. Expressionsvektoren, die mit den Zellen des Wirtsorganismus kompatibel sind, sind im Stand der Technik hinreichend bekannt und schließen die jeweils geeigneten Kontrollelemente für die Transkription und Translation von Nukleinsäuresequenzen ein.

[0085] Es können Expressionsvektoren zur Expression in prokaryotischen und eukaryotischen Zellen konstruiert werden. Zu den prokaryotischen Zellen zählen z. B. Bakterien, insbesondere Escherichia coli. Gemäß der vorliegenden Erfindung werden prokaryotische Zellen zum Erhalt einer ausreichenden Menge an Nukleinsäure verwendet. Ein Expressionsvektor enthält im Allgemeinen die operativ assoziierten Kontrollelemente, die funktionell mit der für das Protein kodierenden DNA gemäß der vorliegenden Erfindung verknüpft sind. Die Kontrollelemente sind so ausgewählt, dass sie eine effiziente und gewebespezifische Expression induzieren. Der Promotor kann konstitutiv oder induzierbar sein, je nach dem gewünschten Muster der Expression. Der Promotor kann nativ oder fremden Ursprungs (nicht natürlich in den Zellen vorkommend, in denen er eingesetzt wird) sowie natürlich oder synthetisch sein. Die Auswahl des Promotors hat so zu erfolgen, dass dieser in den Zielzellen des Wirtsorganismus funktioniert. Des Weiteren sind Initiationssignale für die effiziente Translation eines Fusionsproteins eingeschlossen, welche das Startcodon ATG und die entsprechenden Sequenzen umfassen.

Schließt der im Rahmen der vorliegenden Erfindung verwendete Vektor zwei oder mehr Leserahmen ein, so sollten die Leserahmen jeweils unabhängig voneinander funktionell mit den Kontrollelementen verknüpft sein und die Kontrollelemente sollten gleich oder voneinander verschieden sein, je nach der gewünschten Art der Proteinherstellung.

[0086] Vor einer weiteren Beschreibung der vorliegenden Erfindung wird darauf hingewiesen, dass die vorliegende Erfindung nicht auf spezifische, hierin beschriebene Ausführungsformen beschränkt ist, die selbstverständlich gegebenenfalls variiert werden können. Des Weiteren wird darauf hingewiesen, dass die hierin verwendete Terminologie lediglich der Beschreibung bestimmter Ausführungsformen dient und nicht als einschränkend zu interpretieren ist, da der Umfang der vorliegenden Erfindung ausschließlich durch die beigefügten Patentansprüche abgegrenzt wird.

[0087] Ist ein Bereich von Werten angegeben, so wird davon ausgegangen, dass jeder zwischen der oberen und der unteren Grenze des Bereichs liegende Wert sowie jeder weitere beliebige angegebene oder dazwischen liegende Wert innerhalb des angegebenen Bereichs bis zu einem Zehntel der Einheit der unteren Grenze in den Umfang der vorliegenden Erfindung eingeschlossen ist, sofern der entsprechende Kontext nicht ausdrücklich etwas anderes besagt. Die oberen und unteren Grenzen dieser kleineren Bereiche können unabhängig voneinander in die kleineren Bereiche eingeschlossen sein und sind ebenfalls von der vorliegenden Erfindung umfasst, unter Vorbehalt ausdrücklich ausgeschlossener Grenzen innerhalb des angegebenen Bereichs. Schließt der angegebene Bereich eine oder beide der Grenzen ein, so sind Bereiche, die entweder eine oder beide dieser eingeschlossenen Grenzen ausschließen, ebenfalls in der vorliegenden Erfindung eingeschlossen.

[0088] Sofern nicht anderweitig definiert, haben alle hierin verwendeten technischen und wissenschaftlichen Begriffe die Bedeutung, in der sie von einem durchschnittlichen Fachmann auf dem Gebiet der vorliegenden Erfindung verstanden werden. Bei der praktischen Ausführung oder beim Testen der vorliegenden Erfindung können beliebige Verfahren oder Materialien verwendet werden, die den hierin beschriebenen ähnlich oder äquivalent sind, im Folgenden werden jedoch die bevorzugten Verfahren und Materialien beschrieben. Alle hierin genannten Veröffentlichungen sind durch Bezugnahme in die vorliegende Erfindung eingeschlossen und offenbaren und beschreiben die Verfahren und/oder Materialien in dem Zusammenhang, in dem die Veröffentlichungen genannt werden.

[0089] Es sei ferner darauf hingewiesen, dass die hierin sowie in den beigefügten Patentansprüchen im Singular verwendeten unbestimmten und bestimmten Artikel "ein" und "der/die/das", einschließlich aller ihrer grammatikalischen Formen, den jeweiligen Plural ebenfalls einschließen, sofern der jeweilige Kontext dies nicht ausdrücklich ausschließt. So schließt z. B. eine Bezugnahme auf "ein Enzym" eine Vielzahl derartiger Enzyme ein und eine Bezugnahme auf "das Gerüst-Polypeptid" schließt eine Bezugnahme auf ein oder mehrere Gerüst-Polypeptide sowie deren im Stand der Technik bekannte Äquivalente ein, und so weiter. Des Weiteren wird darauf hingewiesen, dass die Patentansprüche so verfasst sein können, dass ein beliebiges Element ausgeschlossen ist. Diese Aussage soll als Grundlage für die Verwendung von ausschließenden Begriffen wie "ausschließlich", "alleinig", "nur" oder dergleichen in Verbindung mit der Aufzählung von Elementen der Patentansprüche bzw. für die Anwendung einer "negativen" Einschränkung dienen.

[0090] Die hierin besprochenen Veröffentlichungen werden nur aufgrund der Tatsache genannt, dass sie bereits vor dem Anmeldetag der vorliegenden Anmeldung offenbart wurden. Nichts herein ist als Eingeständnis dahingehend auszulegen, dass die vorliegende Erfindung nicht den Anspruch hätte, eine solche Veröffentlichung kraft einer früheren Erfindung vorwegzunehmen.

[0091] Des Weiteren können die genannten Veröffentlichungsdaten von den tatsächlichen Veröffentlichungsdaten abweichen, die gegebenenfalls unabhängig festzustellen sind.

BEISPIELE:

KLONIERUNG UND AUFREINIGUNG VON POLYPEPTIDEN

[0092] Die Konstruktion von DNA-Sequenzen für wie vorstehend beschriebene Nukleinsäure bindende Faktoren und Enzyme des Biosynthesewegs oder weitere funktionelle Polypeptide erfolgte auf Grundlage der Aminosäuresequenzen der ausgewählten Proteindomänen unter Verwendung des Tools "Designer" von DNA2.0 Inc. Dieses Tool ermöglicht es dem Benutzer, DNA-Fragmente zu konstruieren und die Expression mittels einer für den betreffenden Organismus spezifischen Codon-Optimierung für den gewünschten Wirt (z. B. E. coli) zu optimieren. Die Gene wurden von GENEART AG (Im Gewerbepark B32, D-93059 Regensburg) bezogen, mit Restriktionsenzymen verdaut und gemäß der im Stand der Technik bekannten Vorgehensweise in einen geeigneten Vektor kloniert, der die geeigneten regulatorischen Sequenzen enthielt. DNA-Sequenzen für gespaltene fluoreszierende Proteine wurden mittels PCR unter Verwendung einer DNA-Sequenz von fluoreszierenden Proteinen als Matrize amplifiziert (Zugangsnummer oder Referenz erforderlich). Die verwendeten Vektoren schließen ein: die kommerziell erhältlichen Vektoren pET, pBluescript, pCD-NA, pSBIA2 (<http://partsregistry.org/Part:pSB1A2>) pSB1C3 (<http://partsregistry.org/wiki/index.php?title=Part:pSB1C3>), pS- B1AK3 (<http://partsregistry.org/wiki/index.php?title=Part:PSB1AK3>) sowie die Plasmidvektoren mit niedriger Kopienzahl pSB4K5 (<http://partsregistry.org/Part:pSB4K5>) oder pSB4C5 (<http://partsregistry.org/Part:pSB4C5>), die alle erforderlichen Merkmale tragen, wie z. B. Antibiotikaresistenz, Replikationsursprung und multiple Klonierungsstelle.

[0093] Zur Herstellung von DNA-Konstrukten wurden molekularbiologische Verfahren angewandt (DNA-Fragmentierung mit Restriktionsenzymen, DNA-Amplifikation unter Anwendung einer PCR (Polymerase-Kettenreaktion), PCR-Ligation, Detektion der DNA-Konzentration, Agarosegelelektrophorese, Aufreinigung von DNA-Fragmenten aus Agarosegels, Ligation von DNA-Fragmenten in einen Vektor, Transformation der chemisch kompetenten Zellen E. coli DH5 α , Isolation von Plasmid-DNA mit kommerziell erhältlichen Kits, Screening und Selektionierung). Alle Verfahren wurden unter sterilen Bedingungen durchgeführt (aseptische Technik). Die Charakterisierung von DNA-Segmenten erfolgte anhand von Restriktionsanalyse und Sequenzierung.

[0094] Verfahren der molekularen Klonierung sind im Stand der Technik hinreichend bekannt sowie in zahlreichen Lehrbüchern der Molekularbiologie ausführlich beschrieben, die dem Fachmann ebenfalls bekannt sind.

[0095] Die DNA-Konstrukte und entsprechenden chimären Proteine sind in Tabelle 1 beschrieben. In allen DNA-Konstrukten ist das Startcodon (ATG) vor den Histidin-Tags oder kodierenden Regionen angeordnet. Die für die chimären Proteine kodierenden Konstrukte wurden entweder zur Expression großer Mengen in den Vektor pET19b oder zur Überwachung der Produktion von Carotinoiden und Violacein in vivo in die Vektoren pSB4K5 bzw. pSB4C5 kloniert. Die Expressionskassette schließt in Richtung vom 5'- zum 3'-Ende den T7-Promotor, eine multiple Klonierungsstelle für Fusionsprotein und den T7-Terminator ein. Diese regulatorischen Elemente ermöglichen die Proteinexpression in der prokaryotischen Zelllinie E. coli, die die T7-RNA-Polymerase trägt.

[0096] Die Herstellung der DNA-Konstrukte erfolgte unter Anwendung von molekularbiologischen Verfahren, wie sie im Wesentlichen in zahlreichen Lehrbüchern der Molekularbiologie beschrieben und dem Fachmann bekannt sind. Die Transformation von Plasmiden, Konstrukten und Zwischenkonstrukten erfolgte mittels einer chemischen Transformation in das Bakterium E. coli DH5 α oder BL21 (DE3) pLysS.

[0097] Es wurden mehrere Konstrukte hergestellt, um die Durchführbarkeit der Produktion der in Tabelle 1 aufgeführten chimären Proteine zu demonstrieren.

[0098] Die Transformation von Plasmiden, die für offene Leserahmen der in Tabelle 1 aufgeführten Fusionsproteine (SEQ ID NR: 52 bis SEQ ID NR: 57) kodieren, erfolgte mittels einer

chemischen Transformation in kompetente *E. coli* BL21 (DE3) pLysS-Zellen. Ausgewählte Bakterienkolonien, die mit einem ausgewählten Antibiotikum (Ampicillin) auf LB-Platten kultiviert worden waren, wurden in 10 mL LB-Kulturbrühe inokuliert, die mit dem ausgewählten Antibiotikum ergänzt worden war. Im Anschluss an ein mehrstündiges Wachstum bei 37°C wurden 10 bis 100 µl der Kultur in 100 ml eines ausgewählten Wachstumsmediums inokuliert und über Nacht bei 37 °C unter Schütteln stehen gelassen. Die über Nacht angesetzte Kultur wurde 20 bis 50 Mal verdünnt, bis eine OD₆₀₀ der verdünnten Kultur von zwischen 0,1 und 0,2 erreicht war. Kulturflaschen mit 500 ml der verdünnten Kultur wurden auf der Schüttelvorrichtung platziert und die Bakterien wurden bei 25°C oder 30°C oder 37°C so lange kultiviert, bis eine OD₆₀₀ von zwischen 0,6 und 0,8 erreicht war; dann wurde die Proteinexpression durch die Zugabe von Induktor-IPTG (1 mM) eingeleitet. Vier Stunden nach der Induktion wurde die Kulturbrühe zentrifugiert und die Bakterienzellen wurden in Lysepuffer (10 mM Tris pH 8,0, 1 M NaCl, 0,1% Desoxycholat, 0,1 mM Zinksulfat, 10 mM DTT, ergänzt mit einem Proteaseinhibitorcocktail) resuspendiert und mindestens über Nacht bei -80°C eingefroren. Die aufgetaute Zellsuspension wurde durch Beschallung lysiert und im Anschluss zentrifugiert. Das Präzipitat (Zellmembranen, Einschlusskörperchen) und der Überstand wurden unter Verwendung von SDS-PAGE und Western-Blot hinsichtlich der Expression der erfindungsgemäß hergestellten Konstrukte untersucht, nötigenfalls unter Verwendung von anti-His-Tag-Antikörpern als primäre Antikörper. Die konstruierten Fusionsproteine waren vorwiegend im nicht löslichen Teil (Einschlusskörperchen) zu finden, der sich zu mehr als 80% aus dem ausgewählten Protein zusammensetzte. Die Einschlusskörperchen wurden zwei Mal mit Lysepuffer, zwei Mal mit 1 M Harnstoff in 10 mM Tris pH 8,0 und zwei Mal mit 2 M Harnstoff in 10 mM Tris pH 8,0 gewaschen. Üblicherweise führte diese Behandlung zu einer Proteinreinheit von mehr als 95%. Aufgereinigte Einschlusskörperchen wurden in 8 M Harnstoff in 50 mM Tris pH 8,0 gelöst. Die Proteine wurden mit 8 M Harnstoff in 50 mM Tris pH 8,0 und 250 mM Imidazol eluiert. Die Proteinrückfaltung erfolgte mittels einer Dialyse gegen Puffer enthaltend 50 mM Tris pH 8,0, 500 mM NaF, 500 µM ZnSO₄, 5 mM DTT, 0,005 % Tween und 10 % Glycerin. Die Dialyse wurde mindestens zwei Mal für 12 h bei 4°C ohne Mischen durchgeführt. Nach dem Abschluss des Prozesses wurde das Dialysat vorsichtig aus den Dialyseröhrchen entnommen und in Zentrifugenröhrchen überführt. Nicht rückgefaltetes Protein präzipitierte und wurde mittels Zentrifugation bei 10.000 rpm von dem rückgefalteten Protein in der löslichen Phase getrennt. Die Bestimmung der Proteinkonzentration erfolgte unter Verwendung eines Spektrophotometers sowie des Bradford-Proteinassays und die Größe der Proteine wurde mittels SDS-PAGE/Coomassie-Färbung und Western-Blot-Analyse unter Einsatz von anti-His-Antikörpern bestimmt.

[0099] Tabelle 1 : Zusammensetzung von Plasmiden, die zur Veranschaulichung der vorliegenden Erfindung hergestellt wurden.

SEQ ID NR:	SEQ ID NR:	Zusammensetzung des Konstrukts	Plasmid rückgrat
DNA	Protein	Bindungsfaktoren	
1	2	Znf_Gli1	pSBAK3
3	4	Znf_HivC	pSBAK3
5	6	Znf_Zif268	pSBAK3
7	8	Znf_Jazz	pSBAK3
9	10	Znf_PSBII	pSBAK3
11	12	Znf_Blues	pSBAK3
13	14	Znf_Tyr	pSBAK3
15	16	NicTAL	pSBAK3
Enzyme des Biosynthesewegs oder weitere funktionelle Polypeptide			
17	18	vioA	pSBIA2
19	20	vioB	pSBIA2
39	40	vioC	pSB1A2

41	42	vioD	pSB1A2
43	44	vioE	pSB1A2
45	46	nCFP (mCerulean, N-terminal)	pSBAK3
47	48	nYFP (mCitrine, N-terminal)	pSBAK3
49	50	cCFP (mCerulean, C-terminal)	pSBAK3
51	52	cYFP (mCitrine, C-terminal)	pSBAK3
Hilfspolypeptide			
53	54	Peptidlinker	pSBAK3
55		His_tag	pSBAK3
Programm-Nukleinsäuresequenz			
56		Programm-SPR, Split, FRET	pBluescript
57		Programm-NICTAL	pBluescript
58		Programm-Biosynthese (123456)	pBluescript
59		Programm-Biosynthese (12346)	pBluescript
60		Programm-Biosynthese (341256)	pBluescript
Chimäre Proteine			
61	62	Znf_Blues-Peptidlinker-vioA	pSB1A2
63	64	Znf_Zif268-Peptidlinker-vioB	PSB1A2
65	66	Znf_PBSII-Peptidlinker-vioE	PSB1A2
67	68	Znf_HivC-Peptidlinker-vioD	PSB1A2
69	70	Znf_Gli1-Peptidlinker-vioC	pSB1A2
71	72	His_tag-Znf_Gli1-Peptidlinker-nCFP	pSBAK8
73	74	His_tag-Znf_Blues-Peptidlinker-nCFP	pSBAK8
75	76	His_tag-Znf_PBSII_Peptidlinker-nYFP	pSBAK8
77	78	cCFP-Peptidlinker-Znf_HIVC-His_tag	pSBAK8
79	80	cCFP-Peptidlinker-Znf_Jazz-Peptidlinker-His_tag	pSBAK8
81	82	CYFP-Peptidlinker-Znf_Zif268-Peptidlinker-His_tag	pSBAK8

BINDUNG EINES BINDUNGSFAKTORS AN EIN ZIELNUKLEINSÄUREELEMENT

β-Galactosidase-Enzym-Assay

[00100] Die Inkubation von Bakterienkulturen, die die in Tabelle 1 aufgeführten Plasmidkonstrukte trugen, erfolgte in 10 ml LB-Kulturbrühe, ergänzt mit 5 µl IPTG (1 M) (Isopropyl-β-D-thiogalactosid) und ansteigenden Konzentrationen von (0%, 0,0025%, 0,1% und 1%) L-Arabinose, auf einem Rotationsschüttler für 18 Stunden bei 37 °C und 180 rpm. Die Bakterienkonzentration wurde anhand der optischen Dichte (OD₆₀₀) bestimmt. Die Messungen der Aktivität der β-Galactosidase erfolgten in einem ELISA-Lesegerät, das auf 28°C vorgeheizt worden war. 5 µl einer jeden Kultur wurden in die Wells einer 96-Well-Mikrotiterplatte mit transparentem Boden überführt und es erfolgte die Zugabe von 100 µl Z-Puffer mit Chloroform (Z-Puffer: 0,06 M Na₂HPO₄ x 7 H₂O, 0,04 M NaH₂PO₄ x H₂O, 0,1 M KCl, 0,001 M MgSO₄ x 7 H₂O, pH 7; Z-Puffer mit Chloroform: Z-Puffer, 1% β-Mercaptoethanol, 10% Chloroform). Die Lyse der Bakterienzellen erfolgte durch die Zugabe von 50 µl Z-Puffer mit SDS (Z-Puffer, 1,6% SDS), gefolgt von einer Inkubation für 10 min bei 28°C. In jedes Well wurden 50 µl einer 0,4%-igen ONPG-Lösung in Z-Puffer zugegeben und die Veränderungen der OD₄₀₅ wurden über einen Zeitraum von 20 min in Intervallen von 30 sec gemessen.

ERGEBNISSE

[00101] Um zu testen, ob es in vivo zu einer Bindung von NSBF (Zinkfinger Zif268, Blues, PBSII, HivC) oder His-Nictal an ein entsprechendes spezifisches Nukleinsäureziel kommt, wurde im Rahmen der vorliegenden Erfindung ein Reportersystem konstruiert. Der Reporter

enthält ein Plasmid, das sich aus a) lacZ unter einem synthetischen Promotor, der eine DNA-Bindungsstelle für das bestimmte untersuchte DNA bindende Protein enthält, und b) einem entsprechenden DNA bindenden Protein unter der Kontrolle des Arabinose-Promotors zusammensetzt. Die erfolgreiche Bindung des NSBF an den synthetischen Promotor würde die Transkription von lacZ verhindern, was wiederum eine verringerte Aktivität der β -Galactosidase zur Folge hätte. Es wurden E. coli-Kulturen, die Plasmide für verschiedene NSBF enthielten, über Nacht in LB-Medium kultiviert, das mit ansteigenden Konzentrationen von Arabinose ergänzt wurde. Die Aktivität der β -Galactosidase wurde gemäß der Beschreibung im Text gemessen. In allen getesteten Konstrukten nahm die Aktivität der β -Galactosidase mit ansteigenden Konzentrationen von Arabinose ab, was darauf schließen lässt, dass die NSBF (Tabelle 1) in vivo an das Zielnukleinsäureelement innerhalb des synthetischen Promotors binden (Figur 3A). Des Weiteren kam es in der Gegenwart einer Fehlpaarung zwischen dem NSBF und der DNA-Zielsequenz (es wurden z. B. jeweils das Blues-Zielelement - blues_0 mit dem PBSII-Zielelement PBSII_0 und das HivC_0-Zielelement mit dem Gli_0-Zielelement vertauscht) zu keiner Beeinträchtigung der Aktivität der β -Galactosidase, was die Spezifität des Testsystems gemäß der vorliegenden Erfindung bestätigt (Figur 3B).

Oberflächenplasmonresonanz (SPR)

[00102] Die SPR-Analyse wurde auf dem Gerät Biacore T100 (GE Healthcare) durchgeführt. Ein Serie S Sensorchip SA (GE Healthcare) wurde mit einer biotinylierten einzelsträngigen DNA-Sonde bei einer Konzentration von 100 nM immobilisiert, bis eine abschließende Reaktion von etwa 300 RU erreicht war. Die Hybridisierung von komplementärer DNA (Programm-DNA-Sequenz mit Bindungsstellen für die Zinkfinger HivC, Zif268, JAZZ, Blues, PBS, Gli1) erfolgte mittels einer Injektion von zwischen 0,5 und 2 μ M in Laufpuffer (10 mM HE-PES, 150 mM NaCl, 0,1 mM EDTA, 0,005% P20, pH 7,4) für bis zu 300 s, um die abschließende Reaktion von etwa 300 RU zu erzielen. Die Bindung von Analyten wurde anhand der Injektion von unterschiedlichen Konzentrationen über 1 min im Anschluss an die Dissoziation detektiert, die über einen Zeitraum von 5 min beobachtet wurde. Die Regeneration der Oberfläche wurde mittels zweier 30-sekündiger Injektionen von 50 mM NaOH und einer 24-sekündigen Injektion von 0,5% SDS erreicht. Im Anschluss hieran wurde die neue DNA injiziert, um eine frische Bindungsoberfläche zu erhalten.

[00103] Die Immobilisierung des Serie S Sensorchips CM5 (GE Healthcare) erfolgte gemäß den Anweisungen des Herstellers mit 3000 RU Avidin über eine Aminkopplung. Die carboxymethylierte Oberfläche des CM5 Chips wurde mittels einer 7-minütigen Pulsinjektion von NHS:EDC in einem Verhältnis von 1:1 aktiviert. Dann erfolgte die Injektion von Avidin in Na-Acetat pH 5,5 in Form einiger kurzer Pulse, um eine Reaktion von 3000 RU zu erzielen. Das Avidin war nur an die zweite Durchflussszelle gekoppelt, während die erste Durchflussszelle als Referenzoberfläche diente, um die nicht spezifische Bindung von Analyten an die Dextranmatrix auf einem Sensorchip zu subtrahieren. Die Stellen auf der Sensoroberfläche, an denen keine Reaktion erfolgt war, wurden mit einem 7-minütigen Injektionspuls von 1 M Ethanolamin (pH 8,5) blockiert.

ERGEBNISSE:

[00104] Die in vitro Bindung von chimären Proteinen, Biosyntheseeenzymen, die an NSBF auf der Programm-DNA gekoppelt sind (Figur 4). Figur 4A zeigt die Bindung von Programm-DNA, die Bindung von chimärem Protein enthaltend den Gli1-NSBF und die Regeneration des Trägers. Die Hybridisierung von DNA erfolgte mittels einer Injektion von zwischen 0,5 und 2 μ M für bis zu 300 s, um die abschließende Reaktion von 300 RU zu erzielen. Die Bindung von Analyten wurde im Anschluss an eine 5-minütige Dissoziation anhand der Injektion von unterschiedlichen Konzentrationen über 1 min detektiert. Die Regeneration wurde mittels zweier 30-sekündiger Injektionen von 50 mM NaOH und einer 24-sekündigen Injektion von 0,5% SDS erreicht.

[00105] Figur 4B zeigt die Bindung von Analyten an Programm-DNA; frische DNA wurde für jedes der Proteine eingefangen. Die Analyten von oben her sind: Gli1, Zif268, BLUES,

ZNF_HIVC, PBSII und JAZZ.

[00106] Die sequentielle Bindung von NSBF auf der Programm-DNA ist in Figur 4C gezeigt. DNA wurde jeweils zu Beginn des Zyklus eingefangen und die Injektion der Proteine erfolgte über 1 Minute nacheinander in 2 unterschiedlichen Sequenzen: durchgezogene Linie: Znf_Jazz, Znf_Blues, Znf_Zif268, Znf_PBSII, Znf_HivC und Znf_Gli1; gestrichelte Linie: Znf_Gli1, Znf_PBSII, Znf_HivC, Znf_Jazz, Znf_Blues und Znf_Zif268. Anhand der Bindung aller Analyten in der Abfolge wurde gezeigt, dass die Bindung der Analyten schwächer war, wenn Gli1 als erstes Protein injiziert wurde.

EMSA - elektrophoretischer Mobilitäts-Shift-Assay

[00107] Die spezifische DNA-Bindung von synthetischen Zinkfingerdomänen an die Programm-Nukleinsäure wurde unter Anwendung eines elektrophoretischen Mobilitäts-Shift-Assays getestet. 1 µg der aufgereinigten Proteine wurde für 3 Stunden mit ansteigenden Mengen an Programm-Nukleinsäure inkubiert, d. h. mit 500 ng, 750 ng und 1000 ng. Die Proben wurden mit Laborwasser mit hohem Reinheitsgrad auf 20 µl verdünnt, auf ein 2,0% Agarosegel geladen, das zuvor mit Ethidiumbromid gefärbt worden war, und bei 70 V für 40 Minuten laufen gelassen. Die Detektion von Nukleinsäure-Protein-Komplexen erfolgte unter UV-Licht.

ERGEBNISSE:

[00108] Die Migration von mit Ethidiumbromid gefärbter Programm-Nukleinsäure war abhängig von der Gegenwart von NSBF (d. h. der DNA bindenden Zinkfingerdomäne), was zu einer langsameren Migration führte als bei der alleinigen Verwendung von Programm-Nukleinsäure (Figur 5). Diese Daten belegen die Bindung von NSBF an eine spezifische Zielsequenz innerhalb eines DNA-Programms (Figur 5).

Konfokalmikroskopie

[00109] Zur Detektion von chimären, gespaltenen, fluoreszierenden Proteinen (z. B. NSBF gekoppelt an nCFP, NSBF gekoppelt an cCFP) sowie zur FRET-Rekonstitution auf einem Nukleinsäureprogramm in vivo wurde das konfokale Lasermikroskop Leica TCS SP5 verwendet. Transfizierte HEK293-Zellen, die über Nacht bei 37°C auf einem 8-Well- Mikroskopträger kultiviert worden waren, wurden auf einer Objektivlinse platziert, auf der sich ein Tropfen Immersionsöl befand. Die Anregung der Zellen erfolgte bei 433 nm für das CFP und 515 nm für das YFP. Die Detektion des FRET erfolgte gemäß der durch die Software des Herstellers zur Verfügung gestellten Vorgehensweise (die Assistenten FRET AB und FRET SE der Leica LAS AF Computersoftware).

Split, FRET-Experimente

[00110] HEK293-Zellen wurden unter Verwendung des jetPEI™ Transfektionsreagenzprotokolls mit Säuger-Expressionsvektoren, die gespaltene fluoreszierende Fusionsproteine unter der Kontrolle des CM- V-Promotors trugen, transfiziert und abschließend mit 4% Paraformaldehyd fixiert. Die Rekonstitution von gespaltenen fluoreszierenden Proteinen ebenso wie der FRET-Effekt wurden unter einem konfokalen Mikroskop beobachtet (siehe den Abschnitt "Konfokalmikroskopie", supra). Das CFP (Znf_Gli1-Peptidlinker-nCFP und cCFP- Peptidlinker-Znf_HivC) und das YFP (Znf_PBSII-Peptidlinker-nYFP und cYFP-Peptidlinker-Znf_Zif268) wurden nur in Gegenwart einer spezifischen, nicht jedoch einer zufällig zusammengesetzten, Programm-Nukleinsäure rekonstituiert.

ERGEBNISSE:

[00111] Gespaltene fluoreszierende Proteine binden an benachbarte Zielnukleotidsequenzen, wenn Plasmide, die gespaltene fluoreszierende Proteinfusionen tragen, mit einem Plasmid kotransfiziert sind, das eine Ziel-Programm-Nukleinsäure trägt. Da gespaltene fluoreszierende Proteine in vivo nicht zu der zufälligen Bildung eines voll ausgebildeten Chromophors in der Lage sind, entsteht das Fluoreszenzsignal als Folge der Spaltung von Protein-Zinkfinger-Fusionen, die an eine DNA Programm-Nukleinsäuresequenz gebunden sind.

[00112] Rekonstitution von gespaltenem GFP. Figur 6 zeigt die in vivo Rekonstitution von zwei nicht funktionellen Fluorophoren (gespaltenen GFP), die nur in der Gegenwart einer Programm-Nukleinsäuresequenz an NSBF gekoppelt werden (Figur 6A).

[00113] Es wurden jeweils 75 ng an Plasmiden, die Znf_GliI-Peptidlin-ker-nCFP und cCFP-Peptidlinker-Znf_HIVC trugen, mit 350 ng an Programm-DNA in HEK293-Zellen kotransfiziert. Die Detektion des Emissionssignals des Cyan fluoreszierenden Proteins erfolgte mittels einer Anregung der Zellen mit 433 nm Laserlicht (Figur 6B).

[00114] Es wurden jeweils 75 ng an Plasmiden, die Fusionen des gespaltenen gelb fluoreszierenden Proteins (YFP) trugen (Znf_PBSII-Peptidlinker-nYFP und cYFP-Peptidlinker-Znf_Zif268) mit 350 ng an Programm-DNA kotransfiziert. Das Emissionssignal des YFP mit einem Spitzenwert bei 529 nm wurde unter einem konfokalen Mikroskop detektiert (Figur 6C).

[00115] Rekonstitution von gespaltenem GFP in vitro. HEK293-Zellen wurden jeweils mit Znf_GliI-Peptidlinker-nCFP, cCFP-Peptidlin-ker-Znf_HivC und Znf_PBSII-Peptidlinker-nYFP, cYFP-Peptidlinker-Znf_Zif268 kotransfiziert. Die Lyse der Zellen erfolgte nach 72 Stunden in 150 µl passivem Lysepuffer von Promega und im Anschluss wurden 50 µg der Programm-Nukleinsäure eingebracht. Unter Verwendung des Lumineszenzspektrometers LS55 von Perkin Elmer wurde ein funktionelles Fluorophor detektiert, nachdem die Zellysate für 18 h bei 4°C mit dem DNA-Programm inkubiert worden waren.

[00116] FRET-Effizienz. Die FRET-Effizienz von vier chimären Proteinen, gespaltenen GFP in Verknüpfung mit NSBF in der Gegenwart von Programm-DNA, ist in Tabelle 2 gezeigt. Der FRET wurde unter Verwendung des Assistenten FRET AB der mit dem konfokalen Lasermikroskop TCS SP5 von Leica mitgelieferten LAS AF Software gemessen. Die Emissionssignale in den CFP- und YFP-Kanälen wurden angepasst. Der Akzeptor wurde bei 515 nm mit Laser photogeleuchtet und es erfolgte daraus eine Berechnung der FRET-Effizienz innerhalb des photogeleuchteten Bereichs. HEK293-Zellen wurden mit Programm-DNA und Znf_GliI-Peptidlinker-nCFP, cCFP-Peptidlinker-Znf_HIVC, Znf_PBSII-Peptidlinker-nYFP sowie Znf_cYFP-Peptidlinker-Znf_Zif268 transfiziert. Als negative Kontrolle wurden Zellen mit cytosolischem CFP und YFP ohne DNA bindende Faktoren transfiziert und als positive Kontrolle wurden Zellen mit CFP transfiziert, das über einen Peptidlinker mit YFP verknüpft war.

[00117] Tabelle 2: Die FRET-Effizienz von vier gespaltenen GFP, die jeweils mit einem unterschiedlichen DNA bindenden Faktor verknüpft sind.

FRET positive Kontrolle	FRET negative Kontrolle	FRET auf einem DNA Programm
23 ± 2	1 ± 1	4 ± 2

Synthese von Violacein

[00118] Über Nacht angesetzte Kulturen von E. coli, die Plasmide enthielten, die für die chimären Proteine kodierte (siehe Tabelle 1, Konstrukt-Kombination Znf_Blues-Peptidlinker-vioA, Znf_Zif268-Peptidlinker-vioB, Znf_PBSII-Peptidlinker-vioE, Znf_HivC-Peptidlinker-vioD, Znf_GliI-Peptidlinker-vioC), entweder mit oder ohne Plasmid, das für das Nukleinsäureprogramm kodierte (siehe Tabelle 1, Programm-Biosynthese (123456) oder Programm-Biosynthese (341256)), wurden in frischer Luria-Bertani-Brühe bis zu einer OD₆₀₀ von 0,05 verdünnt und bei 30°C in der Gegenwart geeigneter Antibiotika kultiviert. Zu verschiedenen Zeitpunkten wurden Proben entnommen (3 ml insgesamt). Der repräsentative Zeitpunkt (17,5 h) ist in Figur 9 gezeigt. Die Lyse der Bakterien erfolgte durch die Zugabe eines identischen Volumens (1,5 ml) an 10% SDS und Violacein wurde mit Ethylacetat (1:1 vol/vol) extrahiert. Nach kurzem Vortexen wurde die organische Phase gesammelt und das Absorptionsspektrum bei 575 nm wurde gemessen. Die Proben wurden mittels TLC analysiert und es erfolgte die quantitative Bestimmung von Violacein mittels Densitometrie bei 575 nm, nachdem die Spektren direkt auf den HPTLC-

Platten gescannt worden waren. Violacein wurde ebenfalls mittels Massenspektrometrie identifiziert. Es erfolgte ein Vergleich der Mengen an Violacein, die jeweils mit dem Nukleinsäureprogramm, ohne das Nukleinsäureprogramm oder in der Gegenwart eines nicht gerichteten Nukleinsäureprogramms produziert wurden (Figur 9).

[00119] Wenngleich die vorliegende Erfindung unter Bezugnahme auf deren spezifische Ausführungsformen beschrieben wurde, so ist es für den Fachmann dennoch selbstverständlich, dass verschiedene Änderungen vorgenommen werden und Äquivalente eingesetzt werden können, ohne damit vom Grundgedanken und Umfang der vorliegenden Erfindung abzuweichen. Darüber hinaus können zahlreiche Modifikationen vorgenommen werden, um eine bestimmte Situation, ein bestimmtes Material, eine bestimmte Stoffzusammensetzung, einen bestimmten Prozess sowie einen oder mehrere bestimmte Prozessschritte an den Gegenstand, Grundgedanken und Umfang der vorliegenden Erfindung anzupassen. Derartige Modifikationen gelten als vom Schutzzumfang der beigefügten Patentansprüche umfasst.

1
SEQUENZAUFLISTUNG

<110> KEMIJSKI INSTITUT
EN-FIST CENTER OD LICNOSTI

<120> Verbesserte Synthese eines biosynthetischen Produkts mit einer
angeordneten Zusammensetzung der biosynthetischen Enzyme auf
einem Nukleotidsequenzmotiv

<130> R 63917

<160> 82

<170> PatentIn version 3.4

<210> 1
<211> 531
<212> DNA
<213> künstliche Sequenz

<220>
<223> znf-Gli1

<220>
<221> CDS
<222> (1)..(531)
<223> znf-Gli1

<400> 1

atg aag cgt gag cct gaa tct gtg tat gaa act gac tgc cgt tgg gat	48
Met Lys Arg Glu Pro Glu Ser Val Tyr Glu Thr Asp Cys Arg Trp Asp	
1 5 10 15	
ggc tgt agc cag gaa ttt gac tcc caa gag cag ctg gtg cac cac atc	96
Gly Cys Ser Gln Glu Phe Asp Ser Gln Glu Leu Val His His Ile	
20 25 30	
aac agc gag cac atc cac ggg gag cgg aag gag ttc gtg tgc cac tgg	144
Asn Ser Glu His Ile His Gly Glu Arg Lys Glu Phe Val Cys His Trp	
35 40 45	
ggg ggc tgc tcc agg gag ctg agg ccc ttc aaa gcc cag tac atg ctg	192
Gly Gly Cys Ser Arg Glu Leu Arg Pro Phe Lys Ala Gln Tyr Met Leu	
50 55 60	
gtg gtt cac atg cgc aga cac act ggc gag aag cca cac aag tgc acg	240
Val Val His Met Arg Arg His Thr Gly Glu Lys Pro His Lys Cys Thr	
65 70 75 80	
ttt gaa ggg tgc cgg aag tca tac tca cgc ctc gaa aac ctg aag acg	288
Phe Glu Gly Cys Arg Lys Ser Tyr Ser Arg Leu Glu Asn Leu Lys Thr	
85 90 95	
cac ctg cgg tca cac acg ggt gag aag cca tac atg tgt gag cac gag	336
His Leu Arg Ser His Thr Gly Glu Lys Pro Tyr Met Cys Glu His Glu	
100 105 110	
ggc tgt agt aaa gcc ttc agc aat gcc agt gac cga gcc aag cac cag	384
Gly Cys Ser Lys Ala Phe Ser Asn Ala Ser Asp Arg Ala Lys His Gln	
115 120 125	
aat cgg acc cat tcc aat gag aag ccg tat gta tgt aag ctc cct ggc	432
Asn Arg Thr His Ser Asn Glu Lys Pro Tyr Val Cys Lys Leu Pro Gly	
130 135 140	
tgc acc aaa cgc tat aca gat cct agc tcg ctg cga aaa cat gtc aag	480
Cys Thr Lys Arg Tyr Thr Asp Pro Ser Ser Leu Arg Lys His Val Lys	
145 150 155 160	

2

aca gtg cat ggt cct gac gcc cat gtg acc aaa cgg cac cgt ggg gat 528
 Thr Val His Gly Pro Asp Ala His Val Thr Lys Arg His Arg Gly Asp
 165 170 175

taa 531

<210> 2

<211> 176

<212> PRT

<213> künstliche Sequenz

<220>

<223> synthetisches Konstrukt

<400> 2

Met Lys Arg Glu Pro Glu Ser Val Tyr Glu Thr Asp Cys Arg Trp Asp
 1 5 10 15

Gly Cys Ser Gln Glu Phe Asp Ser Gln Glu Gln Leu Val His His Ile
 20 25 30

Asn Ser Glu His Ile His Gly Glu Arg Lys Glu Phe Val Cys His Trp
 35 40 45

Gly Gly Cys Ser Arg Glu Leu Arg Pro Phe Lys Ala Gln Tyr Met Leu
 50 55 60

Val Val His Met Arg Arg His Thr Gly Glu Lys Pro His Lys Cys Thr
 65 70 75 80

Phe Glu Gly Cys Arg Lys Ser Tyr Ser Arg Leu Glu Asn Leu Lys Thr
 85 90 95

His Leu Arg Ser His Thr Gly Glu Lys Pro Tyr Met Cys Glu His Glu
 100 105 110

Gly Cys Ser Lys Ala Phe Ser Asn Ala Ser Asp Arg Ala Lys His Gln
 115 120 125

Asn Arg Thr His Ser Asn Glu Lys Pro Tyr Val Cys Lys Leu Pro Gly
 130 135 140

Cys Thr Lys Arg Tyr Thr Asp Pro Ser Ser Leu Arg Lys His Val Lys
 145 150 155 160

Thr Val His Gly Pro Asp Ala His Val Thr Lys Arg His Arg Gly Asp
 165 170 175

<210> 3

<211> 258

<212> DNA

<213> künstliche Sequenz

<220>

<223> znf-HivC

3

<220>
 <221> CDS
 <222> (1)..(258)
 <223> Znf-HivC
 <400> 3
 atg ccc ttc cag tgt cgc att tgc atg cgg aac ttt tcg ctc agg acg 48
 Met Pro Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser Leu Arg Thr
 1 5 10 15
 gac ctt gac agg cat acc cgt act cat acc ggt gaa aaa ccg ttt cag 96
 Asp Leu Asp Arg His Thr Arg Thr His Thr Gly Glu Lys Pro Phe Gln
 20 25 30
 tgt cgg atc tgt atg cga aat ttc tcc ctc agc cag acc ttg cgc cgc 144
 Cys Arg Ile Cys Met Arg Asn Phe Ser Leu Ser Gln Thr Leu Arg Arg
 35 40 45
 cat cta cgt acg cac acc ggc gag aag cca ttc caa tgc cga ata tgc 192
 His Leu Arg Thr His Thr Gly Glu Lys Pro Phe Gln Cys Arg Ile Cys
 50 55 60
 atg cgc aac ttc agt ctc agg agc aac ctg ggg cgg cac cta aaa acc 240
 Met Arg Asn Phe Ser Leu Arg Ser Asn Leu Gly Arg His Leu Lys Thr
 65 70 75 80
 cac aca gga gaa aaa taa 258
 His Thr Gly Glu Lys
 85

<210> 4
 <211> 85
 <212> PRT
 <213> künstliche Sequenz

<220>
 <223> synthetisches Konstrukt

<400> 4
 Met Pro Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser Leu Arg Thr
 1 5 10 15
 Asp Leu Asp Arg His Thr Arg Thr His Thr Gly Glu Lys Pro Phe Gln
 20 25 30
 Cys Arg Ile Cys Met Arg Asn Phe Ser Leu Ser Gln Thr Leu Arg Arg
 35 40 45
 His Leu Arg Thr His Thr Gly Glu Lys Pro Phe Gln Cys Arg Ile Cys
 50 55 60
 Met Arg Asn Phe Ser Leu Arg Ser Asn Leu Gly Arg His Leu Lys Thr
 65 70 75 80
 His Thr Gly Glu Lys
 85

<210> 5
 <211> 270

4

<212> DNA
<213> künstliche Sequenz

<220>
<223> Znf-Zif268

<220>
<221> CDS
<222> (1)..(270)
<223> Znf-Zif268

<400> 5
atg ccg ggt gaa aaa ccg tat gca tgt ccg gtt gaa agc tgt gat cgt 48
Met Pro Gly Glu Lys Pro Tyr Ala Cys Pro Val Glu Ser Cys Asp Arg
1 5 10 15
cgt ttt agc cgt agt gat gaa ctg acc cgt cat att cgt att cat aca 96
Arg Phe Ser Arg Ser Asp Glu Leu Thr Arg His Ile Arg Ile His Thr
20 25 30
ggt cag aaa ccg ttt cag tgt cgt att tgc atg cgt aat ttt agc cgt 144
Gly Gln Lys Pro Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser Arg
35 40 45
tca gat cat ctg acc acc cat att cgt acc cat act ggc gaa aaa ccg 192
Ser Asp His Leu Thr Thr His Ile Arg Thr His Thr Gly Glu Lys Pro
50 55 60
ttt gcc tgt gat att tgt ggt cgt aaa ttt gca cgt tcc gat gaa cgt 240
Phe Ala Cys Asp Ile Cys Gly Arg Lys Phe Ala Arg Ser Asp Glu Arg
65 70 75 80
aaa cgc cat acc aaa att cat acg ggt taa 270
Lys Arg His Thr Lys Ile His Thr Gly
85

<210> 6
<211> 89
<212> PRT
<213> künstliche Sequenz

<220>
<223> synthetisches Konstrukt

<400> 6
Met Pro Gly Glu Lys Pro Tyr Ala Cys Pro Val Glu Ser Cys Asp Arg
1 5 10 15
Arg Phe Ser Arg Ser Asp Glu Leu Thr Arg His Ile Arg Ile His Thr
20 25 30
Gly Gln Lys Pro Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser Arg
35 40 45
Ser Asp His Leu Thr Thr His Ile Arg Thr His Thr Gly Glu Lys Pro
50 55 60
Phe Ala Cys Asp Ile Cys Gly Arg Lys Phe Ala Arg Ser Asp Glu Arg
65 70 75 80
Lys Arg His Thr Lys Ile His Thr Gly

85

5

<210> 7
<211> 288
<212> DNA
<213> künstliche Sequenz

<220>
<223> Znf-Jazz

<220>
<221> CDS
<222> (1)..(288)
<223> Znf-Jazz

<400> 7
atg gca agt gat gat cgt ccg tat gca tgt ccg gtt gaa agc tgt gat 48
Met Ala Ser Asp Asp Arg Pro Tyr Ala Cys Pro Val Glu Ser Cys Asp
1 5 10 15
cgt cgt ttt agc cgt agt gat gaa ctg acc cgt cat att cgt att cat 96
Arg Arg Phe Ser Arg Ser Asp Glu Leu Thr Arg His Ile Arg Ile His
20 25 30
aca ggt cag aaa cct ttt cag tgt cgt att tgc atg cgt aat ttt agc 144
Thr Gly Gln Lys Pro Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser
35 40 45
tca aga gat gtt ctg cgt cgt cat aat cgt acc cat acg ggt gaa aaa 192
Ser Arg Asp Val Leu Arg Arg His Asn Arg Thr His Thr Gly Glu Lys
50 55 60
ccg ttt gca tgt gat att tgc ggt cgt aaa ttt gca agt cgt gac gtg 240
Pro Phe Ala Cys Asp Ile Cys Gly Arg Lys Phe Ala Ser Arg Asp Val
65 70 75 80
ctg cgt cgc cat aac cgc att cat ctg cgt cag aat gat ctg gaa taa 288
Leu Arg Arg His Asn Arg Ile His Leu Arg Gln Asn Asp Leu Glu
85 90 95

<210> 8
<211> 95
<212> PRT
<213> künstliche Sequenz

<220>
<223> synthetisches Konstrukt

<400> 8
Met Ala Ser Asp Asp Arg Pro Tyr Ala Cys Pro Val Glu Ser Cys Asp
1 5 10 15
Arg Arg Phe Ser Arg Ser Asp Glu Leu Thr Arg His Ile Arg Ile His
20 25 30
Thr Gly Gln Lys Pro Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser
35 40 45
Ser Arg Asp Val Leu Arg Arg His Asn Arg Thr His Thr Gly Glu Lys
50 55 60

Pro Phe Ala Cys Asp Ile Cys Gly Arg Lys Phe Ala Ser Arg Asp Val
 65 70 75 80

Leu Arg Arg His Asn Arg Ile His Leu Arg Gln Asn Asp Leu Glu
 85 90 95

<210> 9
 <211> 264
 <212> DNA
 <213> künstliche Sequenz

<220>
 <223> Znf-PSBII

<220>
 <221> CDS
 <222> (1)..(264)
 <223> Znf-PSBII

<400> 9
 atg ccg ggt gaa aaa ccg tat gca tgt cct gaa tgt ggt aaa agc ttt 48
 Met Pro Gly Glu Lys Pro Tyr Ala Cys Pro Glu Cys Gly Lys Ser Phe
 1 5 10 15
 agc cag cgt gca aat ctg cgt gca cat cag cgt acc cat aca ggt gaa 96
 Ser Gln Arg Ala Asn Leu Arg Ala His Gln Arg Thr His Thr Gly Glu
 20 25 30
 aag cct tat aaa tgc cca gaa tgc ggc aaa agc ttt tca cgt agc gat 144
 Lys Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser Arg Ser Asp
 35 40 45
 cat ctg acc acc cat cag cgc aca cat act ggc gag aaa cct tac aaa 192
 His Leu Thr Thr His Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Lys
 50 55 60
 tgt cca gag tgt ggt aaa tca ttt agc cgt agt gat gtt ctg gtt cgt 240
 Cys Pro Glu Cys Gly Lys Ser Phe Ser Arg Ser Asp Val Leu Val Arg
 65 70 75 80
 cat cag aga acc cac acg ggt taa 264
 His Gln Arg Thr His Thr Gly
 85

<210> 10
 <211> 87
 <212> PRT
 <213> künstliche Sequenz

<220>
 <223> synthetisches Konstrukt

<400> 10
 Met Pro Gly Glu Lys Pro Tyr Ala Cys Pro Glu Cys Gly Lys Ser Phe
 1 5 10 15
 Ser Gln Arg Ala Asn Leu Arg Ala His Gln Arg Thr His Thr Gly Glu
 20 25 30
 Lys Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser Arg Ser Asp
 35 40 45

7

His Leu Thr Thr His Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Lys
50 55 60

Cys Pro Glu Cys Gly Lys Ser Phe Ser Arg Ser Asp Val Leu Val Arg
65 70 75 80

His Gln Arg Thr His Thr Gly
85

<210> 11
<211> 288
<212> DNA
<213> künstliche Sequenz

<220>
<223> Znf-Blues

<220>
<221> CDS
<222> (1)..(288)
<223> Znf-Blues

<400> 11
atg gca agt gat gat cgt ccg tat gca tgt ccg gtt gaa agc tgt gat 48
Met Ala Ser Asp Asp Arg Pro Tyr Ala Cys Pro Val Glu Ser Cys Asp
1 5 10 15
cgt cgt ttt agc cgt cgt gat gtt ctg atg aat cac att cgt att cac 96
Arg Arg Phe Ser Arg Arg Asp Val Leu Met Asn His Ile Arg Ile His
20 25 30
aca ggt cag aaa ccg ttt cag tgt cgt att tgc atg cgt aat ttt agc 144
Thr Gly Gln Lys Pro Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser
35 40 45
cgt agc gat cat ctg acc acc cat att cgt acc cat aca ggt gaa aaa 192
Arg Ser Asp His Leu Thr Thr His Ile Arg Thr His Thr Gly Glu Lys
50 55 60
ccg ttt gca tgt gat att tgc ggt cgc aaa ttt gca aat cgt gat acc 240
Pro Phe Ala Cys Asp Ile Cys Gly Arg Lys Phe Ala Asn Arg Asp Thr
65 70 75 80
ctg acc cgt cat agc aaa att cat ctg cgt cag aat gat ctg gaa taa 288
Leu Thr Arg His Ser Lys Ile His Leu Arg Gln Asn Asp Leu Glu
85 90 95

<210> 12
<211> 95
<212> PRT
<213> künstliche Sequenz

<220>
<223> synthetisches Konstrukt

<400> 12

Met Ala Ser Asp Asp Arg Pro Tyr Ala Cys Pro Val Glu Ser Cys Asp
1 5 10 15

Arg Arg Phe Ser Arg Arg Asp Val Leu Met Asn His Ile Arg Ile His
20 25 30

8

Thr Gly Gln Lys Pro Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser
35 40 45

Arg Ser Asp His Leu Thr Thr His Ile Arg Thr His Thr Gly Glu Lys
50 55 60

Pro Phe Ala Cys Asp Ile Cys Gly Arg Lys Phe Ala Asn Arg Asp Thr
65 70 75 80

Leu Thr Arg His Ser Lys Ile His Leu Arg Gln Asn Asp Leu Glu
85 90 95

<210> 13
 <211> 288
 <212> DNA
 <213> künstliche Sequenz

<220>
 <223> Znf-Tyr

<220>
 <221> CDS
 <222> (1)..(288)
 <223> Znf_Tyr

<400> 13	
atg gaa aaa ccg tat aaa tgt cct gaa tgc ggc aaa agc ttt agc cag	48
Met Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser Gln	
1 5 10 15	
agc agc aat ctg gca cgt cat cag cgt acc cat aca ggt gaa aaa cct	96
Ser Ser Asn Leu Ala Arg His Gln Arg Thr His Thr Gly Glu Lys Pro	
20 25 30	
tac aaa tgc cct gag tgt ggt aaa agc ttt tca cgt agc gat cat ctg	144
Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser Arg Ser Asp His Leu	
35 40 45	
acc aaa cat cag cgc aca cat act ggc gag aaa cca ttt aaa tgc cca	192
Thr Lys His Gln Arg Thr His Thr Gly Glu Lys Pro Phe Lys Cys Pro	
50 55 60	
gaa tgt ggg aaa tca ttt agc cag tca agt aat ctg gct cgc cat caa	240
Glu Cys Gly Lys Ser Phe Ser Gln Ser Ser Asn Leu Ala Arg His Gln	
65 70 75 80	
aga acg cat acg ggt ggt ggt agc ggt ggt ggt tca ggt ggt agt taa	288
Arg Thr His Thr Gly Gly Gly Ser Gly Gly Gly Ser Gly Gly Ser	
85 90 95	

<210> 14
 <211> 95
 <212> PRT
 <213> künstliche Sequenz

<220>
 <223> synthetisches Konstrukt

<400> 14

Met Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser Gln

1 5 10 9 15

Ser Ser Asn Leu Ala Arg His Gln Arg Thr His Thr Gly Glu Lys Pro
20 25 30

Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser Arg Ser Asp His Leu
35 40 45

Thr Lys His Gln Arg Thr His Thr Gly Glu Lys Pro Phe Lys Cys Pro
50 55 60

Glu Cys Gly Lys Ser Phe Ser Gln Ser Ser Asn Leu Ala Arg His Gln
65 70 75 80

Arg Thr His Thr Gly Gly Gly Ser Gly Gly Gly Ser Gly Gly Ser
85 90 95

<210> 15
<211> 1710
<212> DNA
<213> künstliche Sequenz

<220>
<223> NicTAL

<220>
<221> CDS
<222> (1)..(1710)
<223> Nic TAL

<400> 15
atg ggt ccg cct ctg caa ctg gat aca ggt cag ctg ctg aaa att gca 48
Met Gly Pro Pro Leu Gln Leu Asp Thr Gly Gln Leu Lys Ile Ala
1 5 10 15

aaa cgt ggt ggt gtt acc gca gtt gaa gca gtt cat gca tgg cgt aat 96
Lys Arg Gly Gly Val Thr Ala Val Glu Ala Val His Ala Trp Arg Asn
20 25 30

gca ctg aca ggt gca ccg ctg aat ctg aca ccg gaa cag gtt gtt gca 144
Ala Leu Thr Gly Ala Pro Leu Asn Leu Thr Pro Glu Gln Val Val Ala
35 40 45

att gca agc cat gat ggt ggt aaa cag gca ctg gaa acc gtt cag cgt 192
Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg
50 55 60

ctg ctg ccg gtt ctg tgt cag gca cat ggt ctg acc cct gaa cag gtg 240
Leu Leu Pro Val Leu Cys Gln Ala His Gly Leu Thr Pro Glu Gln Val
65 70 75 80

gtg gcc att gcc agt aat ggt ggt ggc aaa cag gcg tta gaa aca gtg 288
Val Ala Ile Ala Ser Asn Gly Gly Gly Lys Gln Ala Leu Glu Thr Val
85 90 95

cag cgc ctg ctg cct gtt tta tgc cag gcc cat ggc ctg aca cca gag 336
Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His Gly Leu Thr Pro Glu
100 105 110

cag gta gtg gcg att gcg agc aat att ggc ggt aaa caa gcc ctt gaa 384
Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu
115 120 125

10

acc	gtg	cag	gca	tta	ctg	ccg	gtg	ctg	tgc	caa	gcg	cac	ggc	ctg	acc	432
Thr	Val	Gln	Ala	Leu	Leu	Pro	Val	Leu	Cys	Gln	Ala	His	Gly	Leu	Thr	
	130					135					140					
cca	gaa	caa	gtt	gtt	gcg	atc	gca	tca	aat	ggc	ggc	ggc	aag	cag	gct	480
Pro	Glu	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Gly	Gly	Gly	Lys	Gln	Ala	
145					150					155					160	
ttg	gaa	acg	gtg	cag	cgt	tta	ctg	cca	gtg	tta	tgt	cag	gcg	cat	ggc	528
Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Ala	His	Gly	
				165					170					175		
tta	aca	ccg	gaa	caa	gtg	gtg	gca	atc	gcc	tca	cat	gat	ggc	gga	aaa	576
Leu	Thr	Pro	Glu	Gln	Val	Val	Ala	Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	
			180					185					190			
caa	gca	tta	gag	act	gtt	caa	cgc	ctg	ctg	cca	gtg	ctt	tgc	cag	gca	624
Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Ala	
		195					200					205				
cac	ggc	tta	aca	cct	gaa	cag	gtc	gta	gcg	att	gca	tca	aat	att	ggc	672
His	Gly	Leu	Thr	Pro	Glu	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Ile	Gly	
	210					215					220					
gga	aaa	cag	gcc	ttg	gag	act	gta	cag	gca	ctg	ctg	cct	gta	ctg	tgt	720
Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Ala	Leu	Leu	Pro	Val	Leu	Cys	
225					230					235					240	
caa	gcc	cac	gga	tta	acg	cca	gaa	cag	gtt	gtg	gct	atc	gcc	agc	aat	768
Gln	Ala	His	Gly	Leu	Thr	Pro	Glu	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	
				245					250					255		
tca	ggc	ggc	aaa	cag	gca	ctc	gag	aca	gtt	cag	gcc	ctg	ctg	ccg	gtc	816
Ser	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Ala	Leu	Leu	Pro	Val	
			260					265					270			
tta	tgt	caa	gct	cac	gga	ctg	act	ccc	gag	cag	gtt	gtc	gcc	att	gcc	864
Leu	Cys	Gln	Ala	His	Gly	Leu	Thr	Pro	Glu	Gln	Val	Val	Ala	Ile	Ala	
		275					280					285				
tca	aat	ggc	ggc	ggc	aaa	caa	gct	ctc	gaa	acg	gta	cag	aga	ctg	ctg	912
Ser	Asn	Gly	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	
	290					295					300					
ccc	gtc	ctg	tgc	cag	gcg	cat	gga	ctt	acg	cct	gag	caa	gtt	gtg	gca	960
Pro	Val	Leu	Cys	Gln	Ala	His	Gly	Leu	Thr	Pro	Glu	Gln	Val	Val	Ala	
305					310					315					320	
att	gca	tct	aat	aat	ggg	ggc	aag	caa	gcg	ctg	gaa	aca	gtt	caa	cgc	1008
Ile	Ala	Ser	Asn	Asn	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	
				325					330					335		
tta	ctg	cct	gtc	ttg	tgc	caa	gca	cat	ggc	tta	acc	cca	gag	cag	gtc	1056
Leu	Leu	Pro	Val	Leu	Cys	Gln	Ala	His	Gly	Leu	Thr	Pro	Glu	Gln	Val	
			340					345					350			
gtt	gct	att	gcc	tct	aac	att	gga	ggc	aaa	cag	gct	ctt	gaa	act	gtc	1104
Val	Ala	Ile	Ala	Ser	Asn	Ile	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	
		355					360					365				
cag	gcc	ctg	tta	cct	gtt	ctg	tgc	cag	gct	cac	ggc	ttg	act	cca	gag	1152
Gln	Ala	Leu	Leu	Pro	Val	Leu	Cys	Gln	Ala	His	Gly	Leu	Thr	Pro	Glu	
	370					375					380					
caa	gtg	gtt	gca	ata	gcc	agc	aac	ggc	ggg	ggc	aaa	caa	gct	tta	gaa	1200
Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Gly	Gly	Gly	Lys	Gln	Ala	Leu	Glu	
	385				390				395						400	

11

acc gtc caa cgt ctg tta cca gtg ctg tgt caa gct cat ggc ctt aca	1248
Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His Gly Leu Thr	
405 410 415	
ccc gaa caa gta gtt gcc att gcg agt aac att ggt ggg aag caa gca	1296
Pro Glu Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala	
420 425 430	
ctt gaa acg gtt cag gcg ctg ctt cca gta tta tgc cag gcg cac gga	1344
Leu Glu Thr Val Gln Ala Leu Leu Pro Val Leu Cys Gln Ala His Gly	
435 440 445	
ctc act cca gaa cag gta gta gca atc gca agt aat aac ggt ggg aaa	1392
Leu Thr Pro Glu Gln Val Val Ala Ile Ala Ser Asn Asn Gly Gly Lys	
450 455 460	
caa gcg ttg gag aca gtc caa aga ctg ctt ccg gtt ctt tgc caa gcc	1440
Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala	
465 470 475 480	
cac ggt ctt aca ccg cag cag gtt gta gct att gct agt aat att gga	1488
His Gly Leu Thr Pro Gln Gln Val Val Ala Ile Ala Ser Asn Ile Gly	
485 490 495	
ggt cgt ccg gca ctg gaa agc att gtt gca cag ctg agc cgt cct gat	1536
Gly Arg Pro Ala Leu Glu Ser Ile Val Ala Gln Leu Ser Arg Pro Asp	
500 505 510	
ccg gca ctg gca gca ctg acc aat gat cat ctg gtt gca ctg gca tgt	1584
Pro Ala Leu Ala Ala Leu Thr Asn Asp His Leu Val Ala Leu Ala Cys	
515 520 525	
ctg ggt ggt cgt cct gcc ctg gat gca gtt aaa aaa ggt ctg ccg cat	1632
Leu Gly Gly Arg Pro Ala Leu Asp Ala Val Lys Lys Gly Leu Pro His	
530 535 540	
gca ccg gca ctg att aaa cgt acc aat cgt cgt atc ccg gaa cgt acc	1680
Ala Pro Ala Leu Ile Lys Arg Thr Asn Arg Arg Ile Pro Glu Arg Thr	
545 550 555 560	
agc cat cgt gtt gca gat cat gca cag taa	1710
Ser His Arg Val Ala Asp His Ala Gln	
565	

<210> 16
 <211> 569
 <212> PRT
 <213> künstliche Sequenz

<220>
 <223> synthetisches Konstrukt

<400> 16

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

12

Leu Leu Pro Val Leu Cys Gln Ala His Gly Leu Thr Pro Glu Gln Val
 65 70 75 80
 Val Ala Ile Ala Ser Asn Gly Gly Gly Lys Gln Ala Leu Glu Thr Val
 85 90 95
 Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His Gly Leu Thr Pro Glu
 100 105 110
 Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu
 115 120 125
 Thr Val Gln Ala Leu Leu Pro Val Leu Cys Gln Ala His Gly Leu Thr
 130 135 140
 Pro Glu Gln Val Val Ala Ile Ala Ser Asn Gly Gly Gly Lys Gln Ala
 145 150 155 160
 Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His Gly
 165 170 175
 Leu Thr Pro Glu Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys
 180 185 190
 Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala
 195 200 205
 His Gly Leu Thr Pro Glu Gln Val Val Ala Ile Ala Ser Asn Ile Gly
 210 215 220
 Gly Lys Gln Ala Leu Glu Thr Val Gln Ala Leu Leu Pro Val Leu Cys
 225 230 235 240
 Gln Ala His Gly Leu Thr Pro Glu Gln Val Val Ala Ile Ala Ser Asn
 245 250 255
 Ser Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Ala Leu Leu Pro Val
 260 265 270
 Leu Cys Gln Ala His Gly Leu Thr Pro Glu Gln Val Val Ala Ile Ala
 275 280 285
 Ser Asn Gly Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu
 290 295 300
 Pro Val Leu Cys Gln Ala His Gly Leu Thr Pro Glu Gln Val Val Ala
 305 310 315 320
 Ile Ala Ser Asn Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg
 325 330 335

13

Leu Leu Pro Val Leu Cys Gln Ala His Gly Leu Thr Pro Glu Gln Val
340 345 350

Val Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu Thr Val
355 360 365

Gln Ala Leu Leu Pro Val Leu Cys Gln Ala His Gly Leu Thr Pro Glu
370 375 380

Gln Val Val Ala Ile Ala Ser Asn Gly Gly Gly Lys Gln Ala Leu Glu
385 390 395 400

Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His Gly Leu Thr
405 410 415

Pro Glu Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala
420 425 430

Leu Glu Thr Val Gln Ala Leu Leu Pro Val Leu Cys Gln Ala His Gly
435 440 445

Leu Thr Pro Glu Gln Val Val Ala Ile Ala Ser Asn Asn Gly Gly Lys
450 455 460

Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala
465 470 475 480

His Gly Leu Thr Pro Gln Gln Val Val Ala Ile Ala Ser Asn Ile Gly
485 490 495

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

Pro Ala Leu Ala Ala Leu Thr Asn Asp His Leu Val Ala Leu Ala Cys
515 520 525

Leu Gly Gly Arg Pro Ala Leu Asp Ala Val Lys Lys Gly Leu Pro His
530 535 540

Ala Pro Ala Leu Ile Lys Arg Thr Asn Arg Arg Ile Pro Glu Arg Thr
545 550 555 560

Ser His Arg Val Ala Asp His Ala Gln
565

<210> 17
<211> 1257
<212> DNA
<213> künstliche Sequenz

<220>
<223> VioA

14

<220>
<221> CDS
<222> (1)..(1257)
<223> VioA

<400> 17

atg aaa cat tct tcc gat atc tgc att gtt ggt gct ggt att tct ggt Met Lys His Ser Ser Asp Ile Cys Ile Val Gly Ala Gly Ile Ser Gly 1 5 10 15	48
ttg acg tgc gca agc cat ctg ctg gac agc ccg gca tgc cgt ggt ctg Leu Thr Cys Ala Ser His Leu Leu Asp Ser Pro Ala Cys Arg Gly Leu 20 25 30	96
agc ctg cgt atc ttt gac atg cag caa gaa gcc ggt ggc cgt atc cgc Ser Leu Arg Ile Phe Asp Met Gln Gln Glu Ala Gly Arg Ile Arg 35 40 45	144
agc aaa atg ctg gat ggt aag gca agc att gaa ctg ggc gca ggt cgc Ser Lys Met Leu Asp Gly Lys Ala Ser Ile Glu Leu Gly Ala Gly Arg 50 55 60	192
tac tcc cct cag ttg cac ccg cat ttc caa agc gca atg cag cac tat Tyr Ser Pro Gln Leu His Pro His Phe Gln Ser Ala Met Gln His Tyr 65 70 75 80	240
agc caa aag agc gaa gtc tat ccg ttc acc cag ttg aag ttc aaa tct Ser Gln Lys Ser Glu Val Tyr Pro Phe Thr Gln Leu Lys Phe Lys Ser 85 90 95	288
cac gtg cag caa aag ctg aag cgc gcc atg aat gaa ctg tcc ccg cgt His Val Gln Gln Lys Leu Lys Arg Ala Met Asn Glu Leu Ser Pro Arg 100 105 110	336
ctg aaa gag cat ggt aaa gag agc ttt ttg cag ttt gtc agc cgt tat Leu Lys Glu His Gly Lys Glu Ser Phe Leu Gln Phe Val Ser Arg Tyr 115 120 125	384
caa ggt cac gat agc gcg gtt ggt atg atc cgc tct atg ggt tac gac Gln Gly His Asp Ser Ala Val Gly Met Ile Arg Ser Met Gly Tyr Asp 130 135 140	432
gca ctg ttc ctg ccg gat atc agc gca gaa atg gcc tac gac att gtg Ala Leu Phe Leu Pro Asp Ile Ser Ala Glu Met Ala Tyr Asp Ile Val 145 150 155 160	480
ggt aag cac ccg gag atc cag agc gtg acg gac aac gac gcg aac caa Gly Lys His Pro Glu Ile Gln Ser Val Thr Asp Asn Asp Ala Asn Gln 165 170 175	528
tgg ttt gca gcg gaa acg ggc ttt gct ggt ctg att cag ggc atc aag Trp Phe Ala Ala Glu Thr Gly Phe Ala Gly Leu Ile Gln Gly Ile Lys 180 185 190	576
gct aag gtt aag gcg gca ggt gcg cgt ttt agc ctg ggt tat cgt ctg Ala Lys Val Lys Ala Ala Gly Ala Arg Phe Ser Leu Gly Tyr Arg Leu 195 200 205	624
ctg agc gtc cgt acc gac ggt gac ggc tac ctg ctg caa ctg gca ggt Leu Ser Val Arg Thr Asp Gly Asp Gly Tyr Leu Leu Gln Leu Ala Gly 210 215 220	672
gac gac ggc tgg aaa ctg gag cac cgt acc cgc cat ctg att ctg gcg Asp Asp Gly Trp Lys Leu Glu His Arg Thr Arg His Leu Ile Leu Ala 225 230 235 240	720
att ccg ccg agc gcg atg gcg ggt ttg aat gtt gat ttt cca gaa gcc Ile Pro Pro Ser Ala Met Ala Gly Leu Asn Val Asp Phe Pro Glu Ala 240 245 250	768

[illegible]

```
<210> 18
<211> 418
<212> PRT
<213> künstliche Sequenz
```

<220>
<223> synthetisches Konstrukt

<400> 18

Met Lys His Ser Ser Asp Ile Cys Ile Val Gly Ala Gly Ile Ser Gly
1 5 10 15

Leu Thr Cys Ala Ser His Leu Leu Asp Ser Pro Ala Cys Arg Gly Leu
20 25 30

Ser Leu Arg Ile Phe Asp Met Gln Gln Glu Ala Gly Gly Arg Ile Arg
35 40 45

Ser Lys Met Leu Asp Gly Lys Ala Ser Ile¹⁶ Glu Leu Gly Ala Gly Arg
 50 55 60
 Tyr Ser Pro Gln Leu His Pro His Phe Gln Ser Ala Met Gln His Tyr
 65 70 75 80
 Ser Gln Lys Ser Glu Val Tyr Pro Phe Thr Gln Leu Lys Phe Lys Ser
 85 90 95
 His Val Gln Gln Lys Leu Lys Arg Ala Met Asn Glu Leu Ser Pro Arg
 100 105 110
 Leu Lys Glu His Gly Lys Glu Ser Phe Leu Gln Phe Val Ser Arg Tyr
 115 120 125
 Gln Gly His Asp Ser Ala Val Gly Met Ile Arg Ser Met Gly Tyr Asp
 130 135 140
 Ala Leu Phe Leu Pro Asp Ile Ser Ala Glu Met Ala Tyr Asp Ile Val
 145 150 155 160
 Gly Lys His Pro Glu Ile Gln Ser Val Thr Asp Asn Asp Ala Asn Gln
 165 170 175
 Trp Phe Ala Ala Glu Thr Gly Phe Ala Gly Leu Ile Gln Gly Ile Lys
 180 185 190
 Ala Lys Val Lys Ala Ala Gly Ala Arg Phe Ser Leu Gly Tyr Arg Leu
 195 200 205
 Leu Ser Val Arg Thr Asp Gly Asp Gly Tyr Leu Leu Gln Leu Ala Gly
 210 215 220
 Asp Asp Gly Trp Lys Leu Glu His Arg Thr Arg His Leu Ile Leu Ala
 225 230 235 240
 Ile Pro Pro Ser Ala Met Ala Gly Leu Asn Val Asp Phe Pro Glu Ala
 245 250 255
 Trp Ser Gly Ala Arg Tyr Gly Ser Leu Pro Leu Phe Lys Gly Phe Leu
 260 265 270
 Thr Tyr Gly Glu Pro Trp Trp Leu Asp Tyr Lys Leu Asp Asp Gln Val
 275 280 285
 Leu Ile Val Asp Asn Pro Leu Arg Lys Ile Tyr Phe Lys Gly Asp Lys
 290 295 300
 Tyr Leu Phe Phe Tyr Thr Asp Ser Glu Met Ala Asn Tyr Trp Arg Gly
 305 310 315 320
 Cys Val Ala Glu Gly Glu Asp Gly Tyr Leu Glu Gln Ile Arg Thr His

Ala Ala

```
<220>
<221> CDS
<222> (1)..(2994)
<223> VioB
```

[illegible]

18

ctg gtc ggt gct cgt ttg gca ctg tgg ggt cac tac aat gat tat ctg Leu Val Gly Ala Arg Leu Ala Leu Trp Gly His Tyr Asn Asp Tyr Leu 115 120 125	384
cgt acc acc ttc aat cgt gct cgt tgg gtc gac agc gac ccg acg cgc Arg Thr Thr Phe Asn Arg Ala Arg Trp Val Asp Ser Asp Pro Thr Arg 130 135 140	432
cgt gac gct gca caa atc tat gcg ggc caa ttc acc att agc ccg gct Arg Asp Ala Ala Gln Ile Tyr Ala Gly Gln Phe Thr Ile Ser Pro Ala 145 150 155 160	480
ggt gcc ggt ccg ggt acg ccg tgg ctg ttt acg gca gac att gat gat Gly Ala Gly Pro Gly Thr Pro Trp Leu Phe Thr Ala Asp Ile Asp Asp 165 170 175	528
agc cat ggt gca cgt tgg acg cgt ggc ggc cac att gca gag cgt ggc Ser His Gly Ala Arg Trp Thr Arg Gly Gly His Ile Ala Glu Arg Gly 180 185 190	576
ggc cac ttc ttg gat gaa gag ttt ggt ctg gca cgc ctg ttt cag ttc Gly His Phe Leu Asp Glu Glu Phe Gly Leu Ala Arg Leu Phe Gln Phe 195 200 205	624
tct gtg ccg aaa gat cac cca cat ttt ctg ttt cac ccg ggt ccg ttt Ser Val Pro Lys Asp His Pro His Phe Leu Phe His Pro Gly Pro Phe 210 215 220	672
gat tcc gag gcc tgg cgt cgt ctg caa ttg gct ctg gag gat gac gac Asp Ser Glu Ala Trp Arg Arg Leu Gln Leu Ala Leu Glu Asp Asp Asp 225 230 235 240	720
gtt ctg ggt ctg acc gtg caa tat gcg ttg ttc aat atg agc acc ccg Val Leu Gly Leu Thr Val Gln Tyr Ala Leu Phe Asn Met Ser Thr Pro 245 250 255	768
cct cag ccg aac agc ccg gtt ttt cac gat atg gtc ggt gtt gtc ggt Pro Gln Pro Asn Ser Pro Val Phe His Asp Met Val Gly Val Val Gly 260 265 270	816
ctg tgg cgt cgt ggt gaa ctg gcg agc tac ccg gct ggt cgt ctg ctg Leu Trp Arg Arg Gly Glu Leu Ala Ser Tyr Pro Ala Gly Arg Leu Leu 275 280 285	864
cgt ccg cgt caa ccg ggt ctg ggt gac ctg acc ctg cgc gtc aac ggt Arg Pro Arg Gln Pro Gly Leu Gly Asp Leu Thr Leu Arg Val Asn Gly 290 295 300	912
ggt cgc gtt gcg ctg aat ttg gcg tgt gcc att ccg ttc agc act cgt Gly Arg Val Ala Leu Asn Leu Ala Cys Ala Ile Pro Phe Ser Thr Arg 305 310 315 320	960
gcc gcg cag cca agc gca ccg gac cgc cga ccc ccg acc tgg gtg cca Ala Ala Gln Pro Ser Ala Pro Asp Arg Arg Pro Arg Thr Trp Val Pro 325 330 335	1008
aac tgc cgc tgg gcg atc tgc tgc tgc gtg atg agg acg gcg cac tgt Asn Cys Arg Trp Ala Ile Cys Cys Cys Val Met Arg Thr Ala His Cys 340 345 350	1056
tgg cac gtg tgc cgc agg ctc tgt acc aag act att gga cga atc acg Trp His Val Cys Arg Arg Leu Thr Cys Thr Lys Thr Ile Gly Arg Ile Thr 355 360 365	1104
gta ttg tgg acc tgc cgc tgc tgc gcg aac cgc gtg gta gct tga ccc Val Leu Trp Thr Cys Arg Cys Cys Ala Asn Arg Val Val Ala Pro 370 375 380	1152

19																
tga	gca	gcg	aac	tgg	cgg	agt	ggc	gtg	agc	aag	act	ggg	tca	ccc	aaa	1200
	Ala	Ala	Asn	Trp	Arg	Ser	Gly	Val	Ser	Lys	Thr	Gly	Ser	Pro	Lys	
		385					390					395				
gcg	acg	cgt	cta	acc	tgt	acc	tgg	agg	cac	cgg	atc	gcc	gtc	acg	gtc	1248
Ala	Thr	Arg	Leu	Thr	Cys	Thr	Trp	Arg	His	Arg	Ile	Ala	Val	Thr	Val	
	400					405					410					
gct	ttt	tcc	ctg	aga	gca	tcg	cgc	tgc	gca	gct	act	ttc	gcg	gtg	aag	1296
Ala	Phe	Ser	Leu	Arg	Ala	Ser	Arg	Cys	Ala	Ala	Thr	Phe	Ala	Val	Lys	
	415				420					425					430	
cgc	gtg	cgc	gtc	cgg	ata	tcc	cgc	atc	gta	tcg	agg	gca	tgg	gcc	tgg	1344
Arg	Val	Arg	Val	Arg	Ile	Ser	Arg	Ile	Val	Ser	Arg	Ala	Trp	Ala	Trp	
				435					440					445		
tcg	gcg	tcg	aat	ctc	gtc	agg	atg	gcg	acg	ctg	cgg	aat	ggc	gtc	tga	1392
Ser	Ala	Ser	Asn	Leu	Val	Arg	Met	Ala	Thr	Leu	Arg	Asn	Gly	Val		
			450					455					460			
cgg	gtc	tgc	gtc	cgg	gtc	cgg	cac	gca	ttg	ttc	tgg	acg	atg	gtg	ccg	1440
Arg	Val	Cys	Val	Arg	Val	Arg	His	Ala	Leu	Phe	Trp	Thr	Met	Val	Pro	
			465					470					475			
agg	cga	tcc	ctc	tgc	gtg	ttc	tgc	ctg	acg	att	ggg	cgc	tgg	atg	acg	1488
Arg	Arg	Ser	Leu	Cys	Val	Phe	Cys	Leu	Thr	Ile	Gly	Arg	Trp	Met	Thr	
		480					485					490				
cga	ccg	tcg	aag	aag	tgg	att	acg	cct	ttt	tgt	acc	gcc	acg	tta	tgg	1536
Arg	Pro	Ser	Lys	Lys	Trp	Ile	Thr	Pro	Phe	Cys	Thr	Ala	Thr	Leu	Trp	
	495					500					505					
cgt	att	acg	agc	tgg	tgt	atc	cat	tca	tga	gcg	aca	agg	tgt	ttt	ccc	1584
Arg	Ile	Thr	Ser	Trp	Cys	Ile	His	Ser		Ala	Thr	Arg	Cys	Phe	Pro	
	510				515						520					
tgg	ctg	atc	gtt	gca	aat	gtg	aaa	cgt	acg	cac	gtc	tga	tgt	ggc	aga	1632
Trp	Leu	Ile	Val	Ala	Asn	Val	Lys	Arg	Thr	His	Val		Cys	Gly	Arg	
	525				530					535						
tgt	gtg	atc	cgc	aga	acc	gca	aca	agt	cct	att	aca	tgc	cga	gca	ccc	1680
Cys	Val	Ile	Arg	Arg	Thr	Ala	Thr	Ser	Pro	Ile	Thr	Cys	Arg	Ala	Pro	
	540				545					550					555	
gcg	aac	tgt	cgg	cac	cga	aag	ctc	gtt	tgt	tct	tga	agt	atc	tgg	ccc	1728
Ala	Asn	Cys	Arg	His	Arg	Lys	Leu	Val	Cys	Ser		Ser	Ile	Trp	Pro	
				560					565						570	
acg	tgg	aag	gcc	agg	cac	gcc	tgc	aag	cac	ctc	cgc	cag	cgg	gtc	cgg	1776
Thr	Trp	Lys	Ala	Arg	His	Ala	Cys	Lys	His	Leu	Arg	Gln	Arg	Val	Arg	
				575					580					585		
cac	gca	ttg	aat	cta	aag	ccc	agt	tgg	cgg	cag	agc	tgc	gta	aag	ccg	1824
His	Ala	Leu	Asn	Leu	Lys	Pro	Ser	Trp	Arg	Gln	Ser	Cys	Val	Lys	Pro	
			590					595					600			
tcg	acc	tgg	agc	tgt	ctg	tga	tgc	tgc	aat	acc	tgt	acg	cgg	cgt	ata	1872
Ser	Thr	Trp	Ser	Cys	Leu		Cys	Cys	Asn	Thr	Cys	Thr	Arg	Arg	Ile	
		605					610						615			
gca	ttc	cga	act	atg	cac	agg	gcc	aac	aac	gtg	ttc	gtg	acg	gtg	cgt	1920
Ala	Phe	Arg	Thr	Met	His	Arg	Ala	Asn	Asn	Val	Phe	Val	Thr	Val	Arg	
		620					625					630				
gga	ccg	ccg	agc	agc	tgc	aac	tgg	cgt	gcg	gta	gcg	gtg	acc	gtc	gcc	1968
Gly	Pro	Pro	Ser	Ser	Cys	Asn	Trp	Arg	Ala	Val	Ala	Val	Thr	Val	Ala	
	635					640					645					
gtg	atg	gcg	gta	ttc	gtc	agc	act	gct	gga	aat	tgc	tca	tga	aga	aat	2016

Val 650	Met	Ala	Val	Phe	Val 655	Ser	Thr	Ala	Gly	Asn 660	Cys	Ser		Arg	Asn	
gat Asp 665	tca Ser	tta Leu	cct Pro	ggg Gly	cgt Arg 670	taa	caa Gln	cct Pro	gct Ala	gat Asp 660	ggc Gly 675	cct Pro	ggg Gly	cga Arg	gcc Ala	2064
gtt Val 680	cta Leu	cgc Arg	ggg Gly	tgt Cys	ccc Pro 685	gct Ala	gat Asp	ggg Gly	cga Arg	agc Ser 690	ggc Gly	acg Thr	tca Ser	ggc Gly	gtt Val 695	2112
tgg Trp	cct Pro	gga Gly	cac His	cga Arg 700	gtt Val	cgc Arg	tct Ser	gga Gly	acc Thr 705	gtt Val	tag	cga Arg	aag Lys	cac His	gct Ala 710	2160
ggc Gly	acg Thr	ttt Phe	tgt Cys	tcg Ser 715	tct Ser	gga Gly	atg Met	gcc Ala	gca Ala 720	ctt Leu	tat Tyr	ccc Pro	agc Ser	acc Thr 725	ggg Gly	2208
caa Gln	atc Ile	cat His	cgc Arg 730	gga Gly	ctg Leu	cta Leu	tgc Cys	cgc Arg 735	cat His	tcg Ser	tca Ser	ggc Gly	gtt Val 740	ttt Phe	gga Gly	2256
tct Ser	gcc Ala	gga Gly 745	ctt Leu	gtt Val	tgg Trp	tgg Trp	cga Arg 750	ggc Gly	agg Arg	taa	gcg Ala	tgg Trp	cgg Arg 755	tga	aca Thr	2304
cca Pro	cct Pro	gtt Val	cct Pro 760	gaa Glu	tga	gct Ala	gac Asp	caa Gln	ccg Pro 765	tgc Cys	gca Ala	tcc Ser	ggg Gly	tta Leu 770	tca Ser	2352
act Thr	gga Gly	agt Ser	ttt Phe 775	cga Arg	tcg Ser	cga Arg	ctc Leu	ggc Gly 780	gct Ala	gtt Val	tgg Trp	tat Tyr	tgc Cys 785	att Ile	tgt Cys	2400
gac Asp	cga Arg	tca Ser 790	ggg Gly	cga Arg	agg Arg	tgg Trp	cgc Arg 795	tct Ser	gga Gly	cag Gln	ccc Pro	gca Ala 800	cta Leu	cga Arg	aca Thr	2448
tag	cca Pro	ttt Phe 805	tca Ser	acg Thr	tct Ser	gcg Ala	tga	aat Asn 810	gag Glu	cgc Arg	gcg Ala	tat Tyr	cat His 815	ggc Gly	tca Ser	2496
aag Lys	cgc Arg	acc Thr 820	gtt Val	cga Arg	acc Thr	ggc Gly	gct Ala 825	gcc Ala	ggc Gly	gtt Val	gcg Ala	taa	tcc Ser 830	ggt Gly	tct Ser	2544
gga Gly	tga	gag Glu	ccc Pro 835	ggg Gly	ttg Leu	cca Pro	acg Thr	tgt Cys 840	cgc Arg	aga Arg	cgg Arg	tcg Ser	tgc Cys 845	gcg Ala	tgc Cys	2592
gct Ala	gat Asp	ggc Gly 850	att Ile	gta Val	cca Pro	agg Arg	cgt Arg 855	tta Leu	tga	gct Ala	gat Asp	gtt Val	tgc Cys 860	gat Asp	gat Asp	2640
ggc Gly	gca Ala	gca Ala 865	ctt Leu	cgc Arg	cgt Arg	gaa Glu	acc Thr 870	gct Ala	ggg Gly	tag	ctt Leu	gcg Ala	tcg Ser 875	cag Gln	ccg Pro	2688
cct Pro	gat Asp	gaa Glu 880	cgc Arg	agc Ser	aat Asn	cga Arg	tct Ser 885	gat Asp	gac Asp	cgg Arg	tct Ser	gtt Val 890	gcg Ala	tcc Ser	gct Ala	2736
gag Glu	ctg Leu 895	cgc Arg	gct Ala	gat Asp	gaa Glu	cct Pro 900	gcc Ala	aag Lys	cgg Arg	cat His	cgc Arg 905	cgg Arg	tcg Ser	cac His	ggc Gly	2784
cgg Arg	tcc Ser	gcc Ala	gct Ala	gcc Ala	ggg Gly	tcc Ser	ggt Gly	tga	cac His	ccg Pro	tag	cta Leu	tga	cga Arg	cta Leu	2832

910 915 21 920

cgc gct ggg ctg tcg cat gct ggc acg ccg ttg cga gcg tct gct gga 2880
 Arg Ala Gly Leu Ser His Ala Gly Thr Pro Leu Arg Ala Ser Ala Gly
 925 930 935

gca ggc gag cat gct gga acc ggg ttg gct gcc gga tgc gca gat gga 2928
 Ala Gly Glu His Ala Gly Thr Gly Leu Ala Ala Gly Cys Ala Asp Gly
 940 945 950

gct gct gga ttt cta tcg tcg cca aat gct gga ctt ggc gtg cgg caa 2976
 Ala Ala Gly Phe Leu Ser Ser Pro Asn Ala Gly Leu Gly Val Arg Gln
 955 960 965 970

act gag cgc gag gcc taa 2994
 Thr Glu Arg Glu Ala
 975

<210> 20
 <211> 382
 <212> PRT
 <213> künstliche Sequenz

<220>
 <223> synthetisches Konstrukt

<400> 20

Met Ser Ile Leu Asp Phe Pro Arg Ile His Phe Arg Gly Trp Ala Arg
 1 5 10 15

Val Asn Ala Pro Thr Ala Asn Arg Asp Pro His Gly His Ile Asp Met
 20 25 30

Ala Ser Asn Thr Val Ala Met Ala Gly Glu Pro Phe Asp Leu Ala Arg
 35 40 45

His Pro Thr Glu Phe His Arg His Leu Arg Ser Leu Gly Pro Arg Phe
 50 55 60

Gly Leu Asp Gly Arg Ala Asp Pro Glu Gly Pro Phe Ser Leu Ala Glu
 65 70 75 80

Gly Tyr Asn Ala Ala Gly Asn Asn His Phe Ser Trp Glu Ser Ala Thr
 85 90 95

Val Ser His Val Gln Trp Asp Gly Gly Glu Ala Asp Arg Gly Asp Gly
 100 105 110

Leu Val Gly Ala Arg Leu Ala Leu Trp Gly His Tyr Asn Asp Tyr Leu
 115 120 125

Arg Thr Thr Phe Asn Arg Ala Arg Trp Val Asp Ser Asp Pro Thr Arg
 130 135 140

Arg Asp Ala Ala Gln Ile Tyr Ala Gly Gln Phe Thr Ile Ser Pro Ala
 145 150 155 160

Gly Ala Gly Pro Gly Thr Pro Trp Leu Phe Thr Ala Asp Ile Asp Asp
 165 170 175 22

Ser His Gly Ala Arg Trp Thr Arg Gly Gly His Ile Ala Glu Arg Gly
 180 185 190

Gly His Phe Leu Asp Glu Glu Phe Gly Leu Ala Arg Leu Phe Gln Phe
 195 200 205

Ser Val Pro Lys Asp His Pro His Phe Leu Phe His Pro Gly Pro Phe
 210 215 220

Asp Ser Glu Ala Trp Arg Arg Leu Gln Leu Ala Leu Glu Asp Asp Asp
 225 230 235 240

Val Leu Gly Leu Thr Val Gln Tyr Ala Leu Phe Asn Met Ser Thr Pro
 245 250 255

Pro Gln Pro Asn Ser Pro Val Phe His Asp Met Val Gly Val Val Gly
 260 265 270

Leu Trp Arg Arg Gly Glu Leu Ala Ser Tyr Pro Ala Gly Arg Leu Leu
 275 280 285

Arg Pro Arg Gln Pro Gly Leu Gly Asp Leu Thr Leu Arg Val Asn Gly
 290 295 300

Gly Arg Val Ala Leu Asn Leu Ala Cys Ala Ile Pro Phe Ser Thr Arg
 305 310 315 320

Ala Ala Gln Pro Ser Ala Pro Asp Arg Arg Pro Arg Thr Trp Val Pro
 325 330 335

Asn Cys Arg Trp Ala Ile Cys Cys Cys Val Met Arg Thr Ala His Cys
 340 345 350

Trp His Val Cys Arg Arg Leu Cys Thr Lys Thr Ile Gly Arg Ile Thr
 355 360 365

Val Leu Trp Thr Cys Arg Cys Cys Ala Asn Arg Val Val Ala
 370 375 380

<210> 21
 <211> 78
 <212> PRT
 <213> künstliche Sequenz

<220>
 <223> synthetisches Konstrukt

<400> 21

Ala Ala Asn Trp Arg Ser Gly Val Ser Lys Thr Gly Ser Pro Lys Ala
 1 5 10 15

23

Thr Arg Leu Thr Cys Thr Trp Arg His Arg Ile Ala Val Thr Val Ala
20 25 30

Phe Ser Leu Arg Ala Ser Arg Cys Ala Ala Thr Phe Ala Val Lys Arg
35 40 45

Val Arg Val Arg Ile Ser Arg Ile Val Ser Arg Ala Trp Ala Trp Ser
50 55 60

Ala Ser Asn Leu Val Arg Met Ala Thr Leu Arg Asn Gly Val
65 70 75

<210> 22
<211> 57
<212> PRT
<213> künstliche Sequenz

<220>
<223> synthetisches Konstrukt

<400> 22

Arg Val Cys Val Arg Val Arg His Ala Leu Phe Trp Thr Met Val Pro
1 5 10 15

Arg Arg Ser Leu Cys Val Phe Cys Leu Thr Ile Gly Arg Trp Met Thr
20 25 30

Arg Pro Ser Lys Lys Trp Ile Thr Pro Phe Cys Thr Ala Thr Leu Trp
35 40 45

Arg Ile Thr Ser Trp Cys Ile His Ser
50 55

<210> 23
<211> 18
<212> PRT
<213> künstliche Sequenz

<220>
<223> synthetisches Konstrukt

<400> 23

Ala Thr Arg Cys Phe Pro Trp Leu Ile Val Ala Asn Val Lys Arg Thr
1 5 10 15

His Val

<210> 24
<211> 30
<212> PRT
<213> künstliche Sequenz

<220>
<223> synthetisches Konstrukt

24

<400> 24

Cys Gly Arg Cys Val Ile Arg Arg Thr Ala Thr Ser Pro Ile Thr Cys
1 5 10 15

Arg Ala Pro Ala Asn Cys Arg His Arg Lys Leu Val Cys Ser
20 25 30

<210> 25

<211> 42

<212> PRT

<213> künstliche Sequenz

<220>

<223> synthetisches Konstrukt

<400> 25

Ser Ile Trp Pro Thr Trp Lys Ala Arg His Ala Cys Lys His Leu Arg
1 5 10 15

Gln Arg Val Arg His Ala Leu Asn Leu Lys Pro Ser Trp Arg Gln Ser
20 25 30

Cys Val Lys Pro Ser Thr Trp Ser Cys Leu
35 40

<210> 26

<211> 54

<212> PRT

<213> künstliche Sequenz

<220>

<223> synthetisches Konstrukt

<400> 26

Cys Cys Asn Thr Cys Thr Arg Arg Ile Ala Phe Arg Thr Met His Arg
1 5 10 15

Ala Asn Asn Val Phe Val Thr Val Arg Gly Pro Pro Ser Ser Cys Asn
20 25 30

Trp Arg Ala Val Ala Val Thr Val Ala Val Met Ala Val Phe Val Ser
35 40 45

Thr Ala Gly Asn Cys Ser
50

<210> 27

<211> 8

<212> PRT

<213> künstliche Sequenz

<220>

<223> synthetisches Konstrukt

<400> 27

Arg Asn Asp Ser Leu Pro Gly Arg

25

1

5

<210> 28
<211> 36
<212> PRT
<213> künstliche Sequenz

<220>
<223> synthetisches Konstrukt

<400> 28

Gln Pro Ala Asp Gly Pro Gly Arg Ala Val Leu Arg Gly Cys Pro Ala
1 5 10 15

Asp Gly Arg Ser Gly Thr Ser Gly Val Trp Pro Gly His Arg Val Arg
20 25 30

Ser Gly Thr Val
35

<210> 29
<211> 46
<212> PRT
<213> künstliche Sequenz

<220>
<223> synthetisches Konstrukt

<400> 29

Arg Lys His Ala Gly Thr Phe Cys Ser Ser Gly Met Ala Ala Leu Tyr
1 5 10 15

Pro Ser Thr Gly Gln Ile His Arg Gly Leu Leu Cys Arg His Ser Ser
20 25 30

Gly Val Phe Gly Ser Ala Gly Leu Val Trp Trp Arg Gly Arg
35 40 45

<210> 30
<211> 6
<212> PRT
<213> künstliche Sequenz

<220>
<223> synthetisches Konstrukt

<400> 30

Thr Pro Pro Val Pro Glu
1 5

<210> 31
<211> 42
<212> PRT
<213> künstliche Sequenz

<220>
<223> synthetisches Konstrukt

26

<400> 31

Ala Asp Gln Pro Cys Ala Ser Gly Leu Ser Thr Gly Ser Phe Arg Ser
1 5 10 15

Arg Leu Gly Ala Val Trp Tyr Cys Ile Cys Asp Arg Ser Gly Arg Arg
20 25 30

Trp Arg Ser Gly Gln Pro Ala Leu Arg Thr
35 40

<210> 32

<211> 6

<212> PRT

<213> künstliche Sequenz

<220>

<223> synthetisches Konstrukt

<400> 32

Pro Phe Ser Thr Ser Ala
1 5

<210> 33

<211> 20

<212> PRT

<213> künstliche Sequenz

<220>

<223> synthetisches Konstrukt

<400> 33

Asn Glu Arg Ala Tyr His Gly Ser Lys Arg Thr Val Arg Thr Gly Ala
1 5 10 15

Ala Gly Val Ala
20

<210> 34

<211> 4

<212> PRT

<213> künstliche Sequenz

<220>

<223> synthetisches Konstrukt

<400> 34

Ser Gly Ser Gly
1

<210> 35

<211> 23

<212> PRT

<213> künstliche Sequenz

<220>

<223> synthetisches Konstrukt

<400> 35

27

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

Gly Ile Val Pro Arg Arg Leu
20

<210> 36
<211> 16
<212> PRT
<213> künstliche Sequenz

<220>
<223> synthetisches Konstrukt

<400> 36

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

<210> 37
<211> 45
<212> PRT
<213> künstliche Sequenz

<220>
<223> synthetisches Konstrukt

<400> 37

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

Ser Val Ala Ser Ala Glu Leu Arg Ala Asp Glu Pro Ala Lys Arg His
20 25 30

Arg Arg Ser His Gly Arg Ser Ala Ala Ala Gly Ser Gly
35 40 45

<210> 38
<211> 55
<212> PRT
<213> künstliche Sequenz

<220>
<223> synthetisches Konstrukt

<400> 38

Arg Leu Arg Ala Gly Leu Ser His Ala Gly Thr Pro Leu Arg Ala Ser
1 5 10 15

Ala Gly Ala Gly Glu His Ala Gly Thr Gly Leu Ala Ala Gly Cys Ala
20 25 30

Asp Gly Ala Ala Gly Phe Leu Ser Ser Pro Asn Ala Gly Leu Gly Val
35 40 45

Arg Gln Thr Glu Arg Glu Ala
50 55

<210> 39
<211> 1290
<212> DNA
<213> künstliche Sequenz

<220>
<223> vioc

<220>
<221> CDS
<222> (1)..(1290)
<223> vioc

<400> 39
atg aaa cgt gcg att atc gtt ggt ggc ggc ctg gcg ggt ggc ctg acc 48
Met Lys Arg Ala Ile Ile Val Gly Gly Gly Leu Ala Gly Gly Leu Thr
1 5 10 15
gcg atc tac ctg gcg aag cgt ggc tac gaa gtg cac gtc gtg gag aag 96
Ala Ile Tyr Leu Ala Lys Arg Gly Tyr Glu Val His Val Val Glu Lys
20 25 30
cgt ggt gat cct ctg cgc gat ctg agc tct tac gtg gac gtt gtt agc 144
Arg Gly Asp Pro Leu Arg Asp Leu Ser Ser Tyr Val Asp Val Val Ser
35 40 45
agc cgt gcg atc ggc gtg agc atg acc gtt cgt ggt atc aag agc gtt 192
Ser Arg Ala Ile Gly Val Ser Met Thr Val Arg Gly Ile Lys Ser Val
50 55 60
ttg gct gcg ggc att ccg cgt gca gag ctg gat gcg tgt ggc gaa ccg 240
Leu Ala Ala Gly Ile Pro Arg Ala Glu Leu Asp Ala Cys Gly Glu Pro
65 70 75 80
atc gtg gca atg gct ttc tcc gtg ggt ggt cag tat cgc atg cgc gaa 288
Ile Val Ala Met Ala Phe Ser Val Gly Gly Gln Tyr Arg Met Arg Glu
85 90 95
ctg aag ccg ttg gag gat ttc cgt ccg ctg agc ttg aac cgt gcg gcg 336
Leu Lys Pro Leu Glu Asp Phe Arg Pro Leu Ser Leu Asn Arg Ala Ala
100 105 110
ttt caa aag ctg ctg aac aaa tac gcg aac ctg gca ggc gtt cgt tac 384
Phe Gln Lys Leu Leu Asn Lys Tyr Ala Asn Leu Ala Gly Val Arg Tyr
115 120 125
tac ttt gag cat aag tgc ctg gat gtt gac ctg gat ggt aag agc gtg 432
Tyr Phe Glu His Lys Cys Leu Asp Val Asp Leu Asp Gly Lys Ser Val
130 135 140
ttg att cag ggc aaa gat ggt cag ccg cag cgt ctg caa ggt gac atg 480
Leu Ile Gln Gly Lys Asp Gly Gln Pro Gln Arg Leu Gln Gly Asp Met
145 150 155 160
att atc ggt gcg gat ggc gcc cac agc gcc gtc cgt cag gcg atg cag 528
Ile Ile Gly Ala Asp Gly Ala His Ser Ala Val Arg Gln Ala Met Gln
165 170 175
agc ggc ctg cgt cgt ttc gag ttc cag caa acg ttc ttc cgc cat ggc 576
Ser Gly Leu Arg Arg Phe Glu Phe Gln Gln Thr Phe Phe Arg His Gly
180 185 190
tac aaa acc ctg gtt ttg ccg gac gcg caa gca ctg ggt tac cgt aaa 624
Tyr Lys Thr Leu Val Leu Pro Asp Ala Gln Ala Leu Gly Tyr Arg Lys
195 200 205

29
 gac acg ctg tac ttt ttc ggc atg gat tcc ggt ggc ctg ttc gcg ggt 672
 Asp Thr Leu Tyr Phe Phe Gly Met Asp Ser Gly Gly Leu Phe Ala Gly
 210 215 220
 cgt gcg gct acg atc cca gat ggt agc gtc agc atc gcc gtt tgc ctg 720
 Arg Ala Ala Thr Ile Pro Asp Gly Ser Val Ser Ile Ala Val Cys Leu
 225 230 235 240
 ccg tac tcg ggt agc cct tcc ctg acg acc acc gac gaa ccg acg atg 768
 Pro Tyr Ser Gly Ser Pro Ser Leu Thr Thr Thr Asp Glu Pro Thr Met
 245 250 255
 cgt gcg ttc ttc gat cgt tac ttc ggt ggc ctg ccg cgt gac gcg cgt 816
 Arg Ala Phe Phe Asp Arg Tyr Phe Gly Gly Leu Pro Arg Asp Ala Arg
 260 265 270
 gac gaa atg ctg cgt cag ttt ctg gcg aag ccg agc aac gac ctg att 864
 Asp Glu Met Leu Arg Gln Phe Leu Ala Lys Pro Ser Asn Asp Leu Ile
 275 280 285
 aac gtg cgc tct agc acc ttt cac tat aag ggt aat gtg ctg ttg ctg 912
 Asn Val Arg Ser Ser Thr Phe His Tyr Lys Gly Asn Val Leu Leu Leu
 290 295 300
 ggt gat gct gcg cat gcg act gcg ccg ttc ctg ggt cag ggt atg aac 960
 Gly Asp Ala Ala His Ala Thr Ala Pro Phe Leu Gly Gln Gly Met Asn
 305 310 315 320
 atg gcg ctg gag gac gcc cgc acg ttt gtc gag ctg ctg gac cgc cac 1008
 Met Ala Leu Glu Asp Ala Arg Thr Phe Val Glu Leu Leu Asp Arg His
 325 330 335
 cag ggc gac caa gac aaa gcc ttt ccg gag ttc acg gag ctg cgc aaa 1056
 Gln Gly Asp Gln Asp Lys Ala Phe Pro Glu Phe Thr Glu Leu Arg Lys
 340 345 350
 gtc cag gca gac gca atg caa gac atg gct cgc gcc aac tat gac gtt 1104
 Val Gln Ala Asp Ala Met Gln Asp Met Ala Arg Ala Asn Tyr Asp Val
 355 360 365
 ttg agc tgc tcg aac ccg atc ttt ttc atg cgt gcg cgt tac acg cgt 1152
 Leu Ser Cys Ser Asn Pro Ile Phe Phe Met Arg Ala Arg Tyr Thr Arg
 370 375 380
 tac atg cat tcc aag ttt ccg ggc ctg tat ccg ccg gat atg gcc gag 1200
 Tyr Met His Ser Lys Phe Pro Gly Leu Tyr Pro Pro Asp Met Ala Glu
 385 390 395 400
 aaa ctg tac ttt acg agc gag ccg tac gat cgt ctg caa caa atc cag 1248
 Lys Leu Tyr Phe Thr Ser Glu Pro Tyr Asp Arg Leu Gln Gln Ile Gln
 405 410 415
 cgt aaa cag aat gtt tgg tac aag att ggt cgc gtg aat tga 1290
 Arg Lys Gln Asn Val Trp Tyr Lys Ile Gly Arg Val Asn
 420 425

<210> 40
 <211> 429
 <212> PRT
 <213> künstliche Sequenz

<220>
 <223> synthetisches Konstrukt

<400> 40

Met Lys Arg Ala Ile Ile Val Gly Gly Gly Leu Ala Gly Gly Leu Thr
 1 5 10 15

30

Ala Ile Tyr Leu Ala Lys Arg Gly Tyr Glu Val His Val Val Glu Lys
20 25 30

Arg Gly Asp Pro Leu Arg Asp Leu Ser Ser Tyr Val Asp Val Val Ser
35 40 45

Ser Arg Ala Ile Gly Val Ser Met Thr Val Arg Gly Ile Lys Ser Val
50 55 60

Leu Ala Ala Gly Ile Pro Arg Ala Glu Leu Asp Ala Cys Gly Glu Pro
65 70 75 80

Ile Val Ala Met Ala Phe Ser Val Gly Gly Gln Tyr Arg Met Arg Glu
85 90 95

Leu Lys Pro Leu Glu Asp Phe Arg Pro Leu Ser Leu Asn Arg Ala Ala
100 105 110

Phe Gln Lys Leu Leu Asn Lys Tyr Ala Asn Leu Ala Gly Val Arg Tyr
115 120 125

Tyr Phe Glu His Lys Cys Leu Asp Val Asp Leu Asp Gly Lys Ser Val
130 135 140

Leu Ile Gln Gly Lys Asp Gly Gln Pro Gln Arg Leu Gln Gly Asp Met
145 150 155 160

Ile Ile Gly Ala Asp Gly Ala His Ser Ala Val Arg Gln Ala Met Gln
165 170 175

Ser Gly Leu Arg Arg Phe Glu Phe Gln Gln Thr Phe Phe Arg His Gly
180 185 190

Tyr Lys Thr Leu Val Leu Pro Asp Ala Gln Ala Leu Gly Tyr Arg Lys
195 200 205

Asp Thr Leu Tyr Phe Phe Gly Met Asp Ser Gly Gly Leu Phe Ala Gly
210 215 220

Arg Ala Ala Thr Ile Pro Asp Gly Ser Val Ser Ile Ala Val Cys Leu
225 230 235 240

Pro Tyr Ser Gly Ser Pro Ser Leu Thr Thr Thr Asp Glu Pro Thr Met
245 250 255

Arg Ala Phe Phe Asp Arg Tyr Phe Gly Gly Leu Pro Arg Asp Ala Arg
260 265 270

Asp Glu Met Leu Arg Gln Phe Leu Ala Lys Pro Ser Asn Asp Leu Ile
275 280 285

31

Asn Val Arg Ser Ser Thr Phe His Tyr Lys Gly Asn Val Leu Leu Leu
290 295 300

Gly Asp Ala Ala His Ala Thr Ala Pro Phe Leu Gly Gln Gly Met Asn
305 310 315 320

Met Ala Leu Glu Asp Ala Arg Thr Phe Val Glu Leu Leu Asp Arg His
325 330 335

Gln Gly Asp Gln Asp Lys Ala Phe Pro Glu Phe Thr Glu Leu Arg Lys
340 345 350

Val Gln Ala Asp Ala Met Gln Asp Met Ala Arg Ala Asn Tyr Asp Val
355 360 365

Leu Ser Cys Ser Asn Pro Ile Phe Phe Met Arg Ala Arg Tyr Thr Arg
370 375 380

Tyr Met His Ser Lys Phe Pro Gly Leu Tyr Pro Pro Asp Met Ala Glu
385 390 395 400

Lys Leu Tyr Phe Thr Ser Glu Pro Tyr Asp Arg Leu Gln Gln Ile Gln
405 410 415

Arg Lys Gln Asn Val Trp Tyr Lys Ile Gly Arg Val Asn
420 425

<210> 41
<211> 1122
<212> DNA
<213> künstliche Sequenz

<220>
<223> VioD

<220>
<221> CDS
<222> (1)..(1122)
<223> VioD

<400> 41	
atg aag att ctg gtc att ggt gct ggt cca gct ggt ctg gtt ttc gca	48
Met Lys Ile Leu Val Ile Gly Ala Gly Pro Ala Gly Leu Val Phe Ala	
1 5 10 15	
tcc caa ctg aag cag gca cgc cct ttg tgg gcc att gac atc gtg gag	96
Ser Gln Leu Lys Gln Ala Arg Pro Leu Trp Ala Ile Asp Ile Val Glu	
20 25 30	
aag aat gac gag caa gaa gtg ctg ggc tgg ggt gtc gtg ctg cct ggc	144
Lys Asn Asp Glu Gln Glu Val Leu Gly Trp Gly Val Val Leu Pro Gly	
35 40 45	
cgt ccg ggt cag cac ccg gcg aac ccg ctg tcc tat ctg gat gca ccg	192
Arg Pro Gly Gln His Pro Ala Asn Pro Leu Ser Tyr Leu Asp Ala Pro	
50 55 60	
gag cgt ctg aat ccg caa ttt ctg gag gac ttc aaa ctg gtg cat cat	240

Glu 65	Arg	Leu	Asn	Pro	Gln 70	Phe	Leu	Glu	Asp 75	Phe 32 75	Lys	Leu	Val	His 80	His		
aat Asn	gag Glu	ccg Pro	tcc Ser	ttg Leu	atg Met	tcc Ser	acg Thr	ggc Gly	gtt Val	ttg Leu	ttg Leu	tgc Cys	ggc Gly	gtg Val	gag Glu		288
cgt Arg	cgc Arg	ggt Gly	ctg Leu	gtt Val	cac His	gcg Ala	ctg Leu	cgc Arg	gat Asp	aag Lys	tgc Cys	cgc Arg	agc Ser	caa Gln	ggc Gly		336
att Ile	gct Ala	att Ile	cgt Arg	ttc Phe	gaa Glu	agc Ser	ccg Pro	ttg Leu	ctg Leu	gaa Glu	cac His	ggt Gly	gag Glu	ctg Leu	ccg Pro		384
ctg Leu	gcg Ala	gac Asp	tat Tyr	gat Asp	ctg Leu	gtg Val	gtc Val	ctg Leu	gct Ala	aat Asn	ggt Gly	gtt Val	aat Asn	cac His	aaa Lys		432
acc Thr	gcg Ala	cat His	ttc Phe	acc Thr	gag Glu	gct Ala	ctg Leu	gtc Val	ccg Pro	cag Gln	gtg Val	gac Asp	tac Tyr	ggc Gly	cgc Arg		480
aat Asn	aag Lys	tac Tyr	att Ile	tgg Trp	tat Tyr	ggc Gly	act Thr	agc Ser	cag Gln	ctg Leu	ttc Phe	gat Asp	cag Gln	atg Met	aat Asn		528
ctg Leu	gtt Val	ttt Phe	cgt Arg	acc Thr	cat His	ggt Gly	aaa Lys	gat Asp	atc Ile	ttt Phe	atc Ile	gcg Ala	cat His	gcc Ala	tat Tyr		576
aag Lys	tat Tyr	agc Ser	gat Asp	acc Thr	atg Met	agc Ser	acg Thr	ttc Phe	att Ile	gtc Val	gaa Glu	tgt Cys	agc Ser	gaa Glu	gag Glu		624
act Thr	tac Tyr	gca Ala	cgc Arg	gca Ala	cgc Arg	ctg Leu	ggc Gly	gaa Glu	atg Met	tcc Ser	gaa Glu	gag Glu	gcg Ala	agc Ser	gca Ala		672
gaa Glu	tac Tyr	gtt Val	gct Ala	aag Lys	gtg Val	ttc Phe	cag Gln	gcc Ala	gag Glu	ctg Leu	ggt Gly	ggt Gly	cac His	ggc Gly	ctg Leu		720
gtg Val	agc Ser	cag Gln	ccg Pro	ggt Gly	ctg Leu	ggt Gly	tgg Trp	cgt Arg	aac Asn	ttc Phe	atg Met	acg Thr	ttg Leu	tct Ser	cat His		768
gac Asp	cgt Arg	tgt Cys	cat His	gat Asp	ggt Gly	aag Lys	ttg Leu	gtt Val	ctg Leu	ctg Leu	ggt Gly	gac Asp	gcg Ala	ctg Leu	caa Gln		816
agc Ser	ggt Gly	cac His	ttt Phe	agc Ser	atc Ile	ggc Gly	cac His	ggc Gly	acc Thr	acg Thr	atg Met	gcc Ala	gtg Val	gtg Val	gtg Val		864
gcg Ala	cag Gln	ctg Leu	ctg Leu	gtt Val	aaa Lys	gcg Ala	ctg Leu	tgt Cys	acc Thr	gaa Glu	gat Asp	ggt Gly	gtg Val	cct Pro	gcc Ala		912
gcg Ala	ctg Leu	aaa Lys	cgt Arg	ttc Phe	gaa Glu	gag Glu	cgt Arg	gcc Ala	ctg Leu	ccg Pro	ctg Leu	gtg Val	cag Gln	ttg Leu	ttc Phe		960
cgt Arg	ggc Gly	cac His	gca Ala	gac Asp	aac Asn	agc Ser	cgc Arg	gtt Val	tgg Trp	ttc Phe	gaa Glu	acc Thr	gtc Val	gaa Glu	gag Glu		1008
cgc Arg	atg Met	cac His	ctg Leu	tcc Ser	tcg Ser	gcg Ala	gaa Glu	ttt Phe	gtg Val	caa Gln	agc Ser	ttc Phe	gac Asp	gca Ala	cgc Arg		1056

340 345 33 350

cgc aaa agc ctg ccg ccg atg ccg gaa gca ctg gcg cag aat ctg cgt 1104
 Arg Lys Ser Leu Pro Pro Met Pro Glu Ala Leu Ala Gln Asn Leu Arg
 355 360 365

tat gct ttg cag cgc tga 1122
 Tyr Ala Leu Gln Arg
 370

<210> 42
 <211> 373
 <212> PRT
 <213> künstliche Sequenz

<220>
 <223> synthetisches Konstrukt

<400> 42

Met Lys Ile Leu val ile Gly Ala Gly Pro Ala Gly Leu val phe Ala
 1 5 10 15

Ser Gln Leu Lys Gln Ala Arg Pro Leu Trp Ala ile Asp ile val Glu
 20 25 30

Lys Asn Asp Glu Gln Glu val Leu Gly Trp Gly val val Leu Pro Gly
 35 40 45

Arg Pro Gly Gln His Pro Ala Asn Pro Leu Ser Tyr Leu Asp Ala Pro
 50 55 60

Glu Arg Leu Asn Pro Gln Phe Leu Glu Asp phe Lys Leu val His His
 65 70 75 80

Asn Glu Pro Ser Leu Met Ser Thr Gly val Leu Leu Cys Gly val Glu
 85 90 95

Arg Arg Gly Leu val His Ala Leu Arg Asp Lys Cys Arg Ser Gln Gly
 100 105 110

Ile Ala ile Arg phe Glu Ser Pro Leu Leu Glu His Gly Glu Leu Pro
 115 120 125

Leu Ala Asp Tyr Asp Leu val val Leu Ala Asn Gly val Asn His Lys
 130 135 140

Thr Ala His phe Thr Glu Ala Leu val Pro Gln val Asp Tyr Gly Arg
 145 150 155 160

Asn Lys Tyr ile Trp Tyr Gly Thr Ser Gln Leu phe Asp Gln Met Asn
 165 170 175

Leu val phe Arg Thr His Gly Lys Asp ile phe ile Ala His Ala Tyr
 180 185 190

Lys Tyr Ser Asp Thr Met Ser Thr Phe Ile Val³⁴ Glu Cys Ser Glu Glu
195 200 205

Thr Tyr Ala Arg Ala Arg Leu Gly Glu Met Ser Glu Glu Ala Ser Ala
210 215 220

Glu Tyr Val Ala Lys Val Phe Gln Ala Glu Leu Gly Gly His Gly Leu
225 230 235 240

Val Ser Gln Pro Gly Leu Gly Trp Arg Asn Phe Met Thr Leu Ser His
245 250 255

Asp Arg Cys His Asp Gly Lys Leu Val Leu Leu Gly Asp Ala Leu Gln
260 265 270

Ser Gly His Phe Ser Ile Gly His Gly Thr Thr Met Ala Val Val Val
275 280 285

Ala Gln Leu Leu Val Lys Ala Leu Cys Thr Glu Asp Gly Val Pro Ala
290 295 300

Ala Leu Lys Arg Phe Glu Glu Arg Ala Leu Pro Leu Val Gln Leu Phe
305 310 315 320

Arg Gly His Ala Asp Asn Ser Arg Val Trp Phe Glu Thr Val Glu Glu
325 330 335

Arg Met His Leu Ser Ser Ala Glu Phe Val Gln Ser Phe Asp Ala Arg
340 345 350

Arg Lys Ser Leu Pro Pro Met Pro Glu Ala Leu Ala Gln Asn Leu Arg
355 360 365

Tyr Ala Leu Gln Arg
370

<210> 43
<211> 576
<212> DNA
<213> künstliche Sequenz

<220>
<223> VioE

<220>
<221> CDS
<222> (1)..(576)
<223> VioE

<400> 43
atg gag aac cgt gag cca cca ctg ttg cca gcc cgt tgg agc agc gcc 48
Met Glu Asn Arg Glu Pro Pro Leu Leu Pro Ala Arg Trp Ser Ser Ala
1 5 10 15

tat gtc tct tat tgg agc ccg atg ctg ccg gat gac cag ctg acc agc 96
Tyr Val Ser Tyr Trp Ser Pro Met Leu Pro Asp Asp Gln Leu Thr Ser

20										25					30					
ggc	tat	tgc	tgg	ttc	gac	tat	gaa	cgt	gac	atc	tgt	cgt	att	gac	ggc	144				
Gly	Tyr	Cys	Trp	Phe	Asp	Tyr	Glu	Arg	Asp	Ile	Cys	Arg	Ile	Asp	Gly					
		35					40					45								
ctg	ttc	aat	ccg	tgg	agc	gag	cgt	gat	act	ggg	tat	cgc	ctg	tgg	atg	192				
Leu	Phe	Asn	Pro	Trp	Ser	Glu	Arg	Asp	Thr	Gly	Tyr	Arg	Leu	Trp	Met					
	50					55				60										
tgc	gag	gtt	ggg	aat	gcg	gcc	agc	ggc	cgt	acc	tgg	aaa	caa	aaa	gtc	240				
Ser	Glu	Val	Gly	Asn	Ala	Ala	Ser	Gly	Arg	Thr	Trp	Lys	Gln	Lys	Val					
	65				70					75					80					
gcc	tat	ggg	cgt	gag	cgt	acc	gcc	ctg	ggg	gaa	cag	ctg	tgt	gag	cgt	288				
Ala	Tyr	Gly	Arg	Glu	Arg	Thr	Ala	Leu	Gly	Glu	Gln	Leu	Cys	Glu	Arg					
				85					90					95						
ccg	ctg	gat	gat	gag	act	ggc	cct	ttt	gcc	gaa	ttg	ttc	ctg	cca	cgc	336				
Pro	Leu	Asp	Asp	Glu	Thr	Gly	Pro	Phe	Ala	Glu	Leu	Phe	Leu	Pro	Arg					
			100					105					110							
gat	gtc	ctg	cgc	cgt	ctg	ggg	gcc	cgt	cac	att	ggc	cgt	cgc	gtg	gtt	384				
Asp	Val	Leu	Arg	Arg	Leu	Gly	Ala	Arg	His	Ile	Gly	Arg	Arg	Val	Val					
		115				120						125								
ctg	ggg	cgc	gaa	gcg	gac	ggg	tgg	cgt	tac	cag	cgc	cca	ggg	aaa	ggg	432				
Leu	Gly	Arg	Glu	Ala	Asp	Gly	Trp	Arg	Tyr	Gln	Arg	Pro	Gly	Lys	Gly					
	130					135				140										
ccg	agc	acc	ctg	tac	ctg	gat	gcg	gcg	agc	ggc	act	cca	ctg	cgc	atg	480				
Pro	Ser	Thr	Leu	Tyr	Leu	Asp	Ala	Ala	Ser	Gly	Thr	Pro	Leu	Arg	Met					
	145				150					155					160					
gtc	acc	ggc	gat	gaa	gcg	tgc	cgt	gca	agc	ctg	cgt	gat	ttt	ccg	aat	528				
Val	Thr	Gly	Asp	Glu	Ala	Ser	Arg	Ala	Ser	Leu	Arg	Asp	Phe	Pro	Asn					
				165					170					175						
gtg	agc	gag	gcg	gag	atc	ccg	gac	gcg	gtt	ttc	gcg	gcc	aag	cgc	taa	576				
Val	Ser	Glu	Ala	Glu	Ile	Pro	Asp	Ala	Val	Phe	Ala	Ala	Lys	Arg						
			180					185					190							

<210> 44
 <211> 191
 <212> PRT
 <213> künstliche Sequenz

<220>
 <223> synthetisches Konstrukt

<400> 44

Met Glu Asn Arg Glu Pro Pro Leu Leu Pro Ala Arg Trp Ser Ser Ala
 1 5 10 15

Tyr Val Ser Tyr Trp Ser Pro Met Leu Pro Asp Asp Gln Leu Thr Ser
 20 25 30

Gly Tyr Cys Trp Phe Asp Tyr Glu Arg Asp Ile Cys Arg Ile Asp Gly
 35 40 45

Leu Phe Asn Pro Trp Ser Glu Arg Asp Thr Gly Tyr Arg Leu Trp Met
 50 55 60

Ser Glu Val Gly Asn Ala Ala Ser Gly Arg³⁶ Thr Trp Lys Gln Lys Val
65 70 75 80

Ala Tyr Gly Arg Glu Arg Thr Ala Leu Gly Glu Gln Leu Cys Glu Arg
85 90 95

Pro Leu Asp Asp Glu Thr Gly Pro Phe Ala Glu Leu Phe Leu Pro Arg
100 105 110

Asp Val Leu Arg Arg Leu Gly Ala Arg His Ile Gly Arg Arg Val Val
115 120 125

Leu Gly Arg Glu Ala Asp Gly Trp Arg Tyr Gln Arg Pro Gly Lys Gly
130 135 140

Pro Ser Thr Leu Tyr Leu Asp Ala Ala Ser Gly Thr Pro Leu Arg Met
145 150 155 160

Val Thr Gly Asp Glu Ala Ser Arg Ala Ser Leu Arg Asp Phe Pro Asn
165 170 175

Val Ser Glu Ala Glu Ile Pro Asp Ala Val Phe Ala Ala Lys Arg
180 185 190

<210> 45
<211> 516
<212> DNA
<213> künstliche Sequenz

<220>
<223> nCFP (N-terminal mCerulean)

<220>
<221> CDS
<222> (1)..(516)
<223> nCFP (N-terminal mCerulean)

<400> 45
gtg agc aag ggc gag gag ctg ttc acc ggg gtg gtg ccc atc ctg gtc 48
Val Ser Lys Gly Glu Glu Leu Phe Thr Gly Val Val Pro Ile Leu Val
1 5 10 15

gag ctg gac ggc gac gta aac ggc cac aag ttc agc gtg tcc ggc gag 96
Glu Leu Asp Gly Asp Val Asn Gly His Lys Phe Ser Val Ser Gly Glu
20 25 30

ggc gag ggc gat gcc acc tac ggc aag ctg acc ctg aag ttc atc tgc 144
Gly Glu Gly Asp Ala Thr Tyr Gly Lys Leu Thr Leu Lys Phe Ile Cys
35 40 45

acc acc ggc aag ctg ccc gtg ccc tgg ccc acc ctc gtg acc acc ctg 192
Thr Thr Gly Lys Leu Pro Val Pro Trp Pro Thr Leu Val Thr Thr Leu
50 55 60

acc tgg ggc gtg cag tgc ttc gcc cgc tac ccc gac cac atg aag cag 240
Thr Trp Gly Val Gln Cys Phe Ala Arg Tyr Pro Asp His Met Lys Gln
65 70 75 80

cac gac ttc ttc aag tcc gcc atg ccc gaa ggc tac gtc cag gag cgc 288
His Asp Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr Val Gln Glu Arg

85	90	37	95	
acc atc ttc ttc aag gac gac ggc aac tac aag acc cgc gcc gag gtg				336
Thr Ile Phe Phe Lys Asp Asp Gly Asn Tyr Lys Thr Arg Ala Glu Val				
			110	
aag ttc gag ggc gac acc ctg gtg aac cgc atc gag ctg aag ggc atc				384
Lys Phe Glu Gly Asp Thr Leu Val Asn Arg Ile Glu Leu Lys Gly Ile				
			125	
gac ttc aag gag gac ggc aac atc ctg ggg cac aag ctg gag tac aac				432
Asp Phe Lys Glu Asp Gly Asn Ile Leu Gly His Lys Leu Glu Tyr Asn				
			140	
gcc atc agc gac aac gtc tat atc acc gcc gac aag cag aag aac ggc				480
Ala Ile Ser Asp Asn Val Tyr Ile Thr Ala Asp Lys Gln Lys Asn Gly				
			160	
atc aag gcc aac ttc aag atc cgc cac aac atc taa				516
Ile Lys Ala Asn Phe Lys Ile Arg His Asn Ile				
			170	

<210> 46
 <211> 171
 <212> PRT
 <213> künstliche Sequenz

<220>
 <223> synthetisches Konstrukt

<400> 46

Val Ser Lys Gly Glu Glu Leu Phe Thr Gly Val Val Pro Ile Leu Val
 1 5 10 15

Glu Leu Asp Gly Asp Val Asn Gly His Lys Phe Ser Val Ser Gly Glu
 20 25 30

Gly Glu Gly Asp Ala Thr Tyr Gly Lys Leu Thr Leu Lys Phe Ile Cys
 35 40 45

Thr Thr Gly Lys Leu Pro Val Pro Trp Pro Thr Leu Val Thr Thr Leu
 50 55 60

Thr Trp Gly Val Gln Cys Phe Ala Arg Tyr Pro Asp His Met Lys Gln
 65 70 75 80

His Asp Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr Val Gln Glu Arg
 85 90 95

Thr Ile Phe Phe Lys Asp Asp Gly Asn Tyr Lys Thr Arg Ala Glu Val
 100 105 110

Lys Phe Glu Gly Asp Thr Leu Val Asn Arg Ile Glu Leu Lys Gly Ile
 115 120 125

Asp Phe Lys Glu Asp Gly Asn Ile Leu Gly His Lys Leu Glu Tyr Asn
 130 135 140

Ala Ile Ser Asp Asn Val Tyr Ile Thr Ala Asp Lys Gln Lys Asn Gly
145 150 155 160

Ile Lys Ala Asn Phe Lys Ile Arg His Asn Ile
165 170

<210> 47
<211> 516
<212> DNA
<213> künstliche Sequenz

<220>
<223> nYFP (N-terminal mCitrine)

<220>
<221> CDS
<222> (1)..(516)
<223> nYFP (N-terminal mCitrine)

<400> 47
gtg agc aag ggc gag gag ctg ttc acc ggg gtg gtg ccc atc ctg gtc 48
Val Ser Lys Gly Glu Glu Leu Phe Thr Gly Val Val Pro Ile Leu Val
1 5 10 15
gag ctg gac ggc gac gta aac ggc cac aag ttc agc gtg tcc ggc gag 96
Glu Leu Asp Gly Asp Val Asn Gly His Lys Phe Ser Val Ser Gly Glu
20 25 30
ggc gag ggc gat gcc acc tac ggc aag ctg acc ctg aag ttc atc tgc 144
Gly Glu Gly Asp Ala Thr Tyr Gly Lys Leu Thr Leu Lys Phe Ile Cys
35 40 45
acc acc ggc aag ctg ccc gtg ccc tgg ccc acc ctc gtg acc acc ttc 192
Thr Thr Gly Lys Leu Pro Val Pro Trp Pro Thr Leu Val Thr Thr Phe
50 55 60
ggc tac ggc ctg atg tgc ttc gcc cgc tac ccc gac cac atg aag cag 240
Gly Tyr Gly Leu Met Cys Phe Ala Arg Tyr Pro Asp His Met Lys Gln
65 70 75 80
cac gac ttc ttc aag tcc gcc atg ccc gaa ggc tac gtc cag gag cgc 288
His Asp Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr Val Gln Glu Arg
85 90 95
acc atc ttc ttc aag gac gac ggc aac tac aag acc cgc gcc gag gtg 336
Thr Ile Phe Phe Lys Asp Asp Gly Asn Tyr Lys Thr Arg Ala Glu Val
100 105 110
aag ttc gag ggc gac acc ctg gtg aac cgc atc gag ctg aag ggc atc 384
Lys Phe Glu Gly Asp Thr Leu Val Asn Arg Ile Glu Leu Lys Gly Ile
115 120 125
gac ttc aag gag gac ggc aac atc ctg ggg cac aag ctg gag tac aac 432
Asp Phe Lys Glu Asp Gly Asn Ile Leu Gly His Lys Leu Glu Tyr Asn
130 135 140
tac aac agc cac aac gtc tat atc atg gcc gac aag cag aag aac ggc 480
Tyr Asn Ser His Asn Val Tyr Ile Met Ala Asp Lys Gln Lys Asn Gly
145 150 155 160
atc aag gtg aac ttc aag atc cgc cac aac atc taa 516
Ile Lys Val Asn Phe Lys Ile Arg His Asn Ile
165 170

<210> 48

39

<211> 171
<212> PRT
<213> künstliche Sequenz

<220>
<223> synthetisches Konstrukt

<400> 48

Val Ser Lys Gly Glu Glu Leu Phe Thr Gly Val Val Pro Ile Leu Val
1 5 10 15

Glu Leu Asp Gly Asp Val Asn Gly His Lys Phe Ser Val Ser Gly Glu
20 25 30

Gly Glu Gly Asp Ala Thr Tyr Gly Lys Leu Thr Leu Lys Phe Ile Cys
35 40 45

Thr Thr Gly Lys Leu Pro Val Pro Trp Pro Thr Leu Val Thr Thr Phe
50 55 60

Gly Tyr Gly Leu Met Cys Phe Ala Arg Tyr Pro Asp His Met Lys Gln
65 70 75 80

His Asp Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr Val Gln Glu Arg
85 90 95

Thr Ile Phe Phe Lys Asp Asp Gly Asn Tyr Lys Thr Arg Ala Glu Val
100 105 110

Lys Phe Glu Gly Asp Thr Leu Val Asn Arg Ile Glu Leu Lys Gly Ile
115 120 125

Asp Phe Lys Glu Asp Gly Asn Ile Leu Gly His Lys Leu Glu Tyr Asn
130 135 140

Tyr Asn Ser His Asn Val Tyr Ile Met Ala Asp Lys Gln Lys Asn Gly
145 150 155 160

Ile Lys Val Asn Phe Lys Ile Arg His Asn Ile
165 170

<210> 49
<211> 288
<212> DNA
<213> künstliche Sequenz

<220>
<223> cCFP (C-terminal mCerulean)

<220>
<221> CDS
<222> (1)..(288)
<223> cCFP (C-terminal mCerulean)

<400> 49

atg gcc gac aag cag aag aac ggc atc aag gcc aac ttc aag atc cgc

48

Met Ala Asp Lys Gln Lys Asn Gly Ile Lys Ala Asn Phe Lys Ile Arg
1 5 10 15
cac aac atc gag gac ggc agc gtg cag ctc gcc gac cac tac cag cag 96
His Asn Ile Glu Asp Gly Ser Val Gln Leu Ala Asp His Tyr Gln Gln
20 25 30
aac acc ccc atc ggc gac ggc ccc gtg ctg ctg ccc gac aac cac tac 144
Asn Thr Pro Ile Gly Asp Gly Pro Val Leu Leu Pro Asp Asn His Tyr
35 40 45
ctg agc acc cag tcc aag ctg agc aaa gac ccc aac gag aag cgc gat 192
Leu Ser Thr Gln Ser Lys Ser Lys Asp Pro Asn Glu Lys Arg Asp
50 55 60
cac atg gtc ctg ctg gag ttc gtg acc gcc gcc ggg atc act ctc ggc 240
His Met Val Leu Leu Glu Phe Val Thr Ala Ala Gly Ile Thr Leu Gly
65 70 75 80
atg gac gag ctg tac aag ggt ggc tct ggc ggt ggg agt ggg gga agc 288
Met Asp Glu Leu Tyr Lys Gly Gly Ser Gly Gly Gly Ser Gly Gly Ser
85 90 95

<210> 50
<211> 96
<212> PRT
<213> künstliche Sequenz

<220>
<223> synthetisches Konstrukt

<400> 50

Met Ala Asp Lys Gln Lys Asn Gly Ile Lys Ala Asn Phe Lys Ile Arg
1 5 10 15
His Asn Ile Glu Asp Gly Ser Val Gln Leu Ala Asp His Tyr Gln Gln
20 25 30
Asn Thr Pro Ile Gly Asp Gly Pro Val Leu Leu Pro Asp Asn His Tyr
35 40 45
Leu Ser Thr Gln Ser Lys Leu Ser Lys Asp Pro Asn Glu Lys Arg Asp
50 55 60
His Met Val Leu Leu Glu Phe Val Thr Ala Ala Gly Ile Thr Leu Gly
65 70 75 80
Met Asp Glu Leu Tyr Lys Gly Gly Ser Gly Gly Gly Ser Gly Gly Ser
85 90 95

<210> 51
<211> 288
<212> DNA
<213> künstliche Sequenz

<220>
<223> cYFP (C-terminal mCitrine)

<220>
<221> CDS

41

<222> (1)..(288)

<223> cYFP (C-terminal mCitrine)

<400> 51

atg gcc gac aag cag aag aac ggc atc aag gtg aac ttc aag atc cgc 48
Met Ala Asp Lys Gln Lys Asn Gly Ile Lys Val Asn Phe Lys Ile Arg
1 5 10 15

cac aac atc gag gac ggc agc gtg cag ctc gcc gac cac tac cag cag 96
His Asn Ile Glu Asp Gly Ser Val Gln Leu Ala Asp His Tyr Gln Gln
20 25 30

aac acc ccc atc ggc gac ggc ccc gtg ctg ctg ccc gac aac cac tac 144
Asn Thr Pro Ile Gly Asp Gly Pro Val Leu Leu Pro Asp Asn His Tyr
35 40 45

ctg agc tac cag tcc gcc ctg agc aaa gac ccc aac gag aag cgc gat 192
Leu Ser Tyr Gln Ser Ala Leu Ser Lys Asp Pro Asn Glu Lys Arg Asp
50 55 60

cac atg gtc ctg ctg gag ttc gtg acc gcc gcc ggg atc act ctc ggc 240
His Met Val Leu Leu Glu Phe Val Thr Ala Ala Gly Ile Thr Leu Gly
65 70 75 80

atg gac gag ctg tac aag ggt ggc tct ggc ggt ggg agt ggg gga agc 288
Met Asp Glu Leu Tyr Lys Gly Gly Ser Gly Gly Gly Ser Gly Gly Ser
85 90 95

<210> 52

<211> 96

<212> PRT

<213> künstliche Sequenz

<220>

<223> synthetisches Konstrukt

<400> 52

Met Ala Asp Lys Gln Lys Asn Gly Ile Lys Val Asn Phe Lys Ile Arg
1 5 10 15

His Asn Ile Glu Asp Gly Ser Val Gln Leu Ala Asp His Tyr Gln Gln
20 25 30

Asn Thr Pro Ile Gly Asp Gly Pro Val Leu Leu Pro Asp Asn His Tyr
35 40 45

Leu Ser Tyr Gln Ser Ala Leu Ser Lys Asp Pro Asn Glu Lys Arg Asp
50 55 60

His Met Val Leu Leu Glu Phe Val Thr Ala Ala Gly Ile Thr Leu Gly
65 70 75 80

Met Asp Glu Leu Tyr Lys Gly Gly Ser Gly Gly Gly Ser Gly Gly Ser
85 90 95

<210> 53

<211> 30

<212> DNA

<213> künstliche Sequenz

<220>

42

<223> Linkerpeptid

<220>

<221> CDS

<222> (1)..(30)

<223> Linkerpeptid

<400> 53

ggt ggc tct ggc ggt ggg agt ggg gga agc
Gly Gly Ser Gly Gly Gly Ser Gly Gly Ser
1 5 10

30

<210> 54

<211> 10

<212> PRT

<213> künstliche Sequenz

<220>

<223> synthetisches Konstrukt

<400> 54

Gly Gly Ser Gly Gly Gly Ser Gly Gly Ser
1 5 10

<210> 55

<211> 18

<212> DNA

<213> künstliche Sequenz

<220>

<223> His-tag

<400> 55

caccaccacc accaccac

18

<210> 56

<211> 112

<212> DNA

<213> künstliche Sequenz

<220>

<223> Programm SPR, split, FRET

<400> 56

gttactactc gagcggaatt cgaaggggaa ttgctgctgc ggtgtttgga tggagcgtgg

60

gcgggggtgtg gaaattgatg ctgcattgac caccacaagac gactgcagta ca

112

<210> 57

<211> 35

<212> DNA

<213> künstliche Sequenz

<220>

<223> Programm NICTAL

<400> 57

gttactactc gagcgttcta tcaatgatag ataca

35

<210> 58

<211> 90

<212> DNA

43

<213> künstliche Sequenz

<220>

<223> Programm Biosynthese (123456)

<400> 58

aattcgcggc cgctctagag ctgctgcggg gtttgatgg agcgtgggcg ggggtgtggaa 60

attgatgctg cattgaccac ccaagacgaa 90

<210> 59

<211> 79

<212> DNA

<213> künstliche Sequenz

<220>

<223> Programm Biosynthese (12346)

<400> 59

aattcgcggc cgctctagag ctgctgcggg gtttgatgg agcgtgggcg ggggtgtggaa 60

attgaccacc caagacgaa 79

<210> 60

<211> 90

<212> DNA

<213> künstliche Sequenz

<220>

<223> Programm Biosynthese (341256)

<400> 60

aattcgcggc cgctctagag cgtgggcggg gtgtggaaat tgctgctgcg gtgtttggat 60

ggagatgctg cattgaccac ccaagacgaa 90

<210> 61

<211> 1602

<212> DNA

<213> künstliche Sequenz

<220>

<223> Znf_Blues-Linkerpeptid-vioA

<220>

<221> CDS

<222> (1)..(1602)

<223> Znf_Blues-Linkerpeptid-vioA

<400> 61

aag agg aga aat act aga atg act aga gca agt gat gat cgt ccg tat 48
Lys Arg Arg Asn Thr Arg Met Thr Arg Ala Ser Asp Asp Arg Pro Tyr
1 5 10 15

gca tgt ccg gtt gaa agc tgt gat cgt cgt ttt agc cgt cgt gat gtt 96
Ala Cys Pro Val Glu Ser Cys Asp Arg Arg Phe Ser Arg Arg Asp Val
20 25 30

ctg atg aat cac att cgt att cac aca ggt cag aaa ccg ttt cag tgt 144
Leu Met Asn His Ile Arg Ile His Thr Gly Gln Lys Pro Phe Gln Cys
35 40 45

cgt att tgc atg cgt aat ttt agc cgt agc gat cat ctg acc acc cat 192
Arg Ile Cys Met Arg Asn Phe Ser Arg Ser Asp His Leu Thr Thr His
50 55 60

44

att Ile 65	cgt Arg	acc Thr	cat His	aca Thr	ggt Gly 70	gaa Glu	aaa Lys	ccg Pro	ttt Phe	gca Ala 75	tgt Cys	gat Asp	att Ile	tgc Cys	ggt Gly 80	240
cgc Arg	aaa Lys	ttt Phe	gca Ala	aat Asn 85	cgt Arg	gat Asp	acc Thr	ctg Leu	acc Thr 90	cgt Arg	cat His	agc Ser	aaa Lys	att Ile 95	cat His	288
ctg Leu	cgt Arg	cag Gln	aat Asn 100	gat Asp	ctg Leu	gaa Glu	ggt Gly	ggt Gly 105	agc Ser	ggt Gly	ggt Gly	ggt Gly	tca Ser 110	ggt Gly	ggt Gly	336
agt Ser	act Thr	aga Arg 115	atg Met	aaa Lys	cat His	tct Ser	tcc Ser 120	gat Asp	atc Ile	tgc Cys	att Ile	gtt Val 125	ggt Gly	gct Ala	ggt Gly	384
att Ile 130	tct Ser	ggt Gly	ttg Leu	acg Thr	tgc Cys	gca Ala 135	agc Ser	cat His	ctg Leu	ctg Leu	gac Asp 140	agc Ser	ccg Pro	gca Ala	tgc Cys	432
cgt Arg 145	ggt Gly	ctg Leu	agc Ser	ctg Leu	cgt Arg 150	atc Ile	ttt Phe	gac Asp	atg Met	cag Gln 155	caa Gln	gaa Glu	gcc Ala	ggt Gly	ggc Gly 160	480
cgt Arg	atc Ile	cgc Arg	agc Ser	aaa Lys 165	atg Met	ctg Leu	gat Asp	ggt Gly 170	aag Lys	gca Ala	agc Ser	att Ile	gaa Glu	ctg Leu 175	ggc Gly	528
gca Ala	ggt Gly	cgc Arg	tac Tyr 180	tcc Ser	cct Pro	cag Gln	ttg Leu	cac His 185	ccg Pro	cat His	ttc Phe	caa Gln	agc Ser 190	gca Ala	atg Met	576
cag Gln	cac His	tat Tyr 195	agc Ser	caa Gln	aag Lys	agc Ser	gaa Glu	gtc Val	tat Tyr	ccg Pro	ttc Phe	acc Thr 205	cag Gln	ttg Leu	aag Lys	624
ttc Phe 210	aaa Lys	tct Ser	cac His	gtg Val	cag Gln	caa Gln 215	aag Lys	ctg Leu	aag Lys	cgc Arg	gcc Ala 220	atg Met	aat Asn	gaa Glu	ctg Leu	672
tcc Ser 225	ccg Pro	cgt Arg	ctg Leu	aaa Lys	gag Glu 230	cat His	ggt Gly	aaa Lys	gag Glu	agc Ser 235	ttt Phe	ttg Leu	cag Gln	ttt Phe	gtc Val 240	720
agc Ser	cgt Arg	tat Tyr	caa Gln	ggt Gly 245	cac His	gat Asp	agc Ser	gcg Ala	gtt Val 250	ggt Gly	atg Met	atc Ile	cgc Arg	tct Ser 255	atg Met	768
ggt Gly	tac Tyr	gac Asp	gca Ala 260	ctg Leu	ttc Phe	ctg Leu	ccg Pro	gat Asp 265	atc Ile	agc Ser	gca Ala	gaa Glu	atg Met 270	gcc Ala	tac Tyr	816
gac Asp	att Ile	gtg Val 275	ggt Gly	aag Lys	cac His	ccg Pro	gag Glu 280	atc Ile	cag Gln	agc Ser	gtg Val	acg Thr 285	gac Asp	aac Asn	gac Asp	864
gcg Ala 290	aac Asn	caa Gln	tgg Trp	ttt Phe	gca Ala	gca Ala 295	gaa Glu	acg Thr	ggc Gly	ttt Phe	gct Ala 300	ggt Gly	ctg Leu	att Ile	cag Gln	912
ggc Gly 305	atc Ile	aag Lys	gct Ala	aag Lys	gtt Val 310	aag Lys	gca Ala	ggt Gly	gca Ala 315	cgt Arg	ttt Phe	agc Ser	ctg Leu	ggt Gly 320		960
tat Tyr	cgt Arg	ctg Leu	ctg Leu	agc Ser 325	gtc Val	cgt Arg	acc Thr	gac Asp	ggt Gly 330	gac Asp	ggc Gly	tac Tyr	ctg Leu	ctg Leu 335	caa Gln	1008

45

ctg gca ggt gac gac ggc tgg aaa ctg gag cac cgt acc cgc cat ctg	1056
Leu Ala Gly Asp Asp Gly Trp Lys Leu Glu His Arg Thr Arg His Leu	
340	345
att ctg gcg att ccg ccg agc gcg atg gcg ggt ttg aat gtt gat ttt	1104
Ile Leu Ala Ile Pro Pro Ser Ala Met Ala Gly Leu Asn Val Asp Phe	
355	360
cca gaa gcc tgg tcc ggt gcg cgc tat ggc agc ctg ccg ctg ttt aag	1152
Pro Glu Ala Trp Ser Gly Ala Arg Tyr Gly Ser Leu Pro Leu Phe Lys	
370	375
ggc ttt ctg acg tac ggt gag ccg tgg tgg ttg gac tac aaa ctg gac	1200
Gly Phe Leu Thr Tyr Gly Glu Pro Trp Trp Leu Asp Tyr Lys Leu Asp	
385	390
gat cag gtg ctg att gtt gac aac ccg ctg cgc aaa atc tat ttc aaa	1248
Asp Gln Val Leu Ile Val Asp Asn Pro Leu Arg Lys Ile Tyr Phe Lys	
405	410
ggc gat aag tac ctg ttc ttc tat acc gat agc gag atg gcg aat tac	1296
Gly Asp Lys Tyr Leu Phe Phe Tyr Thr Asp Ser Glu Met Ala Asn Tyr	
420	425
tggt cgc ggt tgt gtc gcg gag ggc gag gac ggt tac ctg gag caa att	1344
Trp Arg Gly Cys Val Ala Glu Gly Glu Asp Gly Tyr Leu Glu Gln Ile	
435	440
cgc acc cat ttg gct agc gca ctg ggt atc gtc cgt gaa cgt atc ccg	1392
Arg Thr His Leu Ala Ser Ala Leu Gly Ile Val Arg Glu Arg Ile Pro	
450	455
caa ccg ctg gca cac gtt cac aag tat tgg gcg cac ggc gtt gag ttt	1440
Gln Pro Leu Ala His Val His Lys Tyr Trp Ala His Gly Val Glu Phe	
465	470
tgc cgt gat tct gat att gac cac ccg agc gca ctg tct cat cgc gac	1488
Cys Arg Asp Ser Asp Ile Asp His Pro Ser Ala Leu Ser His Arg Asp	
485	490
agc ggt atc atc gcg tgc tcc gat gcg tac acg gag cat tgt ggt tgg	1536
Ser Gly Ile Ile Ala Cys Ser Asp Ala Tyr Thr Glu His Cys Gly Trp	
500	505
atg gag ggc ggt ctg ctg agc gcc cgt gag gca agc cgt ctg ctg ttg	1584
Met Glu Gly Gly Leu Leu Ser Ala Arg Glu Ala Ser Arg Leu Leu Leu	
515	520
cag cgt atc gcc gcg tga	1602
Gln Arg Ile Ala Ala	
530	

<210> 62
 <211> 533
 <212> PRT
 <213> künstliche Sequenz

<220>
 <223> synthetisches Konstrukt

<400> 62

Lys Arg Arg Asn Thr Arg Met Thr Arg Ala Ser Asp Asp Arg Pro Tyr
 1 5 10 15

Ala Cys Pro Val Glu Ser Cys Asp Arg Arg Phe Ser Arg Arg Asp Val
 20 25 30

46

Leu Met Asn His Ile Arg Ile His Thr Gly Gln Lys Pro Phe Gln Cys
35 40 45

Arg Ile Cys Met Arg Asn Phe Ser Arg Ser Asp His Leu Thr Thr His
50 55 60

Ile Arg Thr His Thr Gly Glu Lys Pro Phe Ala Cys Asp Ile Cys Gly
65 70 75 80

Arg Lys Phe Ala Asn Arg Asp Thr Leu Thr Arg His Ser Lys Ile His
85 90 95

Leu Arg Gln Asn Asp Leu Glu Gly Gly Ser Gly Gly Gly Ser Gly Gly
100 105 110

Ser Thr Arg Met Lys His Ser Ser Asp Ile Cys Ile Val Gly Ala Gly
115 120 125

Ile Ser Gly Leu Thr Cys Ala Ser His Leu Leu Asp Ser Pro Ala Cys
130 135 140

Arg Gly Leu Ser Leu Arg Ile Phe Asp Met Gln Gln Glu Ala Gly Gly
145 150 155 160

Arg Ile Arg Ser Lys Met Leu Asp Gly Lys Ala Ser Ile Glu Leu Gly
165 170 175

Ala Gly Arg Tyr Ser Pro Gln Leu His Pro His Phe Gln Ser Ala Met
180 185 190

Gln His Tyr Ser Gln Lys Ser Glu Val Tyr Pro Phe Thr Gln Leu Lys
195 200 205

Phe Lys Ser His Val Gln Gln Lys Leu Lys Arg Ala Met Asn Glu Leu
210 215 220

Ser Pro Arg Leu Lys Glu His Gly Lys Glu Ser Phe Leu Gln Phe Val
225 230 235 240

Ser Arg Tyr Gln Gly His Asp Ser Ala Val Gly Met Ile Arg Ser Met
245 250 255

Gly Tyr Asp Ala Leu Phe Leu Pro Asp Ile Ser Ala Glu Met Ala Tyr
260 265 270

Asp Ile Val Gly Lys His Pro Glu Ile Gln Ser Val Thr Asp Asn Asp
275 280 285

Ala Asn Gln Trp Phe Ala Ala Glu Thr Gly Phe Ala Gly Leu Ile Gln
290 295 300

47

Gly Ile Lys Ala Lys Val Lys Ala Ala Gly Ala Arg Phe Ser Leu Gly
305 310 315 320

Tyr Arg Leu Leu Ser Val Arg Thr Asp Gly Asp Gly Tyr Leu Leu Gln
325 330 335

Leu Ala Gly Asp Asp Gly Trp Lys Leu Glu His Arg Thr Arg His Leu
340 345 350

Ile Leu Ala Ile Pro Pro Ser Ala Met Ala Gly Leu Asn Val Asp Phe
355 360 365

Pro Glu Ala Trp Ser Gly Ala Arg Tyr Gly Ser Leu Pro Leu Phe Lys
370 375 380

Gly Phe Leu Thr Tyr Gly Glu Pro Trp Trp Leu Asp Tyr Lys Leu Asp
385 390 395 400

Asp Gln Val Leu Ile Val Asp Asn Pro Leu Arg Lys Ile Tyr Phe Lys
405 410 415

Gly Asp Lys Tyr Leu Phe Phe Tyr Thr Asp Ser Glu Met Ala Asn Tyr
420 425 430

Trp Arg Gly Cys Val Ala Glu Gly Glu Asp Gly Tyr Leu Glu Gln Ile
435 440 445

Arg Thr His Leu Ala Ser Ala Leu Gly Ile Val Arg Glu Arg Ile Pro
450 455 460

Gln Pro Leu Ala His Val His Lys Tyr Trp Ala His Gly Val Glu Phe
465 470 475 480

Cys Arg Asp Ser Asp Ile Asp His Pro Ser Ala Leu Ser His Arg Asp
485 490 495

ser Gly Ile Ile Ala Cys Ser Asp Ala Tyr Thr Glu His Cys Gly Trp
500 505 510

Met Glu Gly Gly Leu Leu Ser Ala Arg Glu Ala Ser Arg Leu Leu Leu
515 520 525

Gln Arg Ile Ala Ala
530

<210> 63
<211> 3300
<212> DNA
<213> künstliche Sequenz

<220>
<223> znf_Zif268-Linkerpeptid-vioB

48

<220>
<221> CDS
<222> (1)..(3300)
<223> Znf_Zif268-Linkerpeptid-vioB

<400> 63
atg ccg ggt gaa aaa ccg tat gca tgt ccg gtt gaa agc tgt gat cgt 48
Met Pro Gly Glu Lys Pro Tyr Ala Cys Pro Val Glu Ser Cys Asp Arg
1 5 10 15
cgt ttt agc cgt agt gat gaa ctg acc cgt cat att cgt att cat aca 96
Arg Phe Ser Arg Ser Asp Glu Leu Thr Arg His Ile Arg Ile His Thr
20 25 30
ggg cag aaa ccg ttt cag tgt cgt att tgc atg cgt aat ttt agc cgt 144
Gly Gln Lys Pro Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser Arg
35 40 45
tca gat cat ctg acc acc cat att cgt acc cat act ggc gaa aaa ccg 192
Ser Asp His Leu Thr Thr His Ile Arg Thr His Thr Gly Glu Lys Pro
50 55 60
ttt gcc tgt gat att tgt ggt cgt aaa ttt gca cgt tcc gat gaa cgt 240
Phe Ala Cys Asp Ile Cys Gly Arg Lys Phe Ala Arg Ser Asp Glu Arg
65 70 75 80
aaa cgc cat acc aaa att cat acg ggt ggt ggt agc ggt ggt ggt tca 288
Lys Arg His Thr Lys Ile His Thr Gly Gly Gly Ser Gly Gly Gly Ser
85 90 95
ggg ggt agt act aga atg agc att ctg gat ttc ccg cgt atc cac ttc 336
Gly Gly Ser Thr Arg Met Ser Ile Leu Asp Phe Pro Arg Ile His Phe
100 105 110
cgt gcc tgg gcc cgt gtc aat gcg ccg acc gcg aac cgc gat ccg cac 384
Arg Gly Trp Ala Arg Val Asn Ala Pro Thr Ala Asn Arg Asp Pro His
115 120 125
ggc cac atc gat atg gcc agc aat acc gtg gcg atg gcg ggt gag ccg 432
Gly His Ile Asp Met Ala Ser Asn Thr Val Ala Met Ala Gly Glu Pro
130 135 140
ttc gac ctg gca cgc cat cct acg gag ttc cac cgt cac ctg cgc tcc 480
Phe Asp Leu Ala Arg His Pro Thr Glu Phe His Arg His Leu Arg Ser
145 150 155 160
ctg ggt ccg cgc ttc ggc ttg gat ggt cgt gct gac ccg gaa ggc ccg 528
Leu Gly Pro Arg Phe Gly Leu Asp Gly Arg Ala Asp Pro Glu Gly Pro
165 170 175
ttc agc ctg gcc gag ggc tac aac gct gcc ggt aac aac cac ttt tcg 576
Phe Ser Leu Ala Glu Gly Tyr Asn Ala Ala Gly Asn Asn His Phe Ser
180 185 190
tgg gag agc gca acc gtt agc cac gtg caa tgg gat ggc ggt gag gcg 624
Trp Glu Ser Ala Thr Val Ser His Val Gln Trp Asp Gly Gly Glu Ala
195 200 205
gat cgt ggt gac ggt ctg gtc ggt gct cgt ttg gca ctg tgg ggt cac 672
Asp Arg Gly Asp Gly Leu Val Gly Ala Arg Leu Ala Leu Trp Gly His
210 215 220
tac aat gat tat ctg cgt acc acc ttc aat cgt gct cgt tgg gtc gac 720
Tyr Asn Asp Tyr Leu Arg Thr Thr Phe Asn Arg Ala Arg Trp Val Asp
225 230 235 240
agc gac ccg acg cgc cgt gac gct gca caa atc tat gcg ggc caa ttc 768
Ser Asp Pro Thr Arg Arg Asp Ala Ala Gln Ile Tyr Ala Gly Gln Phe

245					250					255						
acc	att	agc	ccg	gct	ggt	gcc	ggt	ccg	ggt	acg	ccg	tgg	ctg	ttt	acg	816
Thr	Ile	Ser	Pro	Ala	Gly	Ala	Gly	Pro	Gly	Thr	Pro	Trp	Leu	Phe	Thr	
			260				265				270					
gca	gac	att	gat	gat	agc	cat	ggt	gca	cgt	tgg	acg	cgt	ggc	ggc	cac	864
Ala	Asp	Ile	Asp	Asp	Ser	His	Gly	Ala	Arg	Trp	Thr	Arg	Gly	Gly	His	
			275				280				285					
att	gca	gag	cgt	ggc	ggc	cac	ttc	ttg	gat	gaa	gag	ttt	ggt	ctg	gca	912
Ile	Ala	Glu	Arg	Gly	Gly	His	Phe	Leu	Asp	Glu	Glu	Phe	Gly	Leu	Ala	
			290				295				300					
cgc	ctg	ttt	cag	ttc	tct	gtg	ccg	aaa	gat	cac	cca	cat	ttt	ctg	ttt	960
Arg	Leu	Phe	Gln	Phe	Ser	Val	Pro	Lys	Asp	His	Pro	His	Phe	Leu	Phe	
			305				310				315				320	
cac	ccg	ggt	ccg	ttt	gat	tcc	gag	gcc	tgg	cgt	cgt	ctg	caa	ttg	gct	1008
His	Pro	Gly	Pro	Phe	Asp	Ser	Glu	Ala	Trp	Arg	Arg	Leu	Gln	Leu	Ala	
			325				330				335					
ctg	gag	gat	gac	gac	gtt	ctg	ggt	ctg	acc	gtg	caa	tat	gcg	ttg	ttc	1056
Leu	Glu	Asp	Asp	Asp	Val	Leu	Gly	Leu	Thr	Val	Gln	Tyr	Ala	Leu	Phe	
			340				345				350					
aat	atg	agc	acc	ccg	cct	cag	ccg	aac	agc	ccg	gtt	ttt	cac	gat	atg	1104
Asn	Met	Ser	Thr	Pro	Pro	Gln	Pro	Asn	Ser	Pro	Val	Phe	His	Asp	Met	
			355				360				365					
gtc	ggt	gtt	gtc	ggt	ctg	tgg	cgt	cgt	ggt	gaa	ctg	gcg	agc	tac	ccg	1152
Val	Gly	Val	Val	Gly	Leu	Trp	Arg	Arg	Gly	Glu	Leu	Ala	Ser	Tyr	Pro	
			370				375				380					
gct	ggt	cgt	ctg	ctg	cgt	ccg	cgt	caa	ccg	ggt	ctg	ggt	gac	ctg	acc	1200
Ala	Gly	Arg	Leu	Leu	Arg	Pro	Arg	Gln	Pro	Gly	Leu	Gly	Asp	Leu	Thr	
			385				390				395				400	
ctg	cgc	gtc	aac	ggt	ggt	cgc	gtt	gcg	ctg	aat	ttg	gcg	tgt	gcc	att	1248
Leu	Arg	Val	Asn	Gly	Gly	Arg	Val	Ala	Leu	Asn	Leu	Ala	Cys	Ala	Ile	
			405				410				415					
ccg	ttc	agc	act	cgt	gcc	gcg	cag	cca	agc	gca	ccg	gac	cgc	ctg	acc	1296
Pro	Phe	Ser	Thr	Arg	Ala	Ala	Gln	Pro	Ser	Ala	Pro	Asp	Arg	Leu	Thr	
			420				425				430					
ccg	gac	ctg	ggt	gcc	aaa	ctg	ccg	ctg	ggc	gat	ctg	ctg	ctg	cgt	gat	1344
Pro	Asp	Leu	Gly	Ala	Lys	Leu	Pro	Leu	Gly	Asp	Leu	Leu	Leu	Arg	Asp	
			435				440				445					
gag	gac	ggc	gca	ctg	ttg	gca	cgt	gtg	ccg	cag	gct	ctg	tac	caa	gac	1392
Glu	Asp	Gly	Ala	Leu	Leu	Ala	Arg	Val	Pro	Gln	Ala	Leu	Tyr	Gln	Asp	
			450				455				460					
tat	tgg	acg	aat	cac	ggt	att	gtg	gac	ctg	ccg	ctg	ctg	cgc	gaa	ccg	1440
Tyr	Trp	Thr	Asn	His	Gly	Ile	Val	Asp	Leu	Pro	Leu	Leu	Arg	Glu	Pro	
			465				470				475				480	
cgt	ggt	agc	ttg	acc	ctg	agc	agc	gaa	ctg	gcg	gag	tgg	cgt	gag	caa	1488
Arg	Gly	Ser	Leu	Thr	Leu	Ser	Ser	Glu	Leu	Ala	Glu	Trp	Arg	Glu	Gln	
			485				490				495					
gac	tgg	gtc	acc	caa	agc	gac	gcg	tct	aac	ctg	tac	ctg	gag	gca	ccg	1536
Asp	Trp	Val	Thr	Gln	Ser	Asp	Ala	Ser	Asn	Leu	Tyr	Leu	Glu	Ala	Pro	
			500				505				510					
gat	cgc	cgt	cac	ggt	cgc	ttt	ttc	cct	gag	agc	atc	gcg	ctg	cgc	agc	1584
Asp	Arg	Arg	His	Gly	Arg	Phe	Phe	Pro	Glu	Ser	Ile	Ala	Leu	Arg	Ser	
			515				520				525					

50

tac ttt cgc ggt gaa gcg cgt gcg cgt ccg gat atc ccg cat cgt atc Tyr Phe Arg Gly Glu Ala Arg Ala Arg Pro Asp Ile Pro His Arg Ile 530 535 540	1632
gag ggc atg ggc ctg gtc ggc gtc gaa tct cgt cag gat ggc gac gct Glu Gly Met Gly Leu Val Gly Val Glu Ser Arg Gln Asp Gly Asp Ala 545 550 555	1680
gcg gaa tgg cgt ctg acg ggt ctg cgt ccg ggt ccg gca cgc att gtt Ala Glu Trp Arg Leu Thr Gly Leu Arg Pro Gly Pro Ala Arg Ile Val 565 570 575	1728
ctg gac gat ggt gcc gag gcg atc cct ctg cgt gtt ctg cct gac gat Leu Asp Asp Gly Ala Glu Ala Ile Pro Leu Arg Val Leu Pro Asp Asp 580 585 590	1776
tgg gcg ctg gat gac gcg acc gtc gaa gaa gtg gat tac gcc ttt ttg Trp Ala Leu Asp Asp Ala Thr Val Glu Glu Val Asp Tyr Ala Phe Leu 595 600 605	1824
tac cgc cac gtt atg gcg tat tac gag ctg gtg tat cca ttc atg agc Tyr Arg His Val Met Ala Tyr Tyr Glu Leu Val Tyr Pro Phe Met Ser 610 615 620	1872
gac aag gtg ttt tcc ctg gct gat cgt tgc aaa tgt gaa acg tac gca Asp Lys Val Phe Ser Leu Ala Asp Arg Cys Lys Cys Glu Thr Tyr Ala 625 630 635 640	1920
cgt ctg atg tgg cag atg tgt gat ccg cag aac cgc aac aag tcc tat Arg Leu Met Trp Gln Met Cys Asp Pro Gln Asn Arg Asn Lys Ser Tyr 645 650 655	1968
tac atg ccg agc acc cgc gaa ctg tcg gca ccg aaa gct cgt ttg ttc Tyr Met Pro Ser Thr Arg Glu Leu Ser Ala Pro Lys Ala Arg Leu Phe 660 665 670	2016
ttg aag tat ctg gcc cac gtg gaa ggc cag gca cgc ctg caa gca cct Leu Lys Tyr Leu Ala His Val Glu Gly Gln Ala Arg Leu Gln Ala Pro 675 680 685	2064
ccg cca gcg ggt ccg gca cgc att gaa tct aaa gcc cag ttg gcg gca Pro Pro Ala Gly Pro Ala Arg Ile Glu Ser Lys Ala Gln Leu Ala Ala 690 695 700	2112
gag ctg cgt aaa gcc gtc gac ctg gag ctg tct gtg atg ctg caa tac Glu Leu Arg Lys Ala Val Asp Leu Glu Leu Ser Val Met Leu Gln Tyr 705 710 715 720	2160
ctg tac gcg gcg tat agc att ccg aac tat gca cag ggc caa caa cgt Leu Tyr Ala Ala Tyr Ser Ile Pro Asn Tyr Ala Gln Gly Gln Gln Arg 725 730 735	2208
gtt cgt gac ggt gcg tgg acc gcc gag cag ctg caa ctg gcg tgc ggt Val Arg Asp Gly Ala Trp Thr Ala Glu Gln Leu Gln Leu Ala Cys Gly 740 745 750	2256
agc ggt gac cgt cgc cgt gat ggc ggt att cgt gca gca ctg ctg gaa Ser Gly Asp Arg Arg Arg Asp Gly Gly Ile Arg Ala Ala Leu Leu Glu 755 760 765	2304
att gct cat gaa gaa atg att cat tac ctg gtc gtt aac aac ctg ctg Ile Ala His Glu Glu Met Ile His Tyr Leu Val Val Asn Asn Leu Leu 770 775 780	2352
atg gcc ctg ggc gag ccg ttc tac gcg ggt gtc ccg ctg atg ggc gaa Met Ala Leu Gly Glu Pro Phe Tyr Ala Gly Val Pro Leu Met Gly Glu 785 790 795 800	2400

51															
gcg gca cgt cag gcg ttt ggc ctg gac acc gag ttc gct ctg gaa ccg	2448														
Ala Ala Arg Gln Ala Phe Gly Leu Asp Thr Glu Phe Ala Leu Glu Pro															
	805							810					815		
ttt agc gaa agc acg ctg gca cgt ttt gtt cgt ctg gaa tgg ccg cac	2496														
Phe Ser Glu Ser Thr Leu Ala Arg Phe Val Arg Leu Glu Trp Pro His															
	820							825					830		
ttt atc cca gca ccg ggc aaa tcc atc gcg gac tgc tat gcc gcc att	2544														
Phe Ile Pro Ala Pro Gly Lys Ser Ile Ala Asp Cys Tyr Ala Ala Ile															
	835							840					845		
cgt cag gcg ttt ttg gat ctg ccg gac ttg ttt ggt ggc gag gca ggt	2592														
Arg Gln Ala Phe Leu Asp Leu Pro Asp Leu Phe Gly Gly Glu Ala Gly															
	850							855					860		
aag cgt ggc ggt gaa cac cac ctg ttc ctg aat gag ctg acc aac cgt	2640														
Lys Arg Gly Gly Glu His His Leu Phe Leu Asn Glu Leu Thr Asn Arg															
	865							870					875		880
gcg cat ccg ggt tat caa ctg gaa gtt ttc gat cgc gac tgc gcg ctg	2688														
Ala His Pro Gly Tyr Gln Leu Glu Val Phe Asp Arg Asp Ser Ala Leu															
	885							890					895		
ttt ggt att gca ttt gtg acc gat cag ggc gaa ggt ggc gct ctg gac	2736														
Phe Gly Ile Ala Phe Val Thr Asp Gln Gly Glu Gly Gly Ala Leu Asp															
	900							905					910		
agc ccg cac tac gaa cat agc cat ttt caa cgt ctg cgt gaa atg agc	2784														
Ser Pro His Tyr Glu His Ser His Phe Gln Arg Leu Arg Glu Met Ser															
	915							920					925		
gcg cgt atc atg gct caa agc gca ccg ttc gaa ccg gcg ctg ccg gcg	2832														
Ala Arg Ile Met Ala Gln Ser Ala Pro Phe Glu Pro Ala Leu Pro Ala															
	930							935					940		
ttg cgt aat ccg gtt ctg gat gag agc ccg ggt tgc caa cgt gtc gca	2880														
Leu Arg Asn Pro Val Leu Asp Glu Ser Pro Gly Cys Gln Arg Val Ala															
	945							950					955		960
gac ggt cgt gcg cgt gcg ctg atg gca ttg tac caa ggc gtt tat gag	2928														
Asp Gly Arg Ala Arg Ala Leu Met Ala Leu Tyr Gln Gly Val Tyr Glu															
	965							970					975		
ctg atg ttt gcg atg atg gcg cag cac ttc gcc gtg aaa ccg ctg ggt	2976														
Leu Met Phe Ala Met Met Ala Gln His Phe Ala Val Lys Pro Leu Gly															
	980							985					990		
agc ttg cgt cgc agc cgc ctg atg aac gca gca atc gat ctg atg acc	3024														
Ser Leu Arg Arg Ser Arg Leu Met Asn Ala Ala Ile Asp Leu Met Thr															
	995							1000					1005		
ggt ctg ttg cgt ccg ctg agc tgc gcg ctg atg aac ctg cca agc	3069														
Gly Leu Leu Arg Pro Leu Ser Cys Ala Leu Met Asn Leu Pro Ser															
	1010							1015					1020		
ggc atc gcc ggt cgc acg gcc ggt ccg ccg ctg ccg ggt ccg gtt	3114														
Gly Ile Ala Gly Arg Thr Ala Gly Pro Pro Leu Pro Gly Pro Val															
	1025							1030					1035		
gac acc cgt agc tat gac gac tac gcg ctg ggc tgt cgc atg ctg	3159														
Asp Thr Arg Ser Tyr Asp Asp Tyr Ala Leu Gly Cys Arg Met Leu															
	1040							1045					1050		
gca cgc cgt tgc gag cgt ctg ctg gag cag gcg agc atg ctg gaa	3204														
Ala Arg Arg Cys Glu Arg Leu Leu Glu Gln Ala Ser Met Leu Glu															
	1055							1060					1065		
ccg ggt tgg ctg ccg gat gcg cag atg gag ctg ctg gat ttc tat	3249														

Pro Gly Trp Leu Pro Asp Ala Gln Met Glu⁵² Leu Leu Asp Phe Tyr
 1070 1075 1080
 cgt cgc caa atg ctg gac ttg gcg tgc ggc aaa ctg agc cgc gag 3294
 Arg Arg Gln Met Leu Asp Leu Ala Cys Gly Lys Leu Ser Arg Glu
 1085 1090 1095
 gcc taa 3300
 Aia

<210> 64
 <211> 1099
 <212> PRT
 <213> künstliche Sequenz
 <220>
 <223> synthetisches Konstrukt
 <400> 64

Met Pro Gly Glu Lys Pro Tyr Ala Cys Pro Val Glu Ser Cys Asp Arg
 1 5 10 15
 Arg Phe Ser Arg Ser Asp Glu Leu Thr Arg His Ile Arg Ile His Thr
 20 25 30
 Gly Gln Lys Pro Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser Arg
 35 40 45
 Ser Asp His Leu Thr Thr His Ile Arg Thr His Thr Gly Glu Lys Pro
 50 55 60
 Phe Ala Cys Asp Ile Cys Gly Arg Lys Phe Ala Arg Ser Asp Glu Arg
 65 70 75 80
 Lys Arg His Thr Lys Ile His Thr Gly Gly Gly Ser Gly Gly Gly Ser
 85 90 95
 Gly Gly Ser Thr Arg Met Ser Ile Leu Asp Phe Pro Arg Ile His Phe
 100 105 110
 Arg Gly Trp Ala Arg Val Asn Ala Pro Thr Ala Asn Arg Asp Pro His
 115 120 125
 Gly His Ile Asp Met Ala Ser Asn Thr Val Ala Met Ala Gly Glu Pro
 130 135 140
 Phe Asp Leu Ala Arg His Pro Thr Glu Phe His Arg His Leu Arg Ser
 145 150 155 160
 Leu Gly Pro Arg Phe Gly Leu Asp Gly Arg Ala Asp Pro Glu Gly Pro
 165 170 175
 Phe Ser Leu Ala Glu Gly Tyr Asn Ala Ala Gly Asn Asn His Phe Ser
 180 185 190

53

Trp Glu Ser Ala Thr Val Ser His Val Gln Trp Asp Gly Gly Glu Ala
195 200 205

Asp Arg Gly Asp Gly Leu Val Gly Ala Arg Leu Ala Leu Trp Gly His
210 215 220

Tyr Asn Asp Tyr Leu Arg Thr Thr Phe Asn Arg Ala Arg Trp Val Asp
225 230 235 240

Ser Asp Pro Thr Arg Arg Asp Ala Ala Gln Ile Tyr Ala Gly Gln Phe
245 250 255

Thr Ile Ser Pro Ala Gly Ala Gly Pro Gly Thr Pro Trp Leu Phe Thr
260 265 270

Ala Asp Ile Asp Asp Ser His Gly Ala Arg Trp Thr Arg Gly Gly His
275 280 285

Ile Ala Glu Arg Gly Gly His Phe Leu Asp Glu Glu Phe Gly Leu Ala
290 295 300

Arg Leu Phe Gln Phe Ser Val Pro Lys Asp His Pro His Phe Leu Phe
305 310 315 320

His Pro Gly Pro Phe Asp Ser Glu Ala Trp Arg Arg Leu Gln Leu Ala
325 330 335

Leu Glu Asp Asp Asp Val Leu Gly Leu Thr Val Gln Tyr Ala Leu Phe
340 345 350

Asn Met Ser Thr Pro Pro Gln Pro Asn Ser Pro Val Phe His Asp Met
355 360 365

Val Gly Val Val Gly Leu Trp Arg Arg Gly Glu Leu Ala Ser Tyr Pro
370 375 380

Ala Gly Arg Leu Leu Arg Pro Arg Gln Pro Gly Leu Gly Asp Leu Thr
385 390 395 400

Leu Arg Val Asn Gly Gly Arg Val Ala Leu Asn Leu Ala Cys Ala Ile
405 410 415

Pro Phe Ser Thr Arg Ala Ala Gln Pro Ser Ala Pro Asp Arg Leu Thr
420 425 430

Pro Asp Leu Gly Ala Lys Leu Pro Leu Gly Asp Leu Leu Leu Arg Asp
435 440 445

Glu Asp Gly Ala Leu Leu Ala Arg Val Pro Gln Ala Leu Tyr Gln Asp
450 455 460

Tyr Trp Thr Asn His Gly Ile Val Asp Leu Pro Leu Leu Arg Glu Pro
 465 470 475 480

Arg Gly Ser Leu Thr Leu Ser Ser Glu Leu Ala Glu Trp Arg Glu Gln
 485 490 495

Asp Trp Val Thr Gln Ser Asp Ala Ser Asn Leu Tyr Leu Glu Ala Pro
 500 505 510

Asp Arg Arg His Gly Arg Phe Phe Pro Glu Ser Ile Ala Leu Arg Ser
 515 520 525

Tyr Phe Arg Gly Glu Ala Arg Ala Arg Pro Asp Ile Pro His Arg Ile
 530 535 540

Glu Gly Met Gly Leu Val Gly Val Glu Ser Arg Gln Asp Gly Asp Ala
 545 550 555 560

Ala Glu Trp Arg Leu Thr Gly Leu Arg Pro Gly Pro Ala Arg Ile Val
 565 570 575

Leu Asp Asp Gly Ala Glu Ala Ile Pro Leu Arg Val Leu Pro Asp Asp
 580 585 590

Trp Ala Leu Asp Asp Ala Thr Val Glu Glu Val Asp Tyr Ala Phe Leu
 595 600 605

Tyr Arg His Val Met Ala Tyr Tyr Glu Leu Val Tyr Pro Phe Met Ser
 610 615 620

Asp Lys Val Phe Ser Leu Ala Asp Arg Cys Lys Cys Glu Thr Tyr Ala
 625 630 635 640

Arg Leu Met Trp Gln Met Cys Asp Pro Gln Asn Arg Asn Lys Ser Tyr
 645 650 655

Tyr Met Pro Ser Thr Arg Glu Leu Ser Ala Pro Lys Ala Arg Leu Phe
 660 665 670

Leu Lys Tyr Leu Ala His Val Glu Gly Gln Ala Arg Leu Gln Ala Pro
 675 680 685

Pro Pro Ala Gly Pro Ala Arg Ile Glu Ser Lys Ala Gln Leu Ala Ala
 690 695 700

Glu Leu Arg Lys Ala Val Asp Leu Glu Leu Ser Val Met Leu Gln Tyr
 705 710 715 720

Leu Tyr Ala Ala Tyr Ser Ile Pro Asn Tyr Ala Gln Gly Gln Gln Arg
 725 730 735

Val Arg Asp Gly Ala Trp Thr Ala Glu Gln Leu Gln Leu Ala Cys Gly

740 55 750
 Ser Gly Asp Arg Arg Arg Asp Gly Gly Ile Arg Ala Ala Leu Leu Glu
 755 760 765
 Ile Ala His Glu Glu Met Ile His Tyr Leu Val Val Asn Asn Leu Leu
 770 775 780
 Met Ala Leu Gly Glu Pro Phe Tyr Ala Gly Val Pro Leu Met Gly Glu
 785 790 795 800
 Ala Ala Arg Gln Ala Phe Gly Leu Asp Thr Glu Phe Ala Leu Glu Pro
 805 810 815
 Phe Ser Glu Ser Thr Leu Ala Arg Phe Val Arg Leu Glu Trp Pro His
 820 825 830
 Phe Ile Pro Ala Pro Gly Lys Ser Ile Ala Asp Cys Tyr Ala Ala Ile
 835 840 845
 Arg Gln Ala Phe Leu Asp Leu Pro Asp Leu Phe Gly Gly Glu Ala Gly
 850 855 860
 Lys Arg Gly Gly Glu His His Leu Phe Leu Asn Glu Leu Thr Asn Arg
 865 870 875 880
 Ala His Pro Gly Tyr Gln Leu Glu Val Phe Asp Arg Asp Ser Ala Leu
 885 890 895
 Phe Gly Ile Ala Phe Val Thr Asp Gln Gly Glu Gly Gly Ala Leu Asp
 900 905 910
 Ser Pro His Tyr Glu His Ser His Phe Gln Arg Leu Arg Glu Met Ser
 915 920 925
 Ala Arg Ile Met Ala Gln Ser Ala Pro Phe Glu Pro Ala Leu Pro Ala
 930 935 940
 Leu Arg Asn Pro Val Leu Asp Glu Ser Pro Gly Cys Gln Arg Val Ala
 945 950 955 960
 Asp Gly Arg Ala Arg Ala Leu Met Ala Leu Tyr Gln Gly Val Tyr Glu
 965 970 975
 Leu Met Phe Ala Met Met Ala Gln His Phe Ala Val Lys Pro Leu Gly
 980 985 990
 Ser Leu Arg Arg Ser Arg Leu Met Asn Ala Ala Ile Asp Leu Met Thr
 995 1000 1005
 Gly Leu Leu Arg Pro Leu Ser Cys Ala Leu Met Asn Leu Pro Ser
 1010 1015 1020

56

Gly Ile Ala Gly Arg Thr Ala Gly Pro Pro Leu Pro Gly Pro Val
1025 1030 1035

Asp Thr Arg Ser Tyr Asp Asp Tyr Ala Leu Gly Cys Arg Met Leu
1040 1045 1050

Ala Arg Arg Cys Glu Arg Leu Leu Glu Gln Ala Ser Met Leu Glu
1055 1060 1065

Pro Gly Trp Leu Pro Asp Ala Gln Met Glu Leu Leu Asp Phe Tyr
1070 1075 1080

Arg Arg Gln Met Leu Asp Leu Ala Cys Gly Lys Leu Ser Arg Glu
1085 1090 1095

Ala

<210> 65
 <211> 873
 <212> DNA
 <213> künstliche Sequenz

<220>
 <223> Znf_PBSII-Linkerpeptid-vioE

<220>
 <221> CDS
 <222> (1)..(873)
 <223> Znf_PBSII-Linkerpeptid-vio

<400> 65	
atg ccg ggt gaa aaa ccg tat gca tgt cct gaa tgt ggt aaa agc ttt	48
Met Pro Gly Glu Lys Pro Tyr Ala Cys Pro Glu Cys Gly Lys Ser Phe	
1 5 10 15	
agc cag cgt gca aat ctg cgt gca cat cag cgt acc cat aca ggt gaa	96
Ser Gln Arg Ala Asn Leu Arg Ala His Gln Arg Thr His Thr Gly Glu	
20 25 30	
aag cct tat aaa tgc cca gaa tgc ggc aaa agc ttt tca cgt agc gat	144
Lys Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser Arg Ser Asp	
35 40 45	
cat ctg acc acc cat cag cgc aca cat act ggc gag aaa cct tac aaa	192
His Leu Thr Thr His Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Lys	
50 55 60	
tgt cca gag tgt ggt aaa tca ttt agc cgt agt gat gtt ctg gtt cgt	240
Cys Pro Glu Cys Gly Lys Ser Phe Ser Arg Ser Asp Val Leu Val Arg	
65 70 75 80	
cat cag aga acc cac acg ggt ggt ggt agc ggt ggt ggt tca ggt ggt	288
His Gln Arg Thr His Thr Gly Gly Gly Ser Gly Gly Gly Ser Gly Gly	
85 90 95	
agt act aga atg gag aac cgt gag cca cca ctg ttg cca gcc cgt tgg	336
Ser Thr Arg Met Glu Asn Arg Glu Pro Pro Leu Leu Pro Ala Arg Trp	
100 105 110	

57
 agc agc gcc tat gtc tct tat tgg agc ccg atg ctg ccg gat gac cag 384
 Ser Ser Ala Tyr Val Ser Tyr Trp Ser Pro Met Leu Pro Asp Asp Gln
 115 120 125
 ctg acc agc ggc tat tgc tgg ttc gac tat gaa cgt gac atc tgt cgt 432
 Leu Thr Ser Gly Tyr Cys Trp Phe Asp Tyr Glu Arg Asp Ile Cys Arg
 130 135 140
 att gac ggc ctg ttc aat ccg tgg agc gag cgt gat act ggt tat cgc 480
 Ile Asp Gly Leu Phe Asn Pro Trp Ser Glu Arg Asp Thr Gly Tyr Arg
 145 150 155 160
 ctg tgg atg tcg gag gtt ggt aat gcg gcc agc ggc cgt acc tgg aaa 528
 Leu Trp Met Ser Glu Val Gly Asn Ala Ala Ser Gly Arg Thr Trp Lys
 165 170 175
 caa aaa gtc gcc tat ggt cgt gag cgt acc gcc ctg ggt gaa cag ctg 576
 Gln Lys Val Ala Tyr Gly Arg Glu Arg Thr Ala Leu Gly Glu Gln Leu
 180 185 190
 tgt gag cgt ccg ctg gat gat gag act ggc cct ttt gcc gaa ttg ttc 624
 Cys Glu Arg Pro Leu Asp Asp Glu Thr Gly Pro Phe Ala Glu Leu Phe
 195 200 205
 ctg cca cgc gat gtc ctg cgc cgt ctg ggt gcc cgt cac att ggc cgt 672
 Leu Pro Arg Asp Val Leu Arg Arg Leu Gly Ala Arg His Ile Gly Arg
 210 215 220
 cgc gtg gtt ctg ggt cgc gaa gcg gac ggt tgg cgt tac cag cgc cca 720
 Arg Val Val Leu Gly Arg Glu Ala Asp Gly Trp Arg Tyr Gln Arg Pro
 225 230 235 240
 ggt aaa ggt ccg agc acc ctg tac ctg gat gcg gcg agc ggc act cca 768
 Gly Lys Gly Pro Ser Thr Leu Tyr Leu Asp Ala Ala Ser Gly Thr Pro
 245 250 255
 ctg cgc atg gtc acc ggc gat gaa gcg tcg cgt gca agc ctg cgt gat 816
 Leu Arg Met Val Thr Gly Asp Glu Ala Ser Arg Ala Ser Leu Arg Asp
 260 265 270
 ttt ccg aat gtg agc gag gcg gag atc ccg gac gcg gtt ttc gcg gcc 864
 Phe Pro Asn Val Ser Glu Ala Glu Ile Pro Asp Ala Val Phe Ala Ala
 275 280 285
 aag cgc taa 873
 Lys Arg
 290

<210> 66
 <211> 290
 <212> PRT
 <213> künstliche Sequenz

<220>
 <223> synthetisches Konstrukt

<400> 66

Met Pro Gly Glu Lys Pro Tyr Ala Cys Pro Glu Cys Gly Lys Ser Phe
 1 5 10 15

Ser Gln Arg Ala Asn Leu Arg Ala His Gln Arg Thr His Thr Gly Glu
 20 25 30

Lys Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser Arg Ser Asp
 35 40 45

58

His Leu Thr Thr His Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Lys
 50 55 60
 Cys Pro Glu Cys Gly Lys Ser Phe Ser Arg Ser Asp Val Leu Val Arg
 65 70 75 80
 His Gln Arg Thr His Thr Gly Gly Gly Ser Gly Gly Gly Ser Gly Gly
 85 90 95
 Ser Thr Arg Met Glu Asn Arg Glu Pro Pro Leu Leu Pro Ala Arg Trp
 100 105 110
 Ser Ser Ala Tyr Val Ser Tyr Trp Ser Pro Met Leu Pro Asp Asp Gln
 115 120 125
 Leu Thr Ser Gly Tyr Cys Trp Phe Asp Tyr Glu Arg Asp Ile Cys Arg
 130 135 140
 Ile Asp Gly Leu Phe Asn Pro Trp Ser Glu Arg Asp Thr Gly Tyr Arg
 145 150 155 160
 Leu Trp Met Ser Glu Val Gly Asn Ala Ala Ser Gly Arg Thr Trp Lys
 165 170 175
 Gln Lys Val Ala Tyr Gly Arg Glu Arg Thr Ala Leu Gly Glu Gln Leu
 180 185 190
 Cys Glu Arg Pro Leu Asp Asp Glu Thr Gly Pro Phe Ala Glu Leu Phe
 195 200 205
 Leu Pro Arg Asp Val Leu Arg Arg Leu Gly Ala Arg His Ile Gly Arg
 210 215 220
 Arg Val Val Leu Gly Arg Glu Ala Asp Gly Trp Arg Tyr Gln Arg Pro
 225 230 235 240
 Gly Lys Gly Pro Ser Thr Leu Tyr Leu Asp Ala Ala Ser Gly Thr Pro
 245 250 255
 Leu Arg Met Val Thr Gly Asp Glu Ala Ser Arg Ala Ser Leu Arg Asp
 260 265 270
 Phe Pro Asn Val Ser Glu Ala Glu Ile Pro Asp Ala Val Phe Ala Ala
 275 280 285
 Lys Arg
 290

<210> 67
 <211> 1419
 <212> DNA

59

<213> künstliche Sequenz

<220>

<223> Znf_HivC-Linkerpeptid-vioD

<220>

<221> CDS

<222> (1)..(1419)

<223> Znf_HivC-Linkerpeptid-vioD

<400> 67

atg ccc ttc cag tgt cgc att tgc atg cgg aac ttt tcg ctc agg acg	48
Met Pro Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser Leu Arg Thr	
1 5 10 15	
gac ctt gac agg cat acc cgt act cat acc ggt gaa aaa ccg ttt cag	96
Asp Leu Asp Arg His Thr Arg Thr His Thr Gly Glu Lys Pro Phe Gln	
20 25 30	
tgt cgg atc tgt atg cga aat ttc tcc ctc agc cag acc ttg cgc cgc	144
Cys Arg Ile Cys Met Arg Asn Phe Ser Leu Ser Gln Thr Leu Arg Arg	
35 40 45	
cat cta cgt acg cac acc ggc gag aag cca ttc caa tgc cga ata tgc	192
His Leu Arg Thr His Thr Gly Glu Lys Pro Phe Gln Cys Arg Ile Cys	
50 55 60	
atg cgc aac ttc agt ctc agg agc aac ctg ggg cgg cac cta aaa acc	240
Met Arg Asn Phe Ser Leu Arg Ser Asn Leu Gly Arg His Leu Lys Thr	
65 70 75 80	
cac aca gga gaa aaa act aga ggc ggg tca ggc ggg ggc tca ggc ggc	288
His Thr Gly Glu Lys Thr Arg Gly Gly Ser Gly Gly Gly Ser Gly Gly	
85 90 95	
tca act aga atg aag att ctg gtc att ggt gct ggt cca gct ggt ctg	336
Ser Thr Arg Met Lys Ile Leu Val Ile Gly Ala Gly Pro Ala Gly Leu	
100 105 110	
gtt ttc gca tcc caa ctg aag cag gca cgc cct ttg tgg gcc att gac	384
Val Phe Ala Ser Gln Leu Lys Gln Ala Arg Pro Leu Trp Ala Ile Asp	
115 120 125	
atc gtg gag aag aat gac gag caa gaa gtg ctg ggc tgg ggt gtc gtg	432
Ile Val Glu Lys Asn Asp Glu Gln Glu Val Leu Gly Trp Gly Val Val	
130 135 140	
ctg cct ggc cgt ccg ggt cag cac ccg gcg aac ccg ctg tcc tat ctg	480
Leu Pro Gly Arg Pro Gly Gln His Pro Ala Asn Pro Leu Ser Tyr Leu	
145 150 155 160	
gat gca ccg gag cgt ctg aat ccg caa ttt ctg gag gac ttc aaa ctg	528
Asp Ala Pro Glu Arg Leu Asn Pro Gln Phe Leu Glu Asp Phe Lys Leu	
165 170 175	
gtg cat cat aat gag ccg tcc ttg atg tcc acg ggc gtt ttg ttg tgc	576
Val His His Asn Glu Pro Ser Leu Met Ser Thr Gly Val Leu Leu Cys	
180 185 190	
ggc gtg gag cgt cgc ggt ctg gtt cac gcg ctg cgc gat aag tgc cgc	624
Gly Val Glu Arg Arg Gly Leu Val His Ala Leu Arg Asp Lys Cys Arg	
195 200 205	
agc caa ggc att gct att cgt ttc gaa agc ccg ttg ctg gaa cac ggt	672
Ser Gln Gly Ile Ala Ile Arg Phe Glu Ser Pro Leu Leu Glu His Gly	
210 215 220	
gag ctg ccg ctg gcg gac tat gat ctg gtg gtc ctg gct aat ggt gtt	720

60
 Glu Leu Pro Leu Ala Asp Tyr Asp Leu Val Val Leu Ala Asn Gly Val
 225 230 235 240
 aat cac aaa acc gcg cat ttc acc gag gct ctg gtc ccg cag gtg gac 768
 Asn His Lys Thr Ala His Phe Thr Glu Ala Leu Val Pro Gln Val Asp
 245 250 255
 tac ggc cgc aat aag tac att tgg tat ggc act agc cag ctg ttc gat 816
 Tyr Gly Arg Asn Lys Tyr Ile Trp Tyr Gly Thr Ser Gln Leu Phe Asp
 260 265 270
 cag atg aat ctg gtt ttt cgt acc cat ggt aaa gat atc ttt atc gcg 864
 Gln Met Asn Leu Val Phe Arg Thr His Gly Lys Asp Ile Phe Ile Ala
 275 280 285
 cat gcc tat aag tat agc gat acc atg agc acg ttc att gtc gaa tgt 912
 His Ala Tyr Lys Tyr Ser Asp Thr Met Ser Thr Phe Ile Val Glu Cys
 290 295 300
 agc gaa gag act tac gca cgc gca cgc ctg ggc gaa atg tcc gaa gag 960
 Ser Glu Glu Thr Tyr Ala Arg Ala Arg Leu Gly Glu Met Ser Glu Glu
 305 310 315 320
 gcg agc gca gaa tac gtt gct aag gtg ttc cag gcc gag ctg ggt ggt 1008
 Ala Ser Ala Glu Tyr Val Ala Lys Val Phe Gln Ala Glu Leu Gly Gly
 325 330 335
 cac ggc ctg gtg agc cag ccg ggt ctg ggt tgg cgt aac ttc atg acg 1056
 His Gly Leu Val Ser Gln Pro Gly Leu Gly Trp Arg Asn Phe Met Thr
 340 345 350
 ttg tct cat gac cgt tgt cat gat ggt aag ttg gtt ctg ctg ggt gac 1104
 Leu Ser His Asp Arg Cys His Asp Gly Lys Leu Val Leu Gly Asp
 355 360 365
 gcg ctg caa agc ggt cac ttt agc atc ggc cac ggc acc acg atg gcc 1152
 Ala Leu Gln Ser Gly His Phe Ser Ile Gly His Gly Thr Thr Met Ala
 370 375 380
 gtg gtg gtg gcg cag ctg ctg gtt aaa gcg ctg tgt acc gaa gat ggt 1200
 Val Val Val Ala Gln Leu Leu Val Lys Ala Leu Cys Thr Glu Asp Gly
 385 390 395 400
 gtg cct gcc gcg ctg aaa cgt ttc gaa gag cgt gcc ctg ccg ctg gtg 1248
 Val Pro Ala Ala Leu Lys Arg Phe Glu Glu Arg Ala Leu Pro Leu Val
 405 410 415
 cag ttg ttc cgt ggc cac gca gac aac agc cgc gtt tgg ttc gaa acc 1296
 Gln Leu Phe Arg Gly His Ala Asp Asn Ser Arg Val Trp Phe Glu Thr
 420 425 430
 gtc gaa gag cgc atg cac ctg tcc tcg gcg gaa ttt gtg caa agc ttc 1344
 Val Glu Glu Arg Met His Leu Ser Ser Ala Glu Phe Val Gln Ser Phe
 435 440 445
 gac gca cgc cgc aaa agc ctg ccg ccg atg ccg gaa gca ctg gcg cag 1392
 Asp Ala Arg Arg Lys Ser Leu Pro Pro Met Pro Glu Ala Leu Ala Gln
 450 455 460
 aat ctg cgt tat gct ttg cag cgc tga 1419
 Asn Leu Arg Tyr Ala Leu Gln Arg
 465 470

<210> 68
 <211> 472
 <212> PRT
 <213> künstliche Sequenz

61

<220>

<223> synthetisches Konstrukt

<400> 68

```

Met Pro Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser Leu Arg Thr
1      5      10      15

Asp Leu Asp Arg His Thr Arg Thr His Thr Gly Glu Lys Pro Phe Gln
20      25      30

Cys Arg Ile Cys Met Arg Asn Phe Ser Leu Ser Gln Thr Leu Arg Arg
35      40      45

His Leu Arg Thr His Thr Gly Glu Lys Pro Phe Gln Cys Arg Ile Cys
50      55      60

Met Arg Asn Phe Ser Leu Arg Ser Asn Leu Gly Arg His Leu Lys Thr
65      70      75      80

His Thr Gly Glu Lys Thr Arg Gly Gly Ser Gly Gly Gly Ser Gly Gly
85      90      95

Ser Thr Arg Met Lys Ile Leu Val Ile Gly Ala Gly Pro Ala Gly Leu
100     105     110

Val Phe Ala Ser Gln Leu Lys Gln Ala Arg Pro Leu Trp Ala Ile Asp
115     120     125

Ile Val Glu Lys Asn Asp Glu Gln Glu Val Leu Gly Trp Gly Val Val
130     135     140

Leu Pro Gly Arg Pro Gly Gln His Pro Ala Asn Pro Leu Ser Tyr Leu
145     150     155     160

Asp Ala Pro Glu Arg Leu Asn Pro Gln Phe Leu Glu Asp Phe Lys Leu
165     170     175

Val His His Asn Glu Pro Ser Leu Met Ser Thr Gly Val Leu Leu Cys
180     185     190

Gly Val Glu Arg Arg Gly Leu Val His Ala Leu Arg Asp Lys Cys Arg
195     200     205

Ser Gln Gly Ile Ala Ile Arg Phe Glu Ser Pro Leu Leu Glu His Gly
210     215     220

Glu Leu Pro Leu Ala Asp Tyr Asp Leu Val Val Leu Ala Asn Gly Val
225     230     235     240

Asn His Lys Thr Ala His Phe Thr Glu Ala Leu Val Pro Gln Val Asp
245     250     255

```

Tyr Gly Arg Asn Lys Tyr Ile Trp Tyr Gly Thr Ser Gln Leu Phe Asp
 260 265 270
 Gln Met Asn Leu Val Phe Arg Thr His Gly Lys Asp Ile Phe Ile Ala
 275 280 285
 His Ala Tyr Lys Tyr Ser Asp Thr Met Ser Thr Phe Ile Val Glu Cys
 290 295 300
 Ser Glu Glu Thr Tyr Ala Arg Ala Arg Leu Gly Glu Met Ser Glu Glu
 305 310 315 320
 Ala Ser Ala Glu Tyr Val Ala Lys Val Phe Gln Ala Glu Leu Gly Gly
 325 330 335
 His Gly Leu Val Ser Gln Pro Gly Leu Gly Trp Arg Asn Phe Met Thr
 340 345 350
 Leu Ser His Asp Arg Cys His Asp Gly Lys Leu Val Leu Leu Gly Asp
 355 360 365
 Ala Leu Gln Ser Gly His Phe Ser Ile Gly His Gly Thr Thr Met Ala
 370 375 380
 Val Val Val Ala Gln Leu Leu Val Lys Ala Leu Cys Thr Glu Asp Gly
 385 390 395 400
 Val Pro Ala Ala Leu Lys Arg Phe Glu Glu Arg Ala Leu Pro Leu Val
 405 410 415
 Gln Leu Phe Arg Gly His Ala Asp Asn Ser Arg Val Trp Phe Glu Thr
 420 425 430
 Val Glu Glu Arg Met His Leu Ser Ser Ala Glu Phe Val Gln Ser Phe
 435 440 445
 Asp Ala Arg Arg Lys Ser Leu Pro Pro Met Pro Glu Ala Leu Ala Gln
 450 455 460
 Asn Leu Arg Tyr Ala Leu Gln Arg
 465 470

<210> 69
 <211> 1860
 <212> DNA
 <213> künstliche Sequenz
 <220>
 <223> Znf_Gli1-Linkerpeptid-vioc
 <220>
 <221> CDS
 <222> (1)..(1860)
 <223> Znf_Gli1-Linkerpeptid-vioc

400>	69																
atg Met 1	aag Lys	cgt Arg	gag Glu	cct Pro 5	gaa Glu	tct Ser	gtg Val	tat Tyr	gaa Glu 10	act Thr	gac Asp	tgc Cys	cgt Arg	tgg Trp 15	gat Asp		48
ggc Gly	tgt Cys	agc Ser	cag Gln 20	gaa Glu	ttt Phe	gac Asp	tcc Ser	caa Gln 25	gag Glu	cag Gln	ctg Leu	gtg Val	cac His 30	cac His	atc Ile		96
aac Asn	agc Ser	gag Glu 35	cac His	atc Ile	cac His	ggg Gly	gag Glu 40	cgg Arg	aag Lys	gag Glu	ttc Phe	gtg Val 45	tgc Cys	cac His	tgg Trp		144
ggg Gly	ggc Gly 50	tgc Cys	tcc Ser	agg Arg	gag Glu	ctg Leu 55	agg Arg	ccc Pro	ttc Phe	aaa Lys	gcc Ala 60	cag Gln	tac Tyr	atg Met	ctg Leu		192
gtg Val 65	ggt Val	cac His	atg Met	cgc Arg	aga Arg 70	cac His	act Thr	ggc Gly	gag Glu	aag Lys 75	cca Pro	cac His	aag Lys	tgc Cys	acg Thr 80		240
ttt Phe	gaa Glu	ggg Gly	tgc Cys	cgg Arg 85	aag Lys	tca Ser	tac Tyr	tca Ser	cgc Arg 90	ctc Leu	gaa Glu	aac Asn	ctg Leu	aag Lys 95	acg Thr		288
cac His	ctg Leu	cgg Arg	tca Ser 100	cac His	acg Thr	ggt Gly	gag Glu	aag Lys 105	cca Pro	tac Tyr	atg Met	tgt Cys	gag Glu 110	cac His	gag Glu		336
ggc Gly	tgt Cys	agt Ser 115	aaa Lys	gcc Ala	ttc Phe	agc Ser	aat Asn 120	gcc Ala	agt Ser	gac Asp	cga Arg	gcc Ala 125	aag Lys	cac His	cag Gln		384
aat Asn	cgg Arg 130	acc Thr	cat His	tcc Ser	aat Asn	gag Glu 135	aag Lys	ccg Pro	tat Tyr	gta Val	tgt Cys 140	aag Lys	ctc Leu	cct Pro	ggc Gly		432
tgc Cys 145	acc Thr	aaa Lys	cgc Arg	tat Tyr	aca Thr 150	gat Asp	cct Pro	agc Ser	tcg Ser	ctg Leu 155	cga Arg	aaa Lys	cat His	gtc Val	aag Lys 160		480
aca Thr	gtg Val	cat His	ggg Gly	cct Pro 165	gac Asp	gcc Ala	cat His	gtg Val	acc Thr 170	aaa Lys	cgg Arg	cac His	cgt Arg	ggg Gly 175	gat Asp		528
act Thr	aga Arg	ggc Gly	ggg Gly 180	tca Ser	ggc Gly	ggg Gly	ggc Gly	tca Ser 185	ggc Gly	ggg Gly	tca Ser	act Thr	aga Arg 190	atg Met	aaa Lys		576
cgt Arg	gcg Ala	att Ile 195	atc Ile	gtt Val	ggt Gly	ggc Gly	ggc Gly 200	ctg Leu	gcg Ala	ggt Gly	ggc Gly	ctg Leu 205	acc Thr	gcg Ala	atc Ile		624
tac Tyr	ctg Leu 210	gcg Ala	aag Lys	cgt Arg	ggc Gly	tac Tyr 215	gaa Glu	gtg Val	cac His	gtc Val	gtg Val 220	gag Glu	aag Lys	cgt Arg	ggt Gly		672
gat Asp 225	cct Pro	ctg Leu	cgc Arg	gat Asp	ctg Leu 230	agc Ser	tct Ser	tac Tyr	gtg Val	gac Asp 235	gtt Val	gtt Val	agc Ser	agc Ser	cgt Arg 240		720
gcg Ala	atc Ile	ggc Gly	gtg Val	agc Ser 245	atg Met	acc Thr	gtt Val	cgt Arg	ggt Gly 250	atc Ile	aag Lys	agc Ser	gtt Val	ttg Leu 255	gct Ala		768
gcg Ala	ggc Gly	att Ile	ccg Pro 260	cgt Arg	gca Ala	gag Glu	ctg Leu	gat Asp 265	gcg Ala	tgt Cys	ggc Gly	gaa Glu	ccg Pro 270	atc Ile	gtg Val		816

64

gca atg gct ttc tcc gtg ggt ggt cag tat cgc atg cgc gaa ctg aag Ala Met Ala Phe Ser Val Gly Gly Gln Tyr Arg Met Arg Glu Leu Lys 275 280 285	864
ccg ttg gag gat ttc cgt ccg ctg agc ttg aac cgt gcg gcg ttt caa Pro Leu Glu Asp Phe Arg Pro Leu Ser Leu Asn Arg Ala Ala Phe Gln 290 295 300	912
aag ctg ctg aac aaa tac gcg aac ctg gca ggc gtt cgt tac tac ttt Lys Leu Leu Asn Lys Tyr Ala Asn Leu Ala Gly Val Arg Tyr Tyr Phe 305 310 315 320	960
gag cat aag tgc ctg gat gtt gac ctg gat ggt aag agc gtg ttg att Glu His Lys Cys Leu Asp Val Asp Leu Asp Gly Lys Ser Val Leu Ile 325 330 335	1008
cag ggc aaa gat ggt cag ccg cag cgt ctg caa ggt gac atg att atc Gln Gly Lys Asp Gly Gln Pro Gln Arg Leu Gln Gly Asp Met Ile Ile 340 345 350	1056
ggt gcg gat ggc gcc cac agc gcc gtc cgt cag gcg atg cag agc ggc Gly Ala Asp Gly Ala His Ser Ala Val Arg Gln Ala Met Gln Ser Gly 355 360 365	1104
ctg cgt cgt ttc gag ttc cag caa acg ttc ttc cgc cat ggc tac aaa Leu Arg Arg Phe Glu Phe Gln Gln Thr Phe Phe Arg His Gly Tyr Lys 370 375 380	1152
acc ctg gtt ttg ccg gac gcg caa gca ctg ggt tac cgt aaa gac acg Thr Leu Val Leu Pro Asp Ala Gln Ala Leu Gly Tyr Arg Lys Asp Thr 385 390 395 400	1200
ctg tac ttt ttc ggc atg gat tcc ggt ggc ctg ttc gcg ggt cgt gcg Leu Tyr Phe Phe Gly Met Asp Ser Gly Gly Leu Phe Ala Gly Arg Ala 405 410 415	1248
gct acg atc cca gat ggt agc gtc agc atc gcc gtt tgc ctg ccg tac Ala Thr Ile Pro Asp Gly Ser Val Ser Ile Ala Val Cys Leu Pro Tyr 420 425 430	1296
tcg ggt agc cct tcc ctg acg acc acc gac gaa ccg acg atg cgt gcg Ser Gly Ser Pro Ser Leu Thr Thr Thr Asp Glu Pro Thr Met Arg Ala 435 440 445	1344
ttc ttc gat cgt tac ttc ggt ggc ctg ccg cgt gac gcg cgt gac gaa Phe Phe Asp Arg Tyr Phe Gly Gly Leu Pro Arg Asp Ala Arg Asp Glu 450 455 460	1392
atg ctg cgt cag ttt ctg gcg aag ccg agc aac gac ctg att aac gtg Met Leu Arg Gln Phe Leu Ala Lys Pro Ser Asn Asp Leu Ile Asn Val 465 470 475 480	1440
cgc tct agc acc ttt cac tat aag ggt aat gtg ctg ttg ctg ggt gat Arg Ser Ser Thr Phe His Tyr Lys Gly Asn Val Leu Leu Leu Gly Asp 485 490 495	1488
gct gcg cat gcg act gcg ccg ttc ctg ggt cag ggt atg aac atg gcg Ala Ala His Ala Thr Ala Pro Phe Leu Gly Gln Gly Met Asn Met Ala 500 505 510	1536
ctg gag gac gcc gcg acg ttt gtc gag ctg ctg gac cgc cac cag ggc Leu Glu Asp Ala Arg Thr Phe Val Glu Leu Leu Asp Arg His Gln Gly 515 520 525	1584
gac caa gac aaa gcc ttt ccg gag ttc acg gag ctg cgc aaa gtc cag Asp Gln Asp Lys Ala Phe Pro Glu Phe Thr Glu Leu Arg Lys Val Gln 530 535 540	1632

gca gac gca atg caa gac atg gct cgc gcc aac tat gac gtt ttg agc 1680
 Ala Asp Ala Met Gln Asp Met Ala Arg Ala Asn Tyr Asp Val Leu Ser 560
 545 550 555
 tgc tcg aac ccg atc ttt ttc atg cgt gcg cgt tac acg cgt tac atg 1728
 Cys Ser Asn Pro Ile Phe Phe Met Arg Ala Arg Tyr Thr Arg Tyr Met 575
 565 570
 cat tcc aag ttt ccg gcc ctg tat ccg ccg gat atg gcc gag aaa ctg 1776
 His Ser Lys Phe Pro Gly Leu Tyr Pro Pro Asp Met Ala Glu Lys Leu 590
 580 585
 tac ttt acg agc gag ccg tac gat cgt ctg caa caa atc cag cgt aaa 1824
 Tyr Phe Thr Ser Glu Pro Tyr Asp Arg Leu Gln Gln Ile Gln Arg Lys 605
 595 600
 cag aat gtt tgg tac aag att ggt cgc gtg aat tga 1860
 Gln Asn Val Trp Tyr Lys Ile Gly Arg Val Asn 610 615

<210> 70
 <211> 619
 <212> PRT
 <213> künstliche Sequenz
 <220>
 <223> synthetisches Konstrukt
 <400> 70

Met Lys Arg Glu Pro Glu Ser Val Tyr Glu Thr Asp Cys Arg Trp Asp
 1 5 10 15
 Gly Cys Ser Gln Glu Phe Asp Ser Gln Glu Gln Leu Val His His Ile
 20 25 30
 Asn Ser Glu His Ile His Gly Glu Arg Lys Glu Phe Val Cys His Trp
 35 40 45
 Gly Gly Cys Ser Arg Glu Leu Arg Pro Phe Lys Ala Gln Tyr Met Leu
 50 55 60
 Val Val His Met Arg Arg His Thr Gly Glu Lys Pro His Lys Cys Thr
 65 70 75 80
 Phe Glu Gly Cys Arg Lys Ser Tyr Ser Arg Leu Glu Asn Leu Lys Thr
 85 90 95
 His Leu Arg Ser His Thr Gly Glu Lys Pro Tyr Met Cys Glu His Glu
 100 105 110
 Gly Cys Ser Lys Ala Phe Ser Asn Ala Ser Asp Arg Ala Lys His Gln
 115 120 125
 Asn Arg Thr His Ser Asn Glu Lys Pro Tyr Val Cys Lys Leu Pro Gly
 130 135 140
 Cys Thr Lys Arg Tyr Thr Asp Pro Ser Ser Leu Arg Lys His Val Lys
 145 150 155 160

66

Thr Val His Gly Pro Asp Ala His Val Thr Lys Arg His Arg Gly Asp
165 170 175

Thr Arg Gly Gly Ser Gly Gly Gly Ser Gly Gly Ser Thr Arg Met Lys
180 185 190

Arg Ala Ile Ile Val Gly Gly Gly Leu Ala Gly Gly Leu Thr Ala Ile
195 200 205

Tyr Leu Ala Lys Arg Gly Tyr Glu Val His Val Val Glu Lys Arg Gly
210 215 220

Asp Pro Leu Arg Asp Leu Ser Ser Tyr Val Asp Val Val Ser Ser Arg
225 230 235 240

Ala Ile Gly Val Ser Met Thr Val Arg Gly Ile Lys Ser Val Leu Ala
245 250 255

Ala Gly Ile Pro Arg Ala Glu Leu Asp Ala Cys Gly Glu Pro Ile Val
260 265 270

Ala Met Ala Phe Ser Val Gly Gly Gln Tyr Arg Met Arg Glu Leu Lys
275 280 285

Pro Leu Glu Asp Phe Arg Pro Leu Ser Leu Asn Arg Ala Ala Phe Gln
290 295 300

Lys Leu Leu Asn Lys Tyr Ala Asn Leu Ala Gly Val Arg Tyr Tyr Phe
305 310 315 320

Glu His Lys Cys Leu Asp Val Asp Leu Asp Gly Lys Ser Val Leu Ile
325 330 335

Gln Gly Lys Asp Gly Gln Pro Gln Arg Leu Gln Gly Asp Met Ile Ile
340 345 350

Gly Ala Asp Gly Ala His Ser Ala Val Arg Gln Ala Met Gln Ser Gly
355 360 365

Leu Arg Arg Phe Glu Phe Gln Gln Thr Phe Phe Arg His Gly Tyr Lys
370 375 380

Thr Leu Val Leu Pro Asp Ala Gln Ala Leu Gly Tyr Arg Lys Asp Thr
385 390 395 400

Leu Tyr Phe Phe Gly Met Asp Ser Gly Gly Leu Phe Ala Gly Arg Ala
405 410 415

Ala Thr Ile Pro Asp Gly Ser Val Ser Ile Ala Val Cys Leu Pro Tyr
420 425 430

Ser	Gly	Ser	Pro	Ser	Leu	Thr	Thr	Asp	Glu	Pro	Thr	Met	Arg	Ala
		435					440					445		
Phe	Phe	Asp	Arg	Tyr	Phe	Gly	Gly	Leu	Pro	Arg	Asp	Ala	Arg	Asp
	450					455					460			Glu
Met	Leu	Arg	Gln	Phe	Leu	Ala	Lys	Pro	Ser	Asn	Asp	Leu	Ile	Asn
465					470					475				480
Arg	Ser	Ser	Thr	Phe	His	Tyr	Lys	Gly	Asn	Val	Leu	Leu	Leu	Gly
				485					490					495
Ala	Ala	His	Ala	Thr	Ala	Pro	Phe	Leu	Gly	Gln	Gly	Met	Asn	Met
			500					505					510	Ala
Leu	Glu	Asp	Ala	Arg	Thr	Phe	Val	Glu	Leu	Leu	Asp	Arg	His	Gln
		515					520					525		Gly
Asp	Gln	Asp	Lys	Ala	Phe	Pro	Glu	Phe	Thr	Glu	Leu	Arg	Lys	Val
	530					535					540			Gln
Ala	Asp	Ala	Met	Gln	Asp	Met	Ala	Arg	Ala	Asn	Tyr	Asp	Val	Ser
545					550					555				560
Cys	Ser	Asn	Pro	Ile	Phe	Phe	Met	Arg	Ala	Arg	Tyr	Thr	Arg	Tyr
				565					570					575
His	Ser	Lys	Phe	Pro	Gly	Leu	Tyr	Pro	Pro	Asp	Met	Ala	Glu	Lys
			580					585					590	Leu
Tyr	Phe	Thr	Ser	Glu	Pro	Tyr	Asp	Arg	Leu	Gln	Gln	Ile	Gln	Arg
		595					600					605		Lys
Gln	Asn	Val	Trp	Tyr	Lys	Ile	Gly	Arg	Val	Asn				
	610					615								

```
<220>
<221> CDS
<222> (1)..(1113)
<223> His_tag-Znf_Gli1-Linkerpeptid-nCFP
```

92 / 116

68															
Val	Tyr	Glu	Thr	Asp	Cys	Arg	Trp	Asp	Gly	Cys	Ser	Gln	Glu	Phe	Asp
20			20					25					30		
tcc	caa	gag	cag	ctg	gtg	cac	cac	atc	aac	agc	gag	cac	atc	cac	ggg
Ser	Gln	Glu	Gln	Leu	Val	His	His	Ile	Asn	Ser	Glu	His	Ile	His	Gly
		35					40					45			
gag	cgg	aag	gag	ttc	gtg	tgc	cac	tgg	ggg	ggc	tgc	tcc	agg	gag	ctg
Glu	Arg	Lys	Glu	Phe	Val	Cys	His	Trp	Gly	Gly	Cys	Ser	Arg	Glu	Leu
	50					55					60				
agg	ccc	ttc	aaa	gcc	cag	tac	atg	ctg	gtg	gtt	cac	atg	cgc	aga	cac
Arg	Pro	Phe	Lys	Ala	Pro	Tyr	Met	Leu	Val	Val	His	Met	Arg	Arg	His
	65				70					75					80
act	ggc	gag	aag	cca	cac	aag	tgc	acg	ttt	gaa	ggg	tgc	cgg	aag	tca
Thr	Gly	Glu	Lys	Pro	His	Lys	Cys	Thr	Phe	Glu	Gly	Cys	Arg	Lys	Ser
				85					90					95	
tac	tca	cgc	ctc	gaa	aac	ctg	aag	acg	cac	ctg	cgg	tca	cac	acg	ggt
Tyr	Ser	Arg	Leu	Glu	Asn	Leu	Lys	Thr	His	Leu	Arg	Ser	His	Thr	Gly
			100					105					110		
gag	aag	cca	tac	atg	tgt	gag	cac	gag	ggc	tgt	agt	aaa	gcc	ttc	agc
Glu	Lys	Pro	Tyr	Met	Cys	Glu	His	Glu	Gly	Cys	Ser	Lys	Ala	Phe	Ser
		115					120					125			
aat	gcc	agt	gac	cga	gcc	aag	cac	cag	aat	cgg	acc	cat	tcc	aat	gag
Asn	Ala	Ser	Asp	Arg	Ala	Lys	His	Gln	Asn	Arg	Thr	His	Ser	Asn	Glu
	130					135					140				
aag	ccg	tat	gta	tgt	aag	ctc	cct	ggc	tgc	acc	aaa	cgc	tat	aca	gat
Lys	Pro	Tyr	Val	Cys	Lys	Leu	Pro	Gly	Cys	Thr	Lys	Arg	Tyr	Thr	Asp
	145				150					155					160
cct	agc	tgc	ctg	cga	aaa	cat	gtc	aag	aca	gtg	cat	ggg	cct	gac	gcc
Pro	Ser	Ser	Leu	Arg	Lys	His	Val	Lys	Thr	Val	His	Gly	Pro	Asp	Ala
			165					170					175		
cat	gtg	acc	aaa	cgg	cac	cgt	ggg	gat	act	aga	ggc	ggg	tca	ggc	ggg
His	Val	Thr	Lys	Arg	His	Arg	Gly	Asp	Thr	Arg	Gly	Gly	Ser	Gly	Gly
			180				185						190		
ggc	tca	ggc	ggg	tca	act	aga	gtg	agc	aag	ggc	gag	gag	ctg	ttc	acc
Gly	Ser	Gly	Gly	Ser	Thr	Arg	Val	Ser	Lys	Gly	Glu	Glu	Leu	Phe	Thr
		195					200					205			
ggg	gtg	gtg	ccc	atc	ctg	gtc	gag	ctg	gac	ggc	gac	gta	aac	ggc	cac
Gly	Val	Val	Pro	Ile	Leu	Val	Glu	Leu	Asp	Gly	Asp	Val	Asn	Gly	His
	210					215					220				
aag	ttc	agc	gtg	tcc	ggc	gag	ggc	gag	ggc	gat	gcc	acc	tac	ggc	aag
Lys	Phe	Ser	Val	Ser	Gly	Glu	Gly	Glu	Gly	Asp	Ala	Thr	Tyr	Gly	Lys
	225				230					235				240	
ctg	acc	ctg	aag	ttc	atc	tgc	acc	acc	ggc	aag	ctg	ccc	gtg	ccc	tgg
Leu	Thr	Leu	Lys	Phe	Ile	Cys	Thr	Thr	Gly	Lys	Leu	Pro	Val	Pro	Trp
				245					250				255		
ccc	acc	ctc	gtg	acc	acc	ctg	acc	tgg	ggc	gtg	cag	tgc	ttc	gcc	cgc
Pro	Thr	Leu	Val	Thr	Thr	Leu	Thr	Trp	Gly	Val	Gln	Cys	Phe	Ala	Arg
			260					265					270		
tac	ccc	gac	cac	atg	aag	cag	cac	gac	ttc	ttc	aag	tcc	gcc	atg	ccc
Tyr	Pro	Asp	His	Met	Lys	Gln	His	Asp	Phe	Phe	Lys	Ser	Ala	Met	Pro
		275					280					285			
gaa	ggc	tac	gtc	cag	gag	cgc	acc	atc	ttc	ttc	aag	gac	gac	ggc	aac
Glu	Gly	Tyr	Val	Gln	Glu	Arg	Thr	Ile	Phe	Phe	Lys	Asp	Asp	Gly	Asn

290	295	69	300	
tac aag acc cgc gcc gag gtg aag ttc gag ggc gac acc ctg gtg aac				960
Tyr Lys Thr Arg Ala Glu Val Lys Phe Glu Gly Asp Thr Leu Val Asn				
305	310		315	320
cgc atc gag ctg aag ggc atc gac ttc aag gag gac ggc aac atc ctg				1008
Arg Ile Glu Leu Lys Gly Ile Asp Phe Lys Glu Asp Gly Asn Ile Leu				
	325		330	335
ggg cac aag ctg gag tac aac gcc atc agc gac aac gtc tat atc acc				1056
Gly His Lys Leu Glu Tyr Asn Ala Ile Ser Asp Asn Val Tyr Ile Thr				
	340		345	350
gcc gac aag cag aag aac ggc atc aag gcc aac ttc aag atc cgc cac				1104
Ala Asp Lys Gln Lys Asn Gly Ile Lys Ala Asn Phe Lys Ile Arg His				
	355		360	365
aac atc taa				1113
Asn Ile				
370				

<210> 72
 <211> 370
 <212> PRT
 <213> künstliche Sequenz
 <220>
 <223> synthetisches Konstrukt
 <400> 72

Met Tyr His His His His His His Thr Arg Lys Arg Glu Pro Glu Ser
 1 5 10 15

Val Tyr Glu Thr Asp Cys Arg Trp Asp Gly Cys Ser Gln Glu Phe Asp
 20 25 30

Ser Gln Glu Gln Leu Val His His Ile Asn Ser Glu His Ile His Gly
 35 40 45

Glu Arg Lys Glu Phe Val Cys His Trp Gly Gly Cys Ser Arg Glu Leu
 50 55 60

Arg Pro Phe Lys Ala Gln Tyr Met Leu Val Val His Met Arg Arg His
 65 70 75 80

Thr Gly Glu Lys Pro His Lys Cys Thr Phe Glu Gly Cys Arg Lys Ser
 85 90 95

Tyr Ser Arg Leu Glu Asn Leu Lys Thr His Leu Arg Ser His Thr Gly
 100 105 110

Glu Lys Pro Tyr Met Cys Glu His Glu Gly Cys Ser Lys Ala Phe Ser
 115 120 125

Asn Ala Ser Asp Arg Ala Lys His Gln Asn Arg Thr His Ser Asn Glu
 130 135 140

Lys Pro Tyr Val Cys Lys Leu Pro Gly Cys Thr Lys Arg Tyr Thr Asp
 145 150 155 160
 Pro Ser Ser Leu Arg Lys His Val Lys Thr Val His Gly Pro Asp Ala
 165 170 175
 His Val Thr Lys Arg His Arg Gly Asp Thr Arg Gly Gly Ser Gly Gly
 180 185 190
 Gly Ser Gly Gly Ser Thr Arg Val Ser Lys Gly Glu Glu Leu Phe Thr
 195 200 205
 Gly Val Val Pro Ile Leu Val Glu Leu Asp Gly Asp Val Asn Gly His
 210 215 220
 Lys Phe Ser Val Ser Gly Glu Gly Glu Gly Asp Ala Thr Tyr Gly Lys
 225 230 235 240
 Leu Thr Leu Lys Phe Ile Cys Thr Thr Gly Lys Leu Pro Val Pro Trp
 245 250 255
 Pro Thr Leu Val Thr Thr Leu Thr Trp Gly Val Gln Cys Phe Ala Arg
 260 265 270
 Tyr Pro Asp His Met Lys Gln His Asp Phe Phe Lys Ser Ala Met Pro
 275 280 285
 Glu Gly Tyr Val Gln Glu Arg Thr Ile Phe Phe Lys Asp Asp Gly Asn
 290 295 300
 Tyr Lys Thr Arg Ala Glu Val Lys Phe Glu Gly Asp Thr Leu Val Asn
 305 310 315 320
 Arg Ile Glu Leu Lys Gly Ile Asp Phe Lys Glu Asp Gly Asn Ile Leu
 325 330 335
 Gly His Lys Leu Glu Tyr Asn Ala Ile Ser Asp Asn Val Tyr Ile Thr
 340 345 350
 Ala Asp Lys Gln Lys Asn Gly Ile Lys Ala Asn Phe Lys Ile Arg His
 355 360 365
 Asn Ile
 370

<210> 73
 <211> 864
 <212> DNA
 <213> künstliche Sequenz
 <220>
 <223> His_tag-Znf_Blues-Linkerpeptid-nCFP

71

```

<220>
<221> CDS
<222> (1)..(864)
<223> His_tag-Znf_Blues-Linkerpeptid-nCFP

<400> 73
atg tat cat cac cat cac cat cac act aga gca agt gat gat cgt ccg      48
Met Tyr His His His His His His Thr Arg Ala Ser Asp Asp Arg Pro
1      5      10      15

tat gca tgt ccg gtt gaa agc tgt gat cgt cgt ttt agc cgt cgt gat      96
Tyr Ala Cys Pro Val Glu Ser Cys Asp Arg Arg Phe Ser Arg Arg Asp
20      25      30

gtt ctg atg aat cac att cgt att cac aca ggt cag aaa ccg ttt cag      144
Val Leu Met Asn His Ile Arg Ile His Thr Gly Gln Lys Pro Phe Gln
35      40      45

tgt cgt att tgc atg cgt aat ttt agc cgt agc gat cat ctg acc acc      192
Cys Arg Ile Cys Met Arg Asn Phe Ser Arg Ser Asp His Leu Thr Thr
50      55      60

cat att cgt acc cat aca ggt gaa aaa ccg ttt gca tgt gat att tgc      240
His Ile Arg Thr His Ala Gly Glu Lys Pro Phe Ala Cys Asp Ile Cys
65      70      75      80

ggg cgc aaa ttt gca aat cgt gat acc ctg acc cgt cat agc aaa att      288
Gly Arg Lys Phe Ala Asn Arg Asp Thr Leu Thr Arg His Ser Lys Ile
85      90      95

cat ctg cgt cag aat gat ctg gaa ggt ggt agc ggt ggt ggt tca ggt      336
His Leu Arg Gln Asn Asp Leu Glu Gly Gly Ser Gly Gly Gly Ser Gly
100      105      110

ggg agt act aga gtg agc aag ggc gag gag ctg ttc acc ggg gtg gtg      384
Gly Ser Thr Arg Val Ser Lys Gly Glu Glu Leu Phe Thr Gly Val Val
115      120      125

ccc atc ctg gtc gag ctg gac ggc gac gta aac ggc cac aag ttc agc      432
Pro Ile Leu Val Glu Leu Asp Gly Asp Val Asn Gly His Lys Phe Ser
130      135      140

gtg tcc ggc gag ggc gag ggc gat gcc acc tac ggc aag ctg acc ctg      480
Val Ser Gly Glu Gly Glu Gly Asp Ala Thr Tyr Gly Lys Leu Thr Leu
145      150      155      160

aag ttc atc tgc acc acc ggc aag ctg ccc gtg ccc tgg ccc acc ctc      528
Lys Phe Ile Cys Thr Thr Gly Lys Leu Pro Val Pro Trp Pro Thr Leu
165      170      175

gtg acc acc ctg acc tgg ggc gtg cag tgc ttc gcc cgc tac ccc gac      576
Val Thr Thr Leu Thr Trp Gly Val Gln Cys Phe Ala Arg Tyr Pro Asp
180      185      190

cac atg aag cag cac gac ttc ttc aag tcc gcc atg ccc gaa ggc tac      624
His Met Lys Gln His Asp Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr
195      200      205

gtc cag gag cgc acc atc ttc ttc aag gac gac ggc aac tac aag acc      672
Val Gln Glu Arg Thr Ile Phe Phe Lys Asp Asp Gly Asn Tyr Lys Thr
210      215      220

cgc gcc gag gtg aag ttc gag ggc gac acc ctg gtg aac cgc atc gag      720
Arg Ala Glu Val Lys Phe Glu Gly Asp Thr Leu Val Asn Arg Ile Glu
225      230      235      240

ctg aag ggc atc gac ttc aag gag gac ggc aac atc ctg ggg cac aag      768
Leu Lys Gly Ile Asp Phe Lys Glu Asp Gly Asn Ile Leu Gly His Lys
245      250      255

```

72

ctg gag tac aac gcc atc agc gac aac gtc tat atc acc gcc gac aag 816
Leu Glu Tyr Asn Ala Ile Ser Asp Asn Val Tyr Ile Thr Ala Asp Lys
260 265 270

cag aag aac ggc atc aag gcc aac ttc aag atc cgc cac aac atc taa 864
Gln Lys Asn Gly Ile Lys Ala Asn Phe Lys Ile Arg His Asn Ile
275 280 285

<210> 74
<211> 287
<212> PRT
<213> künstliche Sequenz

<220>
<223> synthetisches Konstrukt

<400> 74

Met Tyr His His His His His His Thr Arg Ala Ser Asp Asp Arg Pro
1 5 10 15

Tyr Ala Cys Pro Val Glu Ser Cys Asp Arg Arg Phe Ser Arg Arg Asp
20 25 30

Val Leu Met Asn His Ile Arg Ile His Thr Gly Gln Lys Pro Phe Gln
35 40 45

Cys Arg Ile Cys Met Arg Asn Phe Ser Arg Ser Asp His Leu Thr Thr
50 55 60

His Ile Arg Thr His Thr Gly Glu Lys Pro Phe Ala Cys Asp Ile Cys
65 70 75 80

Gly Arg Lys Phe Ala Asn Arg Asp Thr Leu Thr Arg His Ser Lys Ile
85 90 95

His Leu Arg Gln Asn Asp Leu Glu Gly Gly Ser Gly Gly Gly Ser Gly
100 105 110

Gly Ser Thr Arg Val Ser Lys Gly Glu Glu Leu Phe Thr Gly Val Val
115 120 125

Pro Ile Leu Val Glu Leu Asp Gly Asp Val Asn Gly His Lys Phe Ser
130 135 140

Val Ser Gly Glu Gly Glu Gly Asp Ala Thr Tyr Gly Lys Leu Thr Leu
145 150 155 160

Lys Phe Ile Cys Thr Thr Gly Lys Leu Pro Val Pro Trp Pro Thr Leu
165 170 175

Val Thr Thr Leu Thr Trp Gly Val Gln Cys Phe Ala Arg Tyr Pro Asp
180 185 190

His Met Lys Gln His Asp Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr

195 200 73 205

Val Gln Glu Arg Thr Ile Phe Phe Lys Asp Asp Gly Asn Tyr Lys Thr
210 215 220

Arg Ala Glu Val Lys Phe Glu Gly Asp Thr Leu Val Asn Arg Ile Glu
225 230 235 240

Leu Lys Gly Ile Asp Phe Lys Glu Asp Gly Asn Ile Leu Gly His Lys
245 250 255

Leu Glu Tyr Asn Ala Ile Ser Asp Asn Val Tyr Ile Thr Ala Asp Lys
260 265 270

Gln Lys Asn Gly Ile Lys Ala Asn Phe Lys Ile Arg His Asn Ile
275 280 285

<210> 75
<211> 840
<212> DNA
<213> künstliche Sequenz

<220>
<223> His_tag-Znf_PBSII_Linkерpeptid-nYFP

<220>
<221> CDS
<222> (1)..(840)
<223> His_tag-Znf_PBSII_Linkерpeptid-nYFP

<400> 75

atg	tat	cat	cac	cat	cac	cat	cac	act	aga	ccg	ggt	gaa	aaa	ccg	tat	48
Met	Tyr	His	His	His	His	His	His	Thr	Arg	Pro	Gly	Glu	Lys	Pro	Tyr	
1				5				10						15		

gca	tgt	cct	gaa	tgt	ggt	aaa	agc	ttt	agc	cag	cgt	gca	aat	ctg	cgt	96
Ala	Cys	Pro	Glu	Cys	Gly	Lys	Ser	Phe	Ser	Gln	Arg	Ala	Asn	Leu	Arg	
			20					25					30			

gca	cat	cag	cgt	acc	cat	aca	ggt	gaa	aag	cct	tat	aaa	tgc	cca	gaa	144
Ala	His	Gln	Arg	Thr	His	Thr	Gly	Glu	Lys	Pro	Tyr	Lys	Cys	Pro	Glu	
			35				40					45				

tgc	ggc	aaa	agc	ttt	tca	cgt	agc	gat	cat	ctg	acc	acc	cat	cag	cgc	192
Cys	Gly	Lys	Ser	Phe	Ser	Arg	Ser	Asp	His	Leu	Thr	Thr	His	Gln	Arg	
	50					55					60					

aca	cat	act	ggc	gag	aaa	cct	tac	aaa	tgt	cca	gag	tgt	ggt	aaa	tca	240
Thr	His	Thr	Gly	Glu	Lys	Pro	Tyr	Lys	Cys	Pro	Glu	Cys	Gly	Lys	Ser	
					70					75					80	

ttt	agc	cgt	agt	gat	gtt	ctg	gtt	cgt	cat	cag	aga	acc	cac	acg	ggt	288
Phe	Ser	Arg	Ser	Asp	Val	Leu	Val	Arg	His	Gln	Arg	Thr	His	Thr	Gly	
				85					90					95		

ggt	ggt	agc	ggt	ggt	ggt	tca	ggt	ggt	agt	act	aga	gtg	agc	aag	ggc	336
Gly	Gly	Ser	Gly	Gly	Gly	Ser	Gly	Gly	Ser	Thr	Arg	Val	Ser	Lys	Gly	
			100					105					110			

gag	gag	ctg	ttc	acc	ggg	gtg	gtg	ccc	atc	ctg	gtc	gag	ctg	gac	ggc	384
Glu	Glu	Leu	Phe	Thr	Gly	Val	Val	Pro	Ile	Leu	Val	Glu	Leu	Asp	Gly	
		115					120					125				

74

gac gta aac ggc cac aag ttc agc gtg tcc ggc gag ggc gag ggc gat Asp Val Asn Gly His Lys Phe Ser Val Ser Gly Glu Gly Glu Gly Asp 130 135 140	432
gcc acc tac ggc aag ctg acc ctg aag ttc atc tgc acc acc ggc aag Ala Thr Tyr Gly Lys Leu Thr Leu Lys Phe Ile Cys Thr Thr Gly Lys 145 150 155 160	480
ctg ccc gtg ccc tgg ccc acc ctc gtg acc acc ttc ggc tac ggc ctg Leu Pro Val Pro Trp Pro Thr Leu Val Thr Thr Phe Gly Tyr Gly Leu 165 170 175	528
atg tgc ttc gcc cgc tac ccc gac cac atg aag cag cac gac ttc ttc Met Cys Phe Ala Arg Tyr Pro Asp His Met Lys Gln His Asp Phe Phe 180 185 190	576
aag tcc gcc atg ccc gaa ggc tac gtc cag gag cgc acc atc ttc ttc Lys Ser Ala Met Pro Glu Gly Tyr Val Gln Glu Arg Thr Ile Phe Phe 195 200 205	624
aag gac gac ggc aac tac aag acc cgc gcc gag gtg aag ttc gag ggc Lys Asp Asp Gly Asn Tyr Lys Thr Arg Ala Glu Val Lys Phe Glu Gly 210 215 220	672
gac acc ctg gtg aac cgc atc gag ctg aag ggc atc gac ttc aag gag Asp Thr Leu Val Asn Arg Ile Glu Leu Lys Gly Ile Asp Phe Lys Glu 225 230 235 240	720
gac ggc aac atc ctg ggg cac aag ctg gag tac aac tac aac agc cac Asp Gly Asn Ile Leu Gly His Lys Leu Glu Tyr Asn Tyr Asn Ser His 245 250 255	768
aac gtc tat atc atg gcc gac aag cag aag aac ggc atc aag gtg aac Asn Val Tyr Ile Met Ala Asp Lys Gln Lys Asn Gly Ile Lys Val Asn 260 265 270	816
ttc aag atc cgc cac aac atc taa Phe Lys Ile Arg His Asn Ile 275	840

<210> 76
 <211> 279
 <212> PRT
 <213> künstliche Sequenz

<220>
 <223> synthetisches Konstrukt

<400> 76

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

```

<210> 77
<211> 573
<212> DNA
<213> künstliche Sequenz

<220>
<223> cCFP-Linkerpeptid-Znf_HIVC-His_tag

<220>
<221> CDS
<222> (1)..(573)
<223> cCFP-Linkerpeptid-Znf_HIVC-His_tag

```

76

<400> 77
 atg gcc gac aag cag aag aac ggc atc aag gcc aac ttc aag atc cgc 48
 Met Ala Asp Lys Gln Lys Asn Gly Ile Lys Ala Asn Phe Lys Ile Arg
 1 5 10 15
 cac aac atc gag gac ggc agc gtg cag ctc gcc gac cac tac cag cag 96
 His Asn Ile Glu Asp Gly Ser Val Gln Leu Ala Asp His Tyr Gln Gln
 20 25 30
 aac acc ccc atc ggc gac ggc ccc gtg ctg ctg ccc gac aac cac tac 144
 Asn Thr Pro Ile Gly Asp Gly Pro Val Leu Leu Pro Asp Asn His Tyr
 35 40 45
 ctg agc acc cag tcc aag ctg agc aaa gac ccc aac gag aag cgc gat 192
 Leu Ser Thr Gln Ser Lys Leu Ser Lys Asp Pro Asn Glu Lys Arg Asp
 50 55 60
 cac atg gtc ctg ctg gag ttc gtg acc gcc gcc ggg atc act ctc ggc 240
 His Met Val Leu Leu Glu Phe Val Thr Ala Ala Gly Ile Thr Leu Gly
 65 70 75 80
 atg gac gag ctg tac aag ggt ggc tct ggc ggt ggg agt ggg gga agc 288
 Met Asp Glu Leu Tyr Lys Gly Gly Ser Gly Gly Gly Ser Gly Gly Ser
 85 90 95
 act aga ccc ttc cag tgt cgc att tgc atg cgg aac ttt tgc ctc agg 336
 Thr Arg Pro Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser Leu Arg
 100 105 110
 acg gac ctt gac agg cat acc cgt act cat acc ggt gaa aaa ccg ttt 384
 Thr Asp Leu Asp Arg His Thr Arg Thr His Thr Gly Glu Lys Pro Phe
 115 120 125
 cag tgt cgg atc tgt atg cga aat ttc tcc ctc agc cag acc ttg cgc 432
 Gln Cys Arg Ile Cys Met Arg Asn Phe Ser Leu Ser Gln Thr Leu Arg
 130 135 140
 cgc cat cta cgt acg cac acc ggc gag aag cca ttc caa tgc cga ata 480
 Arg His Leu Arg Thr His Thr Gly Glu Lys Pro Phe Gln Cys Arg Ile
 145 150 155 160
 tgc atg cgc aac ttc agt ctc agg agc aac ctg ggg cgg cac cta aaa 528
 Cys Met Arg Asn Phe Ser Leu Arg Ser Asn Leu Gly Arg His Leu Lys
 165 170 175
 acc cac aca gga gaa aaa act aga cat cac cat cac cat cac taa 573
 Thr His Thr Gly Glu Lys Thr Arg His His His His His His
 180 185 190

<210> 78
 <211> 190
 <212> PRT
 <213> künstliche Sequenz

<220>
 <223> synthetisches Konstrukt

<400> 78

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

35 40 77 45
 Leu Ser Thr Gln Ser Lys Leu Ser Lys Asp Pro Asn Glu Lys Arg Asp
 50 55 60
 His Met Val Leu Leu Glu Phe Val Thr Ala Ala Gly Ile Thr Leu Gly
 65 70 75 80
 Met Asp Glu Leu Tyr Lys Gly Gly Ser Gly Gly Gly Ser Gly Gly Ser
 85 90 95
 Thr Arg Pro Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser Leu Arg
 100 105 110
 Thr Asp Leu Asp Arg His Thr Arg Thr His Thr Gly Glu Lys Pro Phe
 115 120 125
 Gln Cys Arg Ile Cys Met Arg Asn Phe Ser Leu Ser Gln Thr Leu Arg
 130 135 140
 Arg His Leu Arg Thr His Thr Gly Glu Lys Pro Phe Gln Cys Arg Ile
 145 150 155 160
 Cys Met Arg Asn Phe Ser Leu Arg Ser Asn Leu Gly Arg His Leu Lys
 165 170 175
 Thr His Thr Gly Glu Lys Thr Arg His His His His His His
 180 185 190

<210> 79
 <211> 633
 <212> DNA
 <213> künstliche Sequenz

<220>
 <223> cCFP-Linkerpeptid-Znf_Jazz-Linkerpeptid-His_tag

<220>
 <221> CDS
 <222> (1)..(633)
 <223> cCFP-Linkerpeptid-Znf_Jazz-Linkerpeptid-His_tag

<400> 79
 atg gcc gac aag cag aag aac ggc atc aag gcc aac ttc aag atc cgc 48
 Met Ala Asp Lys Gln Lys Asn Gly Ile Lys Ala Asn Phe Lys Ile Arg
 1 5 10 15
 cac aac atc gag gac ggc agc gtg cag ctg gcc gac cac tac cag cag 96
 His Asn Ile Glu Asp Gly Ser Val Gln Leu Ala Asp His Tyr Gln Gln
 20 25 30
 aac acc ccc atc ggc gac ggc ccc gtg ctg ctg ccc gac aac cac tac 144
 Asn Thr Pro Ile Gly Asp Gly Pro Val Leu Leu Pro Asp Asn His Tyr
 35 40 45
 ctg agc acc cag tcc aag ctg agc aaa gac ccc aac gag aag cgc gat 192
 Leu Ser Thr Gln Ser Lys Leu Ser Lys Asp Pro Asn Glu Lys Arg Asp
 50 55 60

78

cac atg gtc ctg ctg gag ttc gtg acc gcc gcc ggg atc act ctc ggc	240
His Met Val Leu Leu Glu Phe Val Thr Ala Ala Gly Ile Thr Leu Gly	
65 70 75 80	
atg gac gag ctg tac aag ggt ggc tct ggc ggt ggg agt ggg gga agc	288
Met Asp Glu Leu Tyr Lys Gly Gly Ser Gly Gly Gly Ser Gly Gly Ser	
85 90 95	
act aga gca agt gat gat cgt ccg tat gca tgt ccg gtt gaa agc tgt	336
Thr Arg Ala Ser Asp Asp Arg Pro Tyr Ala Cys Pro Val Glu Ser Cys	
100 105 110	
gat cgt cgt ttt agc cgt agt gat gaa ctg acc cgt cat att cgt att	384
Asp Arg Arg Phe Ser Arg Ser Asp Glu Leu Thr Arg His Ile Arg Ile	
115 120 125	
cat aca ggt cag aaa cct ttt cag tgt cgt att tgc atg cgt aat ttt	432
His Thr Gly Gln Lys Pro Phe Gln Cys Arg Ile Cys Met Arg Asn Phe	
130 135 140	
agc tca aga gat gtt ctg cgt cgt cat aat cgt acc cat acg ggt gaa	480
Ser Ser Arg Asp Val Leu Arg Arg His Asn Arg Thr His Thr Gly Glu	
145 150 155 160	
aaa ccg ttt gca tgt gat att tgc ggt cgt aaa ttt gca agt cgt gac	528
Lys Pro Phe Ala Cys Asp Ile Cys Gly Arg Lys Phe Ala Ser Arg Asp	
165 170 175	
gtg ctg cgt cgc cat aac cgc att cat ctg cgt cag aat gat ctg gaa	576
Val Leu Arg Arg His Asn Arg Ile His Leu Arg Gln Asn Asp Leu Glu	
180 185 190	
ggg ggt agc ggt ggt ggt tca ggt ggt agt act aga cac cac cac cac	624
Gly Gly Ser Gly Gly Gly Ser Gly Gly Ser Thr Arg His His His His	
195 200 205	
cac cac tag	633
His His	
210	

<210> 80
 <211> 210
 <212> PRT
 <213> künstliche Sequenz

<220>
 <223> synthetisches Konstrukt

<400> 80

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

```
<210> 81
<211> 615
<212> DNA
<213> künstliche Sequenz
```

```
<220>
<221> CDS
<222> (1)..(615)
<223> cYFP-Linkerpeptid-znf_zif268-Linkerpeptid -His_ tag
```

<400>	81															
atg gcc gac aag cag aag aac ggc atc aag gtg aac ttc aag atc cgc																48
Met Ala Asp Lys Gln Lys Asn Gly Ile Lys Val Asn Phe Lys Ile Arg																
1				5					10					15		
cac aac atc gag gac ggc agc gtg cag ctc gcc gac cac tac cag cag																96
His Asn Ile Glu Asp Gly Ser Val Gln Leu Ala Asp His Tyr Gln Gln																
			20					25					30			
aac acc ccc atc ggc gac ggc ccc gtg ctg ctg ccc gac aac cac tac																144
Asn Thr Pro Ile Gly Asp Gly Pro Val Leu Leu Pro Asp Asn His Tyr																
		35					40					45				
ctg agc tac cag tcc gcc ctg agc aaa gac ccc aac gag aag cgc gat																192
Leu Ser Tyr Gln Ser Ala Leu Ser Lys Asp Pro Asn Glu Lys Arg Asp																
	50					55					60					

80

cac atg gtc ctg ctg gag ttc gtg acc gcc gcc ggg atc act ctc ggc His Met Val Leu Leu Glu Phe Val Thr Ala Ala Gly Ile Thr Leu Gly 65 70 75 80	240
atg gac gag ctg tac aag ggt ggc tct gcc ggt ggg agt ggg gga agc Met Asp Glu Leu Tyr Lys Gly Gly Ser Gly Gly Gly Ser Gly Gly Ser 85 90 95	288
act aga ccg ggt gaa aaa ccg tat gca tgt ccg gtt gaa agc tgt gat Thr Arg Pro Gly Glu Lys Pro Tyr Ala Cys Pro Val Glu Ser Cys Asp 100 105 110	336
cgt cgt ttt agc cgt agt gat gaa ctg acc cgt cat att cgt att cat Arg Arg Phe Ser Arg Ser Asp Glu Leu Thr Arg His Ile Arg Ile His 115 120 125	384
aca ggt cag aaa ccg ttt cag tgt cgt att tgc atg cgt aat ttt agc Thr Gly Gln Lys Pro Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser 130 135 140	432
cgt tca gat cat ctg acc acc cat att cgt acc cat act ggc gaa aaa Arg Ser Asp His Leu Thr Thr His Ile Arg Thr His Thr Gly Glu Lys 145 150 155 160	480
ccg ttt gcc tgt gat att tgt ggt cgt aaa ttt gca cgt tcc gat gaa Pro Phe Ala Cys Asp Ile Cys Gly Arg Lys Phe Ala Arg Ser Asp Glu 165 170 175	528
cgt aaa cgc cat acc aaa att cat acg ggt ggt ggt agc ggt ggt ggt Arg Lys Arg His Thr Lys Ile His Thr Gly Gly Gly Ser Gly Gly Gly 180 185 190	576
tca ggt ggt agt act aga cac cac cac cac cac tag Ser Gly Gly Ser Thr Arg His His His His His His 195 200	615

<210> 82
<211> 204
<212> PRT
<213> künstliche Sequenz

<220>
<223> synthetisches Konstrukt

<400> 82

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

85 90 81 95

Thr Arg Pro Gly Glu Lys Pro Tyr Ala Cys Pro Val Glu Ser Cys Asp
100 105 110

Arg Arg Phe Ser Arg Ser Asp Glu Leu Thr Arg His Ile Arg Ile His
115 120 125

Thr Gly Gln Lys Pro Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser
130 135 140

Arg Ser Asp His Leu Thr Thr His Ile Arg Thr His Thr Gly Glu Lys
145 150 155 160

Pro Phe Ala Cys Asp Ile Cys Gly Arg Lys Phe Ala Arg Ser Asp Glu
165 170 175

Arg Lys Arg His Thr Lys Ile His Thr Gly Gly Gly Ser Gly Gly Gly
180 185 190

Ser Gly Gly Ser Thr Arg His His His His His His
195 200

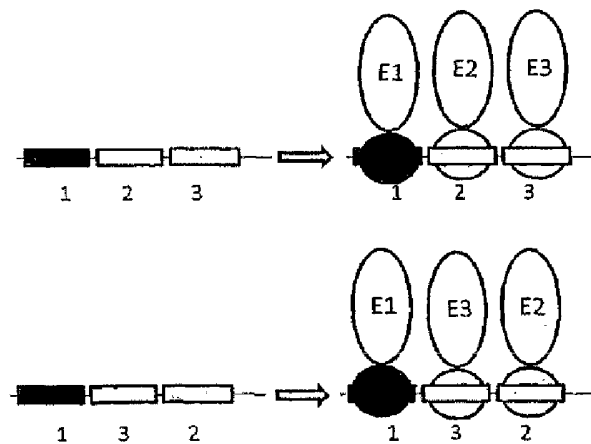
Patentansprüche

1. Verfahren zur Herstellung einer Verbindung über einen Biosyntheseweg durch das Kultivieren einer genetisch modifizierten Wirtszelle, wobei die Zelle mit einer oder mehreren Nukleinsäuren modifiziert ist, umfassend:
 - a) eine Nukleinsäuresequenz, die für mindestens drei bis etwa 100 chimäre Proteine kodiert, wobei sich jedes der chimären Proteine aus mindestens einem Enzym des Biosynthesewegs oder einem weiteren funktionellen Polypeptid sowie mindestens einem Nukleinsäure bindenden Faktor zusammensetzt, wobei das Enzym des Biosynthesewegs oder das weitere funktionelle Polypeptid und der Nukleinsäure bindende Faktor über einen Polypeptidlinker miteinander verbunden sind;
 - b) eine Programm-Nukleinsäuresequenz, eine Nukleinsäuresequenz, die mindestens zwischen drei und etwa 100 Zielnukleinsäureelemente enthält, von denen jedes durch einen anderen der Nukleinsäure bindenden Faktoren aus dem chimären Polypeptid erkannt wird, wobei die Programm-Nukleinsäuresequenz eine Anordnung von Enzymen des Biosynthesewegs oder der weiteren funktionellen Polypeptide entlang der Programm-Nukleinsäuresequenz bereitstellt; und
 - c) ein Substrat für den Biosyntheseweg, das entweder in der Wirtszelle vorliegt oder aus der extrazellulären Umgebung bereitgestellt wird,wobei die Kultivierung die Synthese der chimären Proteine in den genetisch modifizierten Wirtszellen ermöglicht und zur Herstellung der Verbindung führt, wobei das Verfahren weiter das Mischen einer oder mehrerer Verbindungen umfasst, wobei die Verbindungen umfasst sind von:
 - a) drei bis etwa 100 chimäre Proteine, wobei sich jedes der chimären Proteine aus mindestens einem Enzym des Biosynthesewegs oder einem weiteren funktionellen Polypeptid sowie mindestens einem Nukleinsäure bindenden Faktor zusammensetzt, wobei das Enzym des Biosynthesewegs oder das weitere funktionelle Polypeptid und der Nukleinsäure bindende Faktor über einen Polypeptidlinker miteinander verbunden sind;
 - b) eine Programm-Nukleinsäuresequenz, eine Nukleinsäuresequenz, die mindestens zwischen drei und etwa 100 Zielnukleinsäureelemente enthält, von denen jedes durch einen anderen der Nukleinsäure bindenden Faktoren aus dem chimären Polypeptid erkannt wird, wobei die Programm-Nukleinsäuresequenz eine Anordnung von Enzymen des Biosynthesewegs oder der weiteren funktionellen Polypeptide entlang der Programm-Nukleinsäuresequenz bereitstellt; und
 - c) ein Substrat für den Biosyntheseweg und Kofaktoren für Enzyme, die zu dem Gemisch gegeben oder in diesem erzeugt werden, wobei die Programm-Nukleinsäuresequenz entweder in Lösung oder immobilisiert an eine weitere Verbindung oder feste Phase vorliegen kann,charakterisiert dadurch, dass das Verfahren zur Herstellung einer Verbindung über einen Biosyntheseweg, ausgewählt aus der Gruppe bestehend aus dem Biosyntheseweg von Carotinoiden, von Violacein und von Resveratrol oder methyliertem Resveratrol, verwendet wird.
2. Verfahren nach Anspruch 1 zur Synthese von Carotinoiden.
3. Verfahren nach Anspruch 1 zur Synthese von Violacein.
4. Verfahren nach Anspruch 1 zur Synthese von Resveratrol oder methyliertem Resveratrol.
5. Verfahren nach einem der Ansprüche 1 bis 4, wobei die genetisch modifizierte Wirtszelle ausgewählt ist von prokaryotischer oder eukaryotischer Zelle.

Hierzu 9 Blatt Zeichnungen

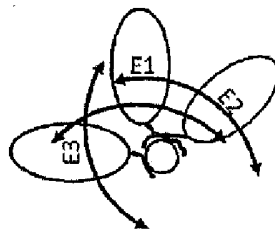
1/9

A



Die Nukleotidsequenz steuert die Anordnung der Enzyme des Biosynthesewegs durch die Interaktion zwischen Nukleotid-Bindungsdomänen und dem Nukleotidgerüst.

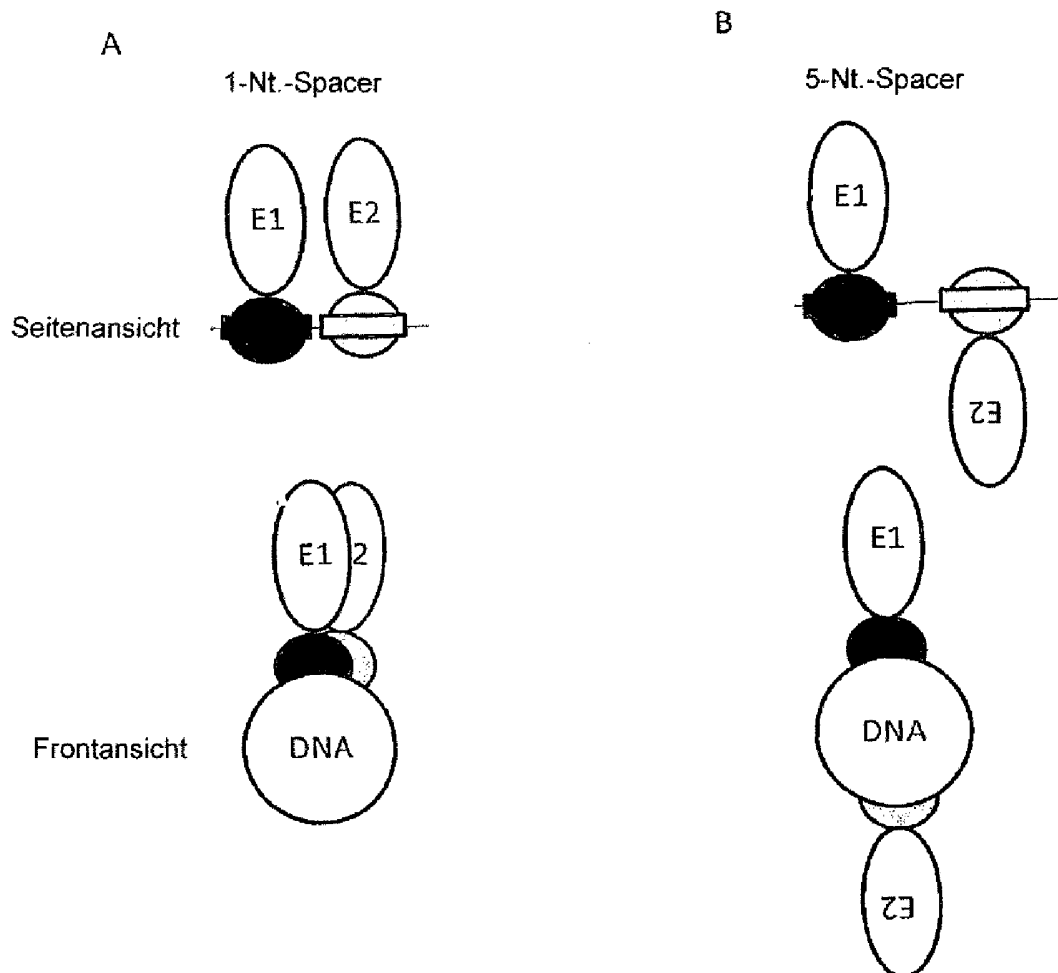
B



Ein Bündel von Enzymen des Biosynthesewegs, durch Peptide mit dem Polypeptidgerüst verbunden.

Figur 1.

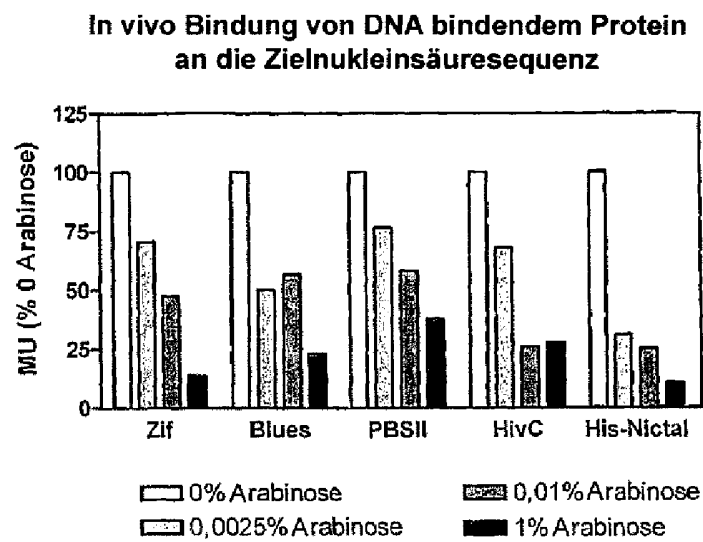
2/9



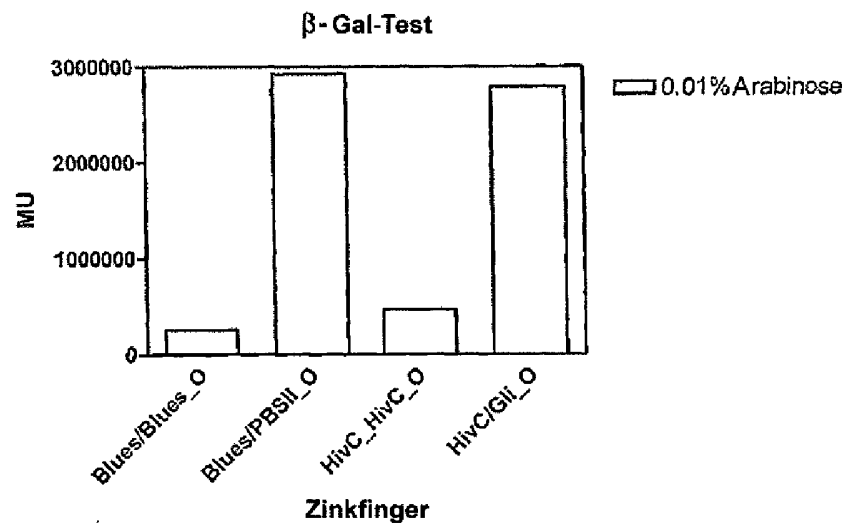
Figur 2.

3/9

A

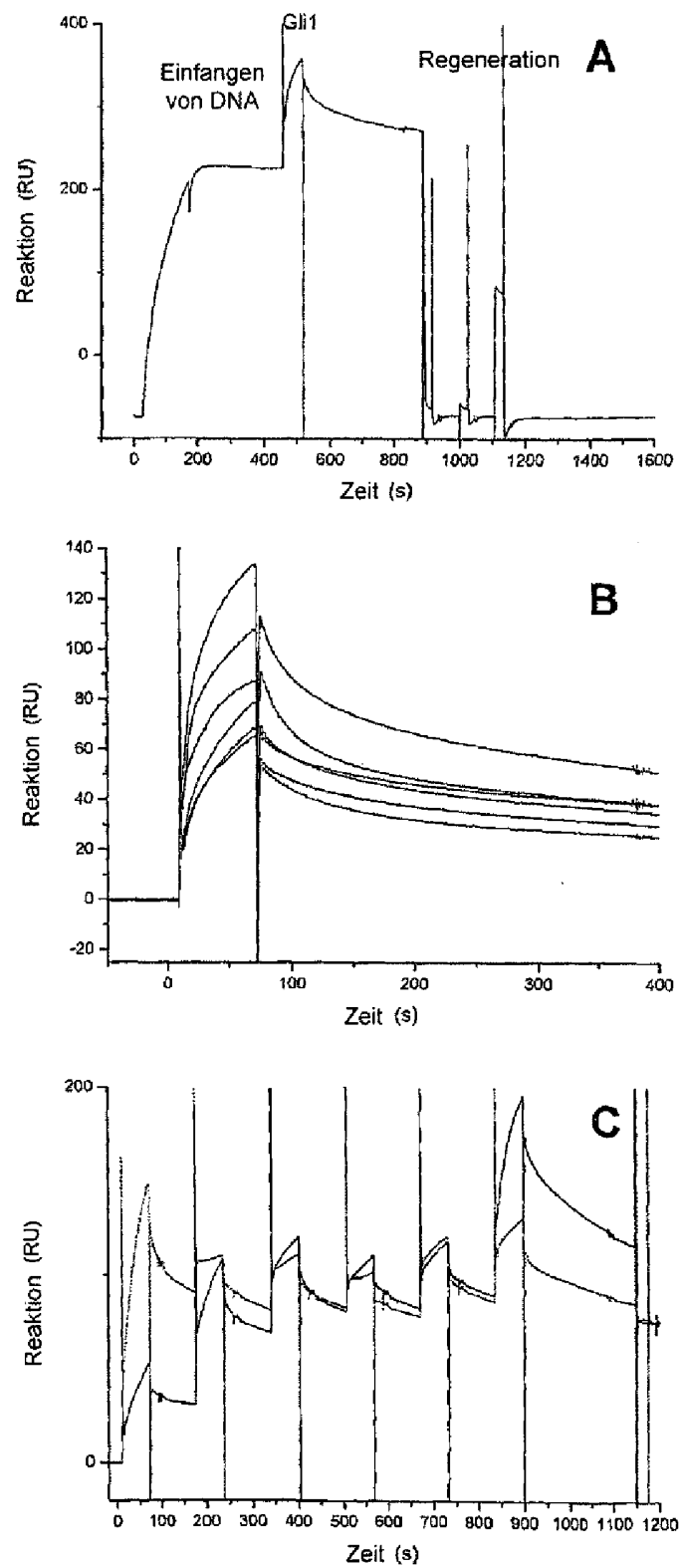


B



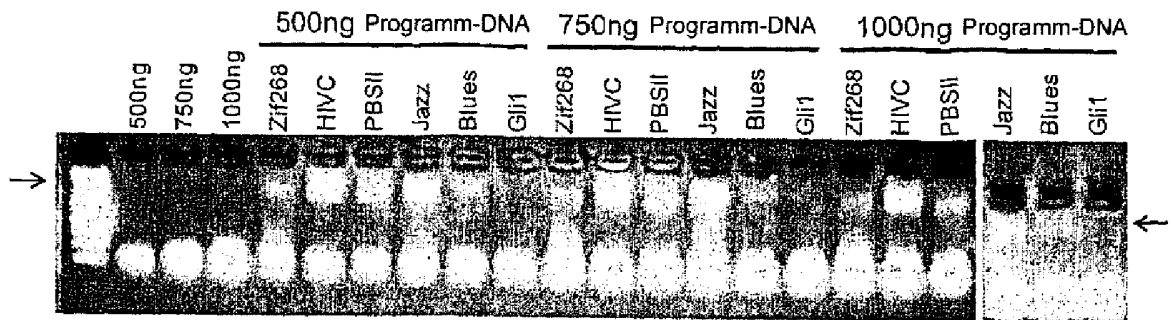
Figur 3.

4/9



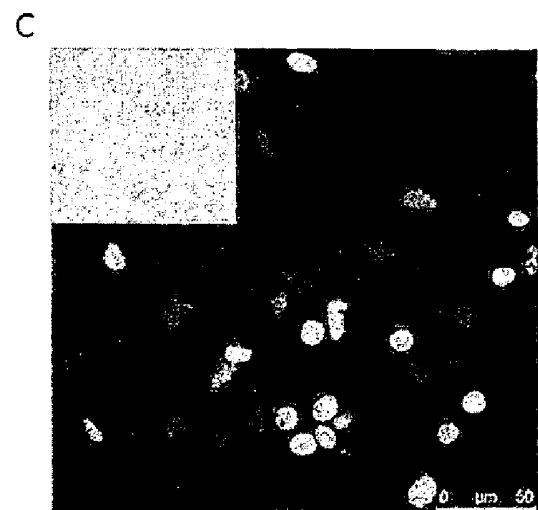
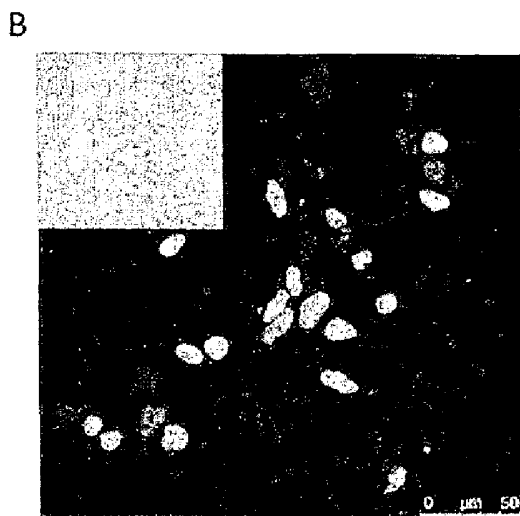
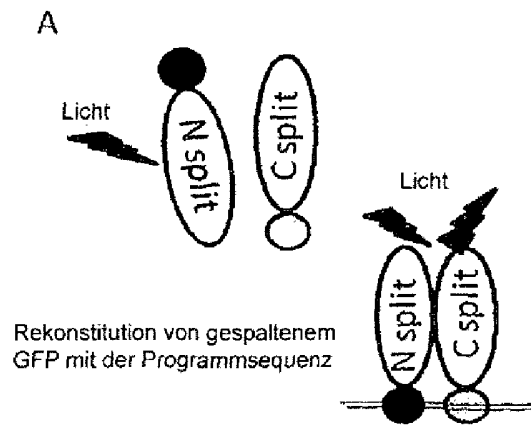
Figur 4.

5/9



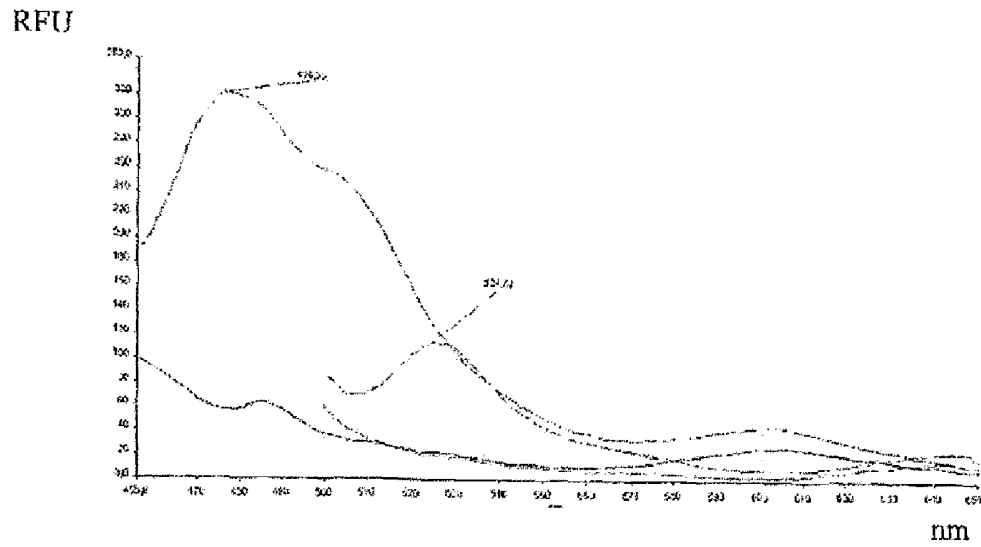
Figur 5.

6/9



Figur 6.

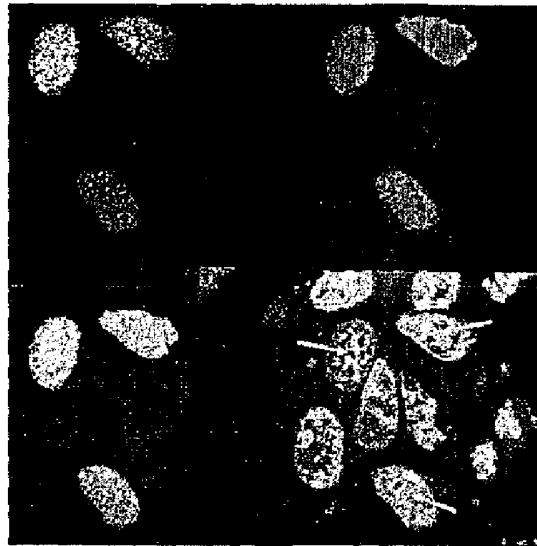
7/9



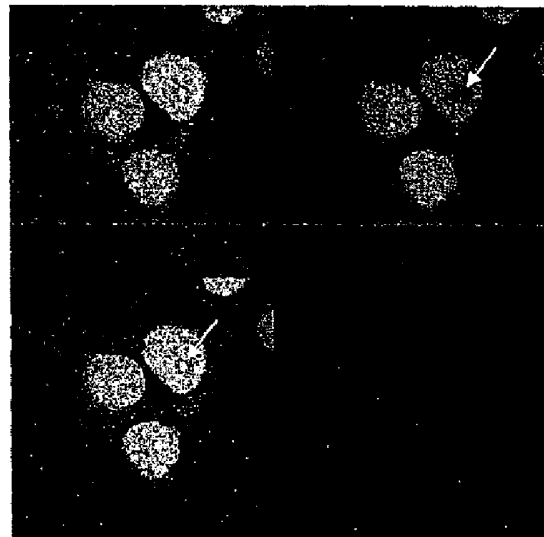
Figur 7.

8/9

A)

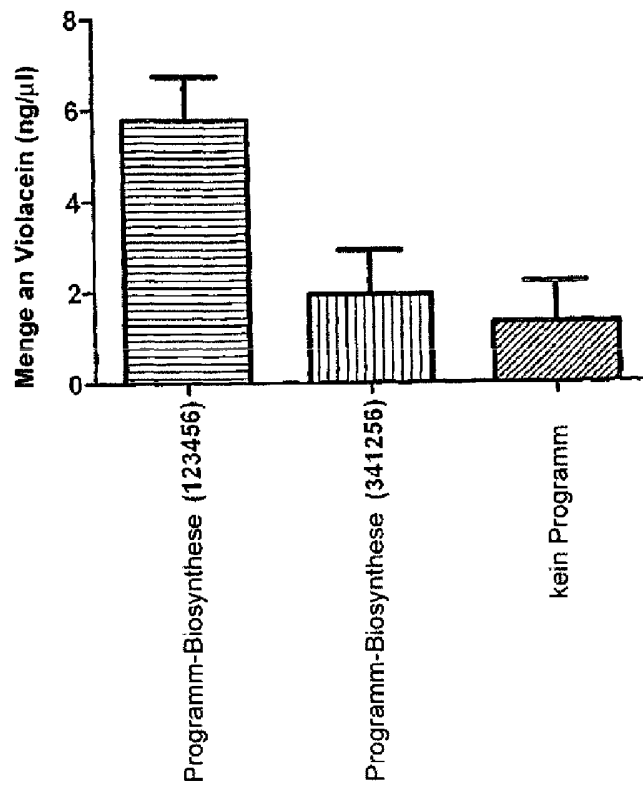


B)



Figur 8

9/9



Figur 9.