(19) österreichisches patentamt

(10) **AT 512 453 A2 2013-08-15**

(12) # Österreichische Patentanmeldung

(21) Anmeldenummer: A 9496/2010
(86) PCT-Anmeldenummer PCT/SI2010/000058
(22) Anmeldetag: 22.10.2010
(43) Veröffentlicht am: 15.08.2013

(51) Int. Cl. : ***C12N 15/10*** (2006.01)
C12N 15/52 (2006.01)
C12N 15/62 (2006.01)

(73) Patentanmelder:
Kemijski institut
1000 Ljubljana (SI)
EN-FIST center odlicnosti
1000 Ljubljana (SI)

(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-Nukleotisequenz 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.

AT 512 453 A2 2013-08-15

Zusammenfassung

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-Nukleotisequenz 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.

Gebiet der Erfindung

Das Gebiet der Erfindung ist eine verbesserte Synthese der Produkte, was mit einer angeordneten Zusammensetzung der chimären Proteine, die aus biosynthetischen Enzymen bzw. anderen funktionellen Polypeptiden sowie Nukleinsäure-bindenden Domänen zusammengesetzt sind, erreicht wird. Die angeordnete Zusammensetzung der chimären Proteine basiert auf einer Bindung an erkenntliche Stellen auf der Programm-Nukleinsäure.

Stand der Technik

Für industrielle Anwendungen sind biosynthetische Wege aus einer großen Anzahl von Enzymen derart geändert, dass sie eine höhere Gewinnung von gewünschten biosynthetischen Produkten erreichen. Verbesserte Wirksamkeit der biosynthetischen Wege mit der Absicht höhere Gewinnungen zu erreichen ist in den letzten Jahren von ausnehmendem Interesse. Bis heute wurden verschiedene Optimisationsstrategien entwickelt. Die Gewinnungen von Endprodukten der biosynthetischen Wege werden wie folgt verbessert: i) mit einer zunehmenden Menge von zugänglichem Substrat und/oder mit einer übermäßigen Expression von Enzymen der begrenzenden Schritte; ii) mit der Zugabe von heterologen Enzymen mit besseren kinetischen Eigenschaften; iii) mit der Hemmung der Verzweigung des biosynthetischen Wegs; iv) mit Kompartimentalisierung der biosynthetischen Wege mit der Ausrichtung der Enzyme eines gewissen biosynthetischen Wegs zu Zellkompartiments bzw. künstliche Kompartiments (z. B. Metabolosome), oder; v) mit Annäherung der biosynthetischen Enzyme mit einer ausgerichteten Zusammsetzung auf einem Proteingerüst (WO 2009/108774). Das Proteingerüst hat zwar viele Vorteile und ein solches Vorgehen wird durch eine Anzahl von verfügbaren Kombinationen von Peptidbindern eingeschränkt und außerdem ermöglicht ein solches Vorgehen keine Kontrolle über die räumliche Verteilung und Ausrichtung der biosynthetischen Enzyme.

Keine der obengenannten Lösungen ermöglicht eine optimale Anordnung der biosynthetischen Enzyme, die eine gewünschte Reihenfolge der biosynthetischen Reaktionen ergeben würde. Die Enzyme der biosynthetischen Wege und anderen funktionellen Polypeptide sind in lebenden Zellen oft mit einer Mikrostelle mit Multienzym-Komplexen oder Ankermechanismen eingeschränkt. Ein solcher Organisationstyp steigert die örtliche Konzentration der Enzyme und die Wirksamkeit der biosynthetischen Wege. Zum Erzielen einer optimalen Konzentration an einzelnen Enzymen und zu einer verbesserten Bildung der

Multienzym-Komplexe wurden zwei verschiedene Zutritte adaptiert: i) Ausrichtung der Enzyme zu Mikrostellen bzw. Targeting von Organellen; und ii) Bildung der Multienzym-Komplexe auf einem Polypeptidgerüst (WO 2009/108774).

Dennoch weist das Proteingerüst viele Nachteile und Einschränkungen auf.

Obwohl die Anzahl und Verteilung der Enzyme in einem Multienzym-Komplex mit einer Sequenz des Polypeptid-Gerüsts programmierbar ist, ist dabei die dreidimensionale Verteilung der Enzyme wegen der Flexibilität der Dimerisationsdomänen unvorhersehbar (Figur 1 und 2). Daneben kann die Gestaltung eines Polypeptidgerüsts mit gebundenen Proteindomänen wegen einer eingeschränkten Anzahl von Dimerisationsdomänen sehr schwierig sein und zusätzlich bestehen für jede Dimerisatonsdomäne spezifische Bedingungen, unter welchen sie faltet und funktionelle Wechselwirkungen bildet.

Kurzfassung der Erfindung

Das Problem der Anordnung der biosynthetischen Wege in eine angeordnete Reihe der biosynthetischen Enzyme kann nicht durch eine Bindung an das Proteingerüst gelöst werden. Die Lösung ist ein Molekül der Nukleinsäure mit Sequenzmotiven, auch Programm-Nukleinsäure genannt, die die Anordnung der chimären Proteine, der biosynthetischen Enzyme oder anderer funktionellen, mit Bindungsdomänen gebundenen Polypeptiden für Nukleinsäure bestimmt.

Die Erfinder haben erfunden, dass die dreidimensionale Anordnung der biosynthetischen Enzyme oder anderer funktionellen Polypeptiden nach Wunsch mit einer Programm-Nukleinsäure bestimmt werden kann. Die Erfinder haben chimäre Polypeptide vorgesehen, die aus Bindungsdomänen für Nukleinsäure und einem biosynthetischen Enzym oder einem anderen funktionellen Polypeptid zusammengesetzt sind. Die genannten chimären Proteine binden sich an das Zielmotiv der Nukleinsäure über eine Nukleinsäurebindungsdomäne. Die Programm-Nukleinsäure bestimmt die Reihenfolge der Zielmotive und damit die Sequenz und die räumliche Verteilung der biosynthetischen Enzyme oder anderer funktionellen Polypeptiden. Ein solcher angeordneter biosynthetischer, durch die Programm-Nukleinsäure gesteuerter Weg der chimären biosynthetischen Enzyme oder anderer funktionellen Polypeptiden erhöht die Gewinnung von biosynthetischem Produkt

dieses Wegs im Vergleich mit dem Proteingerüst, welches einfach Enzym-Cluster ohne einer spezifischen Anordnung bildet.

Die Anwendung der Programm-Nukleinsäure zur Immobilisierung und Bestimmung der Reihenfolge der chimären biosynthetischen Enzyme oder anderer funktionellen Polypeptiden stellt eine oder mehrere von den folgenden Möglichkeiten bereit: i) erhöht die Wirksamkeit des Flusses bzw. des Durchflusses des biosyntethischen Wegs, ii) optimiert den Stoffweckselfluss durch den Weg, iii) vermindert die Stoffwechsellast der Wirtszelle, iv) erlaubt die Möglichkeit alternativer Verzweigungen der Stoffweckselwege und damit Bildung neuer Produkte, v) erlaubt Reprogrammierung der Reihenfolge der chimären Enzyme im Inneren des Organismus, wenn eine verschiedene Programm-Nukleinsäure zugefügt wird.

Die Sequenz einer Programm-Nukleinsäure trägt mehr als zwei Zielelemente. Diese Elemente stellen spezifische Liganden oder Ziele der Nukleinsäurebindungsdomänen dar. Wenn eine Programm-Nukleinsäure in einer Lösung mit chimären Proteinen befindlich ist, binden sich die chimären Proteine auf Zielelemente im Inneren der Programmsequenz und ein Protein-Nukleinsäure-Komplex bilden, was die Bildung eines Multiprotein-Komplexes ermöglicht. Die Anordnung der chimären Proteine innerhalb eines solchen Komplex ist wegen des Folgenden vorhersehbar: i) Anordnung der Zielmotive ist durch die Sequenz einer Nukleinsäure wegen des bekannten dreidimensionalen Aufbaus der Nukleinsäuren bestimmt, ii) da die Reihenfolge der Motive vorhersehbar ist, ermöglicht sie eine angeordnete Anordnung der chimären Proteine, iii) eine solche geordnete Anordnung der chimären Proteine bestimmt die Reihenfolge der biosynthetischen Reaktionen, und diese Reihenfolge wird mit einer Nukleotidsequenz der Programm-Nukleinsäure definiert; damit wird die für die Diffusion der Zwischenprodukte der Reaktion zwischen verschiedenen Enzymen eines biosynthetischen Wegs benötigte Zeit verkürzt. Die Erfinder beschließen, dass es vorteilhaft ist, für eine solche Zusammensetzung der biosynthetischen Wege eine große Anzahl der Nukleinsäurebindungsdomänen vorhanden zu haben. Für Motive aus 9 Nucleotiden bestehen z. B. so viel wie 262144 Kombinationen, die durch verschiedene Bindungsdomänen erkannt werden können, was aber eine sehr hohe Variabilität der Zusammsetzung der biosynthetischen Wege ermöglicht.

Die vorliegende Erfindung weist zahlreiche Vorteile gegenüber dem Proteingerüst auf. Die Programm-Nukleinsäure hat keine Probleme mit der Reifung und eine angeordnete

Nukleotidsequenz kann nach Belieben ausgewählt werden. Die Bindung der Nukleinsäurebindungsdomänen auf eine Ziel-Nukleotidsequenz ist gut charakterisiert. Wegen der Nähe der an Programm-Nukleinsäure gebundenen chimären Proteine sind andere Polypeptide, die die Synthese umlenken könnten, räumlich aus dem Multienzym-Komplex ausgeschlossen. Die Bindung aller Komponenten des biosynthetischen Wegs verläuft unter denselben Bedingungen, weil alle bei der Zusammensetzung der chimären Proteine angewandte Bindungsdomänen auf demselben Typ der Proteine basieren und mit Zielnukleotiden mit der gleichen Art von Interaktionen interagieren. Im Gegenteil müssen beim Proteingerüst verschiedene Typen der Dimerisationsdomänen, die sich richtig falten und unter verschiedenen Bedingungen interagieren, zum Einsatz kommen. Die Erfindung hat des Weiteren den Vorteil, dass sie die Bildung verschiedener Reaktionsprodukte in Abhängigkeit von der Sequenz der Programm-Nukleinsäure, die zu Produktionszellen oder zum Reaktionsmedium zugefügt wird, steuern kann, wo eine verschiedene Sequenz der Programm-DNA eine verschiedene Reihenfolge der Bindung der chimären biosynthetischen Enzymen verursacht.

Die vorliegende Erfindung beschreibt eine Methode zur Herstellung von Verbindungen der biosynthetischen Wege umfassend Kultivierung genetisch geänderter Wirtszellen, die derart geändert wurden, um die chimären aus Enzymen des biosynthetischen Wegs oder anderen funtionellen Polypeptiden zusammengesetzten Proteine zu exprimieren, wobei die Polypeptide mit Nukleinsäurebindungsdomänen verbunden sind; sowie um die Programm-Nukleinsäuresequenz, die die Anordnung der chimären Proteine in die Multienzym-biosynthetische-Komplexe reguliert, zu exprimieren.

Die vorliegende Erfindung bezieht sich auf Prozesse, die mindestens drei bis etwa 100 Schritte der biosynthetischen Reaktionen umfassen. Drei Schritte sind eine Minimumgröße, wo sich die Reihenfolge der Reaktionsprozesse unterscheiden kann. Drei Schritte können zum Beispiel die Reaktionsschritte 1,2,3 auch zu einer Reihenfolge 1,3,2 oder 3,1,2 einordnen.

Eine weitere Ausführung der Erfindung umfasst die Anwendung von chimären aus Enzymen des biosynthetischen Wegs oder anderen funtionellen Polypeptiden zusammengesetzten Proteine, die mit Nukleinsäurebindungsdomänen verbunden sind und in Wirtszellen produziert werden oder werden daraus isoliert, sowie eine Programm-Nukleinsäuresequenz, die die Anordnung der chimären Proteine in die Multienzym-

biosynthetische-Komplexe reguliert. Auf diese Weise ist der biosynthetische Weg *in vitro* zusammengesetzt, entweder in einer Lösung oder mit der Immobilisierung der Programm-DNA auf ein anderes Molekül oder auf eine Festphase. Zur Ausführung einer biosynthetischen Reaktion muss ein *in vitro* zusammengebauter Komplex unter den die Biosynthese fördernde Bedingungen inkubiert werden, und auch das nötige Substrat und die für die biosynthetische Reaktion benötigte Kofaktoren bzw. ein Enzym-Komplex für die Generierung dieser Kofaktoren muss zugegeben werden.

Die vorliegende Erfindung stellt eine Beschreibung der Gestaltung der Programm-Nukleinsäure für *in vivo-* und *in vitro*-Anwendung dar. Es werden genetisch geänderte Wirtszellen dargestellt, die eine Sequenz der Programm-Nukleinsäure und eine Sequenz, die die chimären Proteine kodiert, umfasst, wobei die Proteine aus Enzymen des biosynthetischen Wegs oder anderen funktionellen Polypeptiden zusammengesetzt sind, wobei die Polypeptide mit Nukleinsäurebindungsdomänen verbunden sind, die sich an spezifische Zielelemente im Inneren der Programm-Nukleinsäure binden.

Die vorliegende Erfindung stellt eine die Sequenz der Programm-Nukleinsäure beinhaltende Nukleinsäure zur Anwendung in der Methode zur erhöhten Gewinnung von biosynthetischen Produkten dar.

Die vorliegende Erfindung ermöglicht die Gestaltung von Bioprozessen zur Produktion spezifischer Endprodukte mit der Ausscheidung anderer Polypeptide, die die Synthese zu ungewünschten Produkten wegen der Nähe der funktionellen Domänen der chimären auf die Programm-Nukleinsäure gebundenen Proteine umlenken könnten.

Die Erfindung bezieht sich auf genetisch geänderte Wirtszellen, die chimären Proteine und/oder eine replizierende Sequenz der Programm-Nukleinsäure exprimieren.

Die Erfindung bezieht sich ferner auf eine für die Herstellung einer Verbindung irgendwelches biosyntethischen Wegs angewendete Methode, die die Nähe von biosynthetischen Enzymen zur Synthese dieser Verbindung benötigt.

Die Erfindung bezieht sich ferner auf Prozesse, wie z. B. Verarbeitung von Angaben, chemische Aufspaltung, Signalisierung und Ähnliches.

Zeichnungserklärungen:

Figur 1: eine schematische Darstellung der Anwendung der Programm-Nukleinsäure.

Figur 2: der Einfluss der Spacersequenz zwischen Zielelementen/-motiven einer Nukleinsäure.

Figur 3: Chimäre Proteine, biosynthetische mit Bindungsdomänen für Nukleinsäure verbundenen Enzyme, binden sich an ein Zielelement der Nukleinsäure. Alle getesteten Bindungsdomänen Znf_Zif268, Znf_Blues, Znf_PBSII, Znf_HivC und NicTAL (Tabelle 1) binden sich spezifisch an ein Zielelement der Nukleinsäure und hemmen die Expression des Reporter-Proteins β-Galaktosidase. [A] Znf_Zif268, Znf_Blues, Znf_PBSII, Znf_HivC und NicTAL (Tabelle 1) wurden getestet. [B] Bindungsdomänen spezifisch erkennen Zielelemente der Nukleinsäure. Ein spezifisches Zielelement wurde mit einem für eine andere Bindungsdomäne spezifischen Element ausgetauscht. Die Expression des Reporter-Proteins β-Galaktosidase wurde nicht gehemmt, als ein anderes Zielelement der Nukleinsäure angewendet wurde.

Figur 4: *In vitro* Bindung der Bindungsdomänen für Nukleinsäuren auf eine Ziel-Nukleinsäure. [A] Hybridisierung der Programm-Nukleinsäure erfolgte mit dem Einspritzen von 0,5-2 μM bis 300 s um eine Endantwort von 300 RU zu erreichen. Die Bindung der chimären Proteine wurde mit dem Einspritzen von verschiedenen Konzentrationen der chimären Proteine 1 Minute beobachtet, dann folgte eine Dissoziationsschritt von 5 Sekunden. Regeneration wurde mit zwei 30-Sekunden-Einspritzen von 50 mM NaOH und einem 24-Sekunden-Einspritzen von 0,5 % SDS erreicht. [B] Chimäre Proteine von oben: Znf_Gli1, Znf_Zif268, Znf_Blues, Znf_HivC, Znf_PBSII, Znf_Jazz. [C] DNA wurde am Anfang des Zyklus gebunden und die chimären Proteine wurden 1 Minute hintereinander in zwei verschiedenen Sequenzen eingespritzt: durchgezogene Linie: Znf_Jazz, Znf_Blues, Znf_Zif268, Znf_PBSII, Znf_HivC und Znf_Gli1; unterbrochene Linie: Znf_Gli1, Znf_PBSII, Znf_HivC, Znf_Jazz, Znf_Blues und Znf_Zif268. Die Ergebnisse zeigen, dass die Bindung der chimären in den dargestellten Sequenzen gezeigten Proteine schwächer ist, ob Znf_Gli1 als erster eingespritzt wurde.

Figur 5. *In vitro*-Bindung der Nukleinsäure-Bindungsdomäne auf eine Sequenz der Nukleinsäure - Programm-DNA. Die Bindung einzelner Bindungsdomänen für Nukleinsäure Znf_Jazz, Znf_Blues, Znf_Zif268, Znf_PBSII, Znf_HivC und Znf_Gli1 wurde mit dem

Electrophoretic Mobility Shift Assay auf einem Agarosegel bestimmt. Die Pfeile zeigen die Lage des Protein-DNA-Komplexes.

Figur 6. *In vivo*-Rekonstitution von zwei unfunktionellen Hälften des grünen Fluoreszenzproteins (split_GFP genannt) gebunden mit Bindungsdomänen für Nukleinsäure in Gegenwart von Programm-DNA. [A] Eine schematische Darstellung der Rekonstitution von split_GFPs. [B] 75 ng von Plasmid, das his_Tag-Znf_Gli1-Linkerpeptid- nCFP enthält und 75 ng von Plasmid, das cCFP-Linkerpeptid-Znf_HIVC-his_Tag enthält, wurden mit 350 ng der Programm-Nukleinsäure cotransfektiert (Program SPR, split, FRET; Tabelle 1). CFP-Emission (475 nm) wurde nach Anregung bei 433 nm detektiert. [C] 75 nm jedes Plasmids mit split-YFP-chimären-Protein (his_Tag- Znf_PBSII_Linkerpeptid-nYFP und cYFP-Linkerpeptid -Znf_Zif268-Linkerpeptid -his_Tag) wurden cotransfektiert mit 350 ng des Plasmids mit Programm-Nukleinsäure (Program SPR, split, FRET; Tabelle 1). Das YFP-Emissionssignal bei 529 nm wurde nach Anregung bei 513 nm unter dem Konfokalmikroskop detektiert.

Figur 7. *In vitro*-Rekonstitution von zwei unfunktionellen split-GFPs gebunden mit Bindungsdomänen für Nukleinsäure in Gegenwart von Programm-DNA. Lysate der mit Plasmiden transfektierten HEK293-Zellen, die his_Tag_Znf_Gli1-Linkerpeptid-nCFP und cCFP-Linkerpeptid-Znf_HIVC-his_Tag oder his_Tag-Znf_PBSII_Linkerpeptid-nYFP und cYFP- Linkerpeptid-Znf_Zif268-Linkerpeptid-his_Tag kodierten, wurden mit 50µg der Programm-DNA vermischt (Program SPR, split, FRET; Tabelle 1). Nach 18 Stunden der Inkubation bei 4°C wurde Fluoreszenzspektrum für CFP und YFP beobachtet. Emissionspeaks sind mit Pfeilen dargestellt.

Figur 8. *In vivo*-Rekonstitution von zwei Paaren von Split-Fluoreszenzproteinen (split CFP und split YFP) wurde mit der FRET Methode bestimmt. Die HEK293-Zellen wurden mit Plasmiden, die his_Tag-Znf_Gli1-Linkerpeptid-nCFP, cCFP-Linkerpeptid-Znf_HIVC-his_Tag, his_Tag-Znf_PBSII_Linkerpeptid-nYFP, cYFP-Linkerpeptid-Znf_Zif268-Linkerpeptid-his_Tag kodierten, und mit dem Plasmid, das die Programm-DNA enthielt (Program SPR, split, FRET, Tabelle 1) kotransfektiert. [A. oben links] Darstellung der CFP-Rekonstitution aus Split-CFPs. Anregung bei 453 nm, Emission von 470 bis 510 nm. [A. oben rechts] Darstellung der YFP-Rekonstitution aus Split-YFPs. Anregung bei 415 nm, Emission von 525 bis 560 nm. [A. unten links] Gleichzeitige Darstellung von CFP und YFP von oben.

[A. unten rechts] Die gleichen Zellen mit dem Farbstoff Hoechst 34580 gefärbt. Die Pfeile zeigen die Zellen, die ein FRET-Signal zeigen, was für die Ko-Lokalisierung von allen vier Split-GFP-Fusionen im Inneren des Zellkerns spricht. [B. oben links] Darstellung des Donators, des rekonstituierten CFPs vor Luminophor-Deaktivierung "photobleaching". [A. oben rechts] Darstellung des Akzeptors, des rekonstituierten mCitrinin nach Luminophor-Deaktivierung. Die gebleichte Stelle ist mit Pfeilen gekennzeichnet. [B. unten links] Darstellung des Donators, des rekonstituierten CFPs nach Luminophor-Deaktivierung. Erhöhte Fluoreszenz des Donators nach der Luminophor-Deaktivierung des Akzeptors ist mit Pfeilen gekennzeichnet.

Figur 9. Biosynthese von Violacein in Gegenwart der Programm-DNA. Über Nacht kultivierte Kulturen von *E. coli* mit Plasmiden, die die chimären Proteine Znf_Blues-Linkerpeptid-vioA, Znf_Zif268-Linkerpeptid-vioB, Znf-PBSII-Linkerpeptid-vioE, Znf-HivC-Linkerpeptid-vioD, Znf_Gli1-Linkerpeptid- vioC (Tabelle 1) kodieren, mit oder ohne Plasmid, das Programm-DNA kodiert (Tabelle 1, Biosynthese-Programm (123456) oder Biosynthese-Programm (341256)). Nach 17,5 Stunden erfolgte eine Extraktion, eine TLC-Analyse und die Menge von Violacein wurde bestimmt. Die Menge von Violacein, das in Gegenwart des Programm-DNAs (Biosynthese-Programm 123456) hergestellt wurde, wurde viel größer als im Falle des Programm-DNAs (Biosynthese-Programm 341256) bzw. wenn kein Programm in *E. coli* anwesend war.

Detaillierte Beschreibung der Erfindung:

Definitionen:

Der Begriff "Programm / Programm-Nukleinsäure / Programm-Nukleotidsequenz / Sequenz der Programm-DNA" bezieht sich auf eine Nukleinsäure, die eine Zielsequenz von Nukleotiden einer beliebigen Länge und Anzahl enthält, die von Bindungsdomänen für Nukleinsäure erkannt werden. Diese Zielsequenzen sind mit Nukleotidsequenzen von beliebigen Längen, Spacer genannt, getrennt. Das Programm kann allein in eine Wirtszelle eingebracht werden, oder kann in einen beliebigen Vektor eingesetzt und in der Wirtszelle multipliziert werden.

Der Begriff "Ziel / Zielnukleotidsequenz / Motiv der Zielnukleotidsequenz" wie in dieser Beschreibung benutzt, bezieht sich auf eine Nukleotidsequenz einer beliebigen Länge, die ein Substrat für Bindungsdomänen der Nukleinsäure ist.

Der Begriff "Bindungsfaktor / Nukleinsäurebindungsdomäne" bezieht sich auf irgendwelches Molekül mit der Fähigkeit, sich an Nukleinsäure zu binden. Der Bindungsfaktor kann ein natürliches oder ein künstlich konstruiertes Gesamtprotein oder nur ein Proteinsegment mit der Fähigkeit, sich an eine spezifische Nukleotidsequenz der Nukleinsäure zu binden, sein. Erfindungsgemäß ist ein Nukleinsäure-Bindungselement als ein Träger der Enzyme angewendet, die in den biosyntethischen Weg oder in irgendwelchem anderen funktionellen Polypeptid eingeschlossen sind. Die Bindungsdomäne der Nukleinsäuren ist mit dem Enzymmolekül mit einer chemischen Bindung verbunden. Dies wird mit einem Linkerpeptid erzielt. Die Bindungsdomäne für Nukleinsäuren bindet sich auf eine oder mehrere spezifische Stellen auf Programm der Nukleinsäuresequenz und stellt eine hohe örtliche Konzentration sowie eine richtige räumliche Position der im biosyntethischen Weg eingesetzten Enzyme bereit.

Der Begriff "Spacer / Spacersequenz" bezieht sich auf eine Nukleotidsequenz einer beliebigen Länge, die Zielnukleotidsequenzen trennt.

Der angewendete Begriff "Nukleinsäure" bezieht sich auf eine polymere Form der beliebig langen Nukleotide, Ribonukleotide oder Desoxyribonukleotide und ist nicht beschränkt auf ein-, zwei- oder mehrkettige DNA oder RNA, genomische DNA, cDNA, DNA-RNA-Hybride oder Polymere mit einer Phosphat-Thiol-Rückgrat, aus Purin- und Pyrimidinbasen oder anderer natürlichen, chemischen oder biochemischen modifizierten, unnatürlichen oder derivatisierten Nukleotidbasen.

Der Begriff "Polypeptid", "Protein", "Peptid" bezieht sich auf eine Polymerform der Aminosäuren von beliebigen Längen, die kodierte und nichtkodierte Aminosäuren, chemisch oder biochemisch modifizierte oder derivatisierte Aminosäuren und Polypeptide mit einer modifizierten Peptidrückgrat umfasst. Der Begriff "funktionelles Polypeptid", wie hier angewendet, bezieht sich auf ein Polypeptid einer beliebigen Länge, das irgendwelche Funktion, wie z. B. Bildung einer Struktur, Ausrichtung auf eine spezifische Stelle, Targeting der Organellen, Erleichterung und Auslösung der chemischen Reaktionen, Bindung an andere

funktionelle Polypeptide, exprimiert. Der Begriff "Enzym des biosynthetischen Wegs", wie hier angewendet, bezieht sich auf ein Polypeptid einer beliebigen Länge, das fähig ist, eine neue chemische Bindung zu bilden.

Der Begriff "das chimäre Protein", wie hier angewendet, hat eine allgemeine Bedeutung und bezieht sich auf eine Polymerform der Aminosäuren einer beliebigen Länge, die aus mehr als einem/einer Protein/Domäne/Segment zusammengesetzt ist, möglichst miteinander durch einen beliebig langen, vorzugsweise aus 1 bis 40 Aminosäuren zusammengesetzten Linker, verbunden, und dass wenigstens ein/eine Protein/Domäne/Segment ein Nukleotid-Bindungsfaktor ist, ganz oder eine Bindungsdomäne, und der andere ist ein Enzym in dem biosynthetischen Weg bzw. ein anderes funktionelles Polypeptid, ganz oder eine aktive Domäne.

Der Begriff "heterolog" bezieht sich auf Kontext von genetisch modifizierten Wirtszellen und bezieht sich auf ein Polypeptid, für welches mindestens eine der folgenden Behauptungen gültig ist: (a) Polypeptid ist für die Wirtszelle fremd ("exogen") (in Natur nicht in ihr anwesend); (b) Polypeptid ist in der Natur in einem gewissen Wirtsmikroorganismus oder einer Wirtszelle anwesend ("endogen"), aber wird entweder in unnatürlichen (mehr als erwartet bzw. in größeren Mengen als in der Natur anwesend) Mengen in der Zelle produziert oder sich in der Nukleotidsequenz von der endogenen Nukleotidsequenz derart unterscheidet, dass dasselbe Protein (weist die gleiche bzw. wesentlich ähnliche Aminosäuresequenz auf) als das endogene in unnatürlichen (mehr als erwartet bzw. mehr als in der Natur anwesend) Mengen in der Zelle produziert wird.

Der Begriff "homolog", wie hier angewendet, bezieht sich auf Proteine oder Nukleinsäuren mit einer aufbewahrten Nukleotidsequenz, vorzugsweise mit einer mindestens 50%-igen Bewahrung, mit einer wenigstens 20%-igen Bewahrung, die mit Protein- oder Nukleinsäure-Vergleichstechniken, die den Fachmännern bekannt sind, bestimmt wird. Homologe Proteine sind dadurch charakterisiert, dass sie die gleichen Funktionen in der Zelle ausführen. Homologe Nukleinsäuren kodieren für homologe Proteine.

Der Begriff "rekombinant", wie hier angewendet, bedeutet, dass eine gewisse Nukleinsäure (DNA oder RNA) als ein Produkt verschiedener Kombinationen von Klonieren, Restriktionen und/oder Ligationen entsteht, die zu einem Konstrukt führen, dessen strukturell

kodierende oder nicht-kodierende Sequenzen sich von endogenen Nukleinsäuren in Natursystemen unterscheiden. Im Allgemeinen kann eine DNA-Sequenz, die eine strukturell kodierende Sequenz kodiert, aus cDNA-Fragmenten oder aus kurzen Oligonukleotidlinkern oder aus syntetischen Oligonukleotiden zusammengebaut werden, aus denen eine synthetische Nukleinsäure erhalten werden kann, die sich aus einer rekombinanten Transkriptionseinheit in die Zellen oder in die Nicht-Zell-Transkription und Translationssystem exprimiert. Eine solche Sequenz kann in Form eines offenen Leserahmens eingesetzt werden und es kommt nicht zu Störungen wegen der internen untranslatierten Sequenzen bzw. Intronen, die normalerweise in eukaryotischen Genen anwesend sind. Die genomische DNA, die wichtige Sequenzen umfasst, kann auch zur Bildung eines rekombinanten Gens oder einer Transkriptionseinheit eingesetzt werden. Die Sequenzen einer untranslatierten DNA können sich an der 5'-Ende oder 3'-Ende des offenen Leserahmens befinden, wo solche Sequenzen keinen Einfluss auf die Manipulation oder Expression der kodierenden Regionen ausüben und können als Modulatoren der Produktion von gewünschten Produkten durch verschiedene Mechanisme wirken (siehe "DNA-regulatorische Sequenzen", unten).

Der Begriff "Wirtszelle", wie hier angewendet, bezieht sich auf eine *in vivo* oder *in vitro* eukaryotische oder prokaryotische Zelle, auf einen Zell- oder Vielzellorganismus (Zell-Linie), das als eine einzellige Einheit kultiviert wurde, die ist bzw. wurde als ein Akzeptor der Nukleinsäure angewendet (Expressionsvektor, der eine Nukleotidsequenz umfasst, die für ein oder mehrere Genprodukte des biosynthetischen Wegs kodiert, wie z. B. Genprodukte des Mevalonatwegs) und Abkömmlinge der Originalzelle, die mit der Nukleinsäure genetisch verändert wurde, enthielt. Selbstverständlich, die Abkömmlinge einer Zelle sind nicht dringend den Eltern ganz identisch in der morphologischen Form und im ganzen DNA-Komplement wegen natürlicher, zufälliger oder geplanten Mutationen. "Genetisch veränderte Wirtszelle" (auch "rekombinante Wirtszelle") ist eine Wirtszelle, in welche eine heterologe Nukleinsäure, ein Expressionsvektor eingetragen wurde. Eine prokaryotische Wirtszelle ist zum Beispiel eine genetisch veränderte Wirtszelle (Bakterie), die mit der Eintragung einer heterologen Nukleinsäure entstanden ist, was bedeutet, das es sich um eine exogene Nukleinsäure, die der prokaryotischen Wirtszelle fremd (normalerweise nicht in Natur anwesend) ist, oder um eine rekombinante Nukleinsäure, die normalerweise in einer prokaryotischen Wirtszelle nicht anwesend ist, oder um eine eukaryotische, genetisch modifizierte Wirtszelle, handelt. Eine eukaryotische Wirtszelle wird eine genetisch veränderte

Wirtszelle, indem in eine geeignete eukaryotische Wirtszelle eine heterologe Nukleinsäure eingeführt wurde, d. h. eine exogene Nukleinsäure, die der eukaryotischen Wirtszelle fremd ist, oder eine rekombinante Nukleinsäure, die in einer eukaryotischen Wirtszelle normalerweise nicht anwesend ist.

Der Begriff "biosynthetischer Weg", wie hier angewendet, bezieht sich auf eine Sequenz der enzymatischen oder anderen Reaktionen, durch welche sich eine Verbindung in eine andere umwandelt, da im Organismus oder in vitro neue kovalente Bindungen entstehen.

Der Begriff "in vitro", wie hier angewendet, bezieht sich auf einen Prozess, der nicht in lebenden Organismen bzw. Zellen, sondern in einer kontrollierten Umwelt ausgeführt ist.

Der Begriff "Linkerpeptid" bezieht sich auf eine kurze Aminosäuresequenz, welcher Aufgabe ist nur Trennung einzelner Domänen der Fusionsproteine. Die Rolle des Linkerpeptids in einem Fusionsprotein kann aber auch die Einleitung der Spaltestelle oder der Stelle für Posttranslations-Modifikationen sein, samt der Einleitung der Stellen für ein verbessertes Prozessieren der Antigene. Die Länge des Linkerpeptids ist nicht begrenzt, doch beträgt normalerweise bis zu 20 Aminosäuren.

Eine heterologe Nukleinsäure ist normalerweise in den Expressionsvektor eingetragen. Geeignete Vektoren sind: Plasmide, Virusvektoren und andere. Die mit den Wirtszellen kompatiblen Vektoren sind den Fachmännern gut bekannt und umfassen alle Elemente, die für die Transkription und Translation der Nukleinsäuren notwendig sind. Ein Expressionsvektor enthält normalerweise eine Kassette mit Antibiotikaresistenz, eine Sequenz für das chimäre Protein unterhalb von einem für eine gesteuerte Expression in der Wirtszelle geeigneten Promotor, ein Polyadenilierungssignal und einen Transkriptionsterminator.

Methode zur Synthese eines biosynthetischen Produkts/Vorläufers

Diese Erfindung stellt eine Methode zur Synthese eines Produkts oder Vorläufers des biosynthetischen Wegs in einer genetisch veränderten Wirtszelle oder *in vitro* bereit. Die Methode umfasst Kultivierung der genetisch veränderten Wirtszellen in geeigneten Bedingungen, wobei der genetisch veränderter Wirt umfasst: a) Nukleinsäuren mit einer Nukleotidsequenz, die für drei oder mehrere mit den Enzymen des biosynthetischen Wegs

oder mit anderen funktionellen Polypeptiden verbundenen Bindungsfaktoren kodiert, oder b) Nukleinsäure mit einer oder mehreren Programm-Nukleotidsequenzen.

In dem zweiten Ausführungsbeispiel der Erfindung umfasst die Methode die Bildung eines Produkts *in vitro*. Die *in vitro* Methode umfasst: a) mindestens drei oder mehr Bindungsfaktoren, die mit den chimären Enzymen des biosynthetischen Wegs oder mit anderen funktionellen Polypeptiden verbunden sind, b) eine Nukleinsäure mit einer oder mehreren Programm-Nukleotidsequenzen und c) ein Substrat des ersten Enzyms und die im Gemisch anwesenden Kofaktoren. Der Bindungsfaktor bindet das Zielelement der Nukleinsäure im Inneren der Programm-Nukleotidsequenz mit einer ausreichenden Affinität, was eine Verbindung zwischen dem chimären Enzym des biosynthetischen Wegs und der Programm-Nukleotidsequenz bereitstellt. Die Affinität zwischen dem chimären Enzym des biosynthetischen Wegs und der Programm-Nukleotidsequenz ist hoch genug, dass die chimären Enzyme des biosynthetischen Wegs oder die anderen funktionellen Polypeptide auf der Programm-Nukleotidsequenz immobilisiert sind. Die Sequenz der Programm-Nukleinsäure bestimmt die Reihenfolge der anliegenden Bindungsfaktoren und damit auch die Reihenfolge der chimären Enzymen oder anderer funktionellen Polypeptide. Die Reihenfolge der Zielelemente der Nukleinsäure mit den gebundenen funktionellen Polypeptiden kann in einer solchen Reihenfolge wie in einer gewissen biosynthetischen Reaktion geordnet werden. Ebenso kann die Reihenfolge der Zielelemente der Nukleinsäure mit geeigneten funktionellen Polypeptiden durch Änderung der Reihenfolge der benachbarten Zielsequenzen in der Programm-Nukleotidsequenz geändert werden, was eine verschiedene Sequenz der benachbarten biosynthetischen Enzymen oder anderen funktionellen Polypeptiden bereitstellt. Die Kultivierung der Wirtszellen erfolgt so, dass das Substrat des ersten Enzyms: a) entweder in der Zelle anwesend oder b) zu den Zellen von außen zugegeben ist. Die Enzyme des biosynthetischen Wegs werden in der Zelle synthetisiert und entweder a) das Substrat zu einem Produkt in der Zelle umwandeln, oder b) werden aus der Zelle abgesondert und umwandeln das Substrat außen in das Produkt.

Die Anwendung der Programm-Nukleotidsequenz zur Immobilisierung der Enzyme des biosynthetischen Wegs stellt mindestens eine der folgenden Wirkungen bereit: erhöhte Wirksamkeit des biosynthetischen Wegs und einen optimierten metabolischen Fluss, reduzierte metabolische Belastungen und Konzentration von potentiell toxischen freien

Zwischenprodukten im Zytosol, Möglichkeit einer alternativen Verzweigung der Wegen, die zu neuen Produkten des biosynthetischen Prozesses führt. Die Programm-Nukleotidsequenz ermöglicht eine ausgeglichene Enzymaktivität. Zum Beispiel, eine Programm-Nukleotidsequenz kann derart geändert werden, dass sie mehr Wiederholungen der Zielelemente der Nukleinsäuren für die chimären Polypeptide, die eine niedrigere Aktivität bei der Umwandlung des Substrats aufweisen, beinhaltet, was eine größere Anzahl der Wiederholungen der chimären eine niedrigere Aktivität bei der Umwandlung des Substrats aufweisenden Polypeptide verursacht, als derer mit einer höheren Aktivität. Das ist vorteilhaft in den Fällen, wenn ein Enzym mit einer niedrigeren Aktivität eine Geschwindigkeit begrenzende Stufe des biosynthetischen Wegs katalysiert. Die Anwendung der Programm-Nukleotidsequenz steigt die Wirksamkeit und optimieret den Stoffwechselfluss durch die Verkürzung des Abstands zwischen der in dem biosynthetischen Weg beteiligten Enzyme; wegen einer höher Wirksamkeit und des optimierten Stoffwechselflusses kann die Gewinnung von einem Produkt gleich oder höher sein bei einer kleineren angewendeten Menge des Enzyms. Synthese einer kleineren Menge des Enzyms in der Wirtszelle ist ein Vorteil, da auf die Wirtszelle eine niedrigere Stoffwechselbelastung ausgeübt wird. Da die Umwandlung der Zwischenprodukte des biosynthetischen Wegs durch die Anwendung der Programm-Nukleotidsequenz wirksamer ist, reduziert sich die Menge/Konzentration der Zwischenprodukte des biosynthetischen Wegs im Zytosol oder im Zytoplasma. Eine reduzierte Menge von freien Zwischenprodukten des biosynthetischen Wegs ist vorteilhaft, wenn diese Zwischenprodukte für die Wirtszelle giftig sind.

In einigen Ausführungsbeispielen sind mindestens drei chimäre biosynthetische Enzyme oder andere funktionelle Polypeptide auf der Programm-Nukleotidsequenz immobilisiert. Der erste chimäre funktionelle Enzym oder Polypeptid produziert ein erstes Produkt, das das Substrat des nächsten chimären Enzyms oder anderen funktionellen Polypeptids ist. Chimäre Proteine, Enzyme des biosynthetischen Wegs oder andere funktionellen Polypeptide sind in der gegenseitigen Nähe angeordnet. Auf diese Weise ist die effektive Konzentration des ersten Produkts hoch und der andere chimäre Enzym des biosynthetischen Wegs kann effektiv auf das erste Produkt wirken.

Drei oder mehr (d. h. drei, vier, fünf, sechs, sieben oder mehr) funktionelle Polypeptide können auf der Programm-Sequenz der Nukleotide immobilisiert werden. In

einigen Ausführungsbeispielen umfasst die Programm-Sequenz der Nukleinsäuren (in Richtung 5′ gegen 3′) a) eine Wiederholung des Elements der Zielsequenz der Nukleotiden für das erste chimäre Polypeptid oder Enzym des biosynthetischen Wegs, b) eine Wiederholung des Elements der Zielsequenz der Nukleotiden für das zweite chimäre Polypeptid oder Enzym des biosynthetischen Wegs, und c) eine Wiederholung des Elements der Zielsequenz der Nukleotiden für das dritte chimäre Polypeptid oder Enzym des biosynthetischen Wegs.

In anderen Ausführungsbeispielen umfasst die Programm-Sequenz der Nukleinsäuren (in Richtung 5′ gegen 3′) a) eine Wiederholung des Elements der Zielsequenz der Nukleotiden für das erste chimäre Polypeptid oder Enzym des biosynthetischen Wegs, und b) zwei oder mehr (d. h. zwei, drei, vier oder mehr) Wiederholungen des Elements der Zielsequenz der Nukleotiden für das zweite chimäre Enzym des biosynthetischen Wegs oder das zweite funktionelle Polypeptid. Auf diese Weise kann das Verhältnis jedes chimären funktionellen Polypeptids im biosynthetischen Weg geändert werden. Zum Beispiel, das Verhältnis zwischen dem ersten und dem zweiten chimären Polypeptid des biosynthetischen Wegs kann von etwa 0,1:10 bis etwa 10:0,1 und das Verhältnis zwischen dem zweiten und dem dritten Polypeptid kann von etwa 0,1:10 bis etwa 10:0,1 betragen, usw.

In manchen Ausführungsbeispielen sind auf der Programm-Nukleinsäure mindestens drei oder mehr (d. h. drei, vier, fünf, sechs, sieben oder mehr) chimäre Enzyme des biosynthetischen Wegs immobilisiert. Das erste chimäre Enzym des biosynthetischen Wegs produziert das erste Produkt, das das Substrat für das zweite Produkt ist, das das Substrat für das dritte Produkt des biosynthetischen Wegs ist: in diesen Ausführungsbeispielen diktiert die Sequenz der Zielnukleotide in der Programm-Nukleinsäure die Reihenfolge und die Anzahl der Kopien des chimären Enzyme des biosynthetischen Wegs.

Gestaltung der Programm-Nukleotidsequenz

Eine Programm-Nukleotidsequenz ist derart gestaltet, dass sie die Enzyme des biosynthetischen Wegs zu einem funktionellen Komplex anordnet. Die Programm-Nukleotidsequenz ist aus drei oder mehr Ziel-Nukleotidelementen zur Bindung der Bindungsfaktoren zusammengesetzt. Die Bindung von der mit chimären Enzymen des biosynthetischen Wegs verbundenen Bindungsfaktoren der Nukleinsäuren auf das Ziel-

Nukleotidelement stellt die Immobilisierung des Enzyms auf die Programm-Nukleotidsequenz bereit. Jedes Zielelement verfügt über einen entsprechenden Bindungspartner im chimären Enzym des biosynthetischen Wegs oder im funktionellen Polypeptid. Ein gewisses Ziel-Nukleotidelement kann dem anderen Ziel-Nukleotidelement unmittelbar anliegend oder von ihm mit einem Nukleotidspacer getrennt sein.

Die Programm-Nukleotidsequenz kann *in vivo* in verschiedene Typen von Wirtszellen in verschiedenen Formen eingeführt werden. Beispiele der verschiedenen Formen der Programm-Nukleotidsequenz sind, aber nicht darauf beschränkt, Baculovirus-Vektor, Bakteriophage-Vektor, Plasmide, Phagemide, Cosmide, Fosmide, bakterielle künstliche Chromosomen (ang. Bacterial Artificial Chromosomes), Virus-Vektor (z. B. Virusvektor auf Basis von Vaccinia-Virus, Polyovirus, Adenovirus, AA, SV40, Herpesvirus und ähnliches, aber nicht darauf beschränkt), künstliche Chromosomen auf P1 Basis, Hefepilzplasmide aus Hefepilzen, künstliche Hefepilzchromosome und andere Vektoren. Die Programm-Nukleotidsequenz kann in verschiedene Wirtszellen, auch als ein lineares Nukleinsäuremolekül eingeführt werden. Alle diesen Formen der Programm-Nukleotidsequenzen können auch *in vitro* eingesetzt werden. Wenn die Programm-Nukleotidsequenz mindestens teilweise aus RNA zusammengesetzt ist, kann sie *per se* eingesetzt werden, aber auch als DNA, die die Programm-Nukleotidsequenz kodiert. Die für die Programm-Sequenz der Nukleinsäuren kodierende DNA kann *in vitro* oder *in vivo* angewendet werden und kann in verschiedenen Formen wie oben aufgeführt vorhanden sein, aber nicht auf die oben aufgeführten Beispiele beschränkt.

In einem Produktionssystem kann die Programm-Nukleotidsequenz in einer oder mehreren Kopien vorhanden sein. Mehrere Kopien einer Programm-Nukleotidsequenz können eine größere Anzahl von funktionellen biosynthetischen, aus immobilisierten chimären Enzymen des biosynthetischen Wegs zusammengesetzten Komplexen bereitstellen. Einzelne Kopien der Programm-Nukleotidsequenzen können auf einem oder mehr Nukleinsäuremolekülen vorhanden sein. Eine größere Anzahl von einzelnen Nukleinsäuremolekülen stellt also auch eine größere Anzahl von Kopien der Programm-Nukleotidsequenz im Produktionssystem bereit.

Zum Beispiel, doch nicht nur darauf beschränkt, eine bakterielle Wirtszelle kann durch selbstreplizierende Plasmiden oder irgendwelchem anderen Typ der Programm-

Nukleotidsequenz in einer oder mehreren Kopien enthaltenden Nukleinsäure transformiert werden, was also eine größere Gesamtanzahl der Programm-Nukleotidsequenzen in Produktionssystem bereitstellt. Das eine oder mehrere Kopien der Programm-Nukleotidsequenz enthaltende Plasmid kann im System in einer oder mehreren Kopien vorhanden sein (zum Beispiel, doch nicht nur auf das Plasmid mit einer großen Anzahl von Kopien, durch welches sich die bakterielle Zelle transformiert). Eine größere Anzahl von Plasmiden im Produktionssystem mit einer Kopie oder mehreren Kopien der Programm-Nukleotidsequenz stellt auch eine höhere Gesamtanzahl von Kopien der Programm-Nukleotid-Sequenz im Produktionssystem bereit.

Zum Beispiel, aber nicht darauf beschränkt, eine bakterielle Wirtszelle kann durch eine Nukleotidsequenz transformiert werden, die für ein ein-, zwei-, teilweise zwei-, mehr- oder teilweise mehrkettige RNA-Molekül kodiert. Die Programm-Nukleotidsequenz kann in einer oder mehreren Kopien auf einem einzelnen RNA-Molekül vorhanden sein. Mehrere Kopien der Programm-Nukleotidsequenz auf einem einzelnen RNA-Molekül stellen eine größere Gesamtanzahl von Kopien der Programm-Nukleotidsequenz im Produktionssystem bereit. Auch die für das RNA-Molekül kodierende Sequenz kann in dem Produktionssystem in einer oder mehreren Kopien vorhanden sein. Eine größere Anzahl von Kopien der für das RNA-Molekül kodierenden Sequenz stellt eine größere Anzahl von RNA-Molekülen im Produktionssystem und also auch eine größere Anzahl von Programm-Nukleotidsequenzen im Produktionssystem bereit. Die Gesamtzahl der einzelnen RNA-Moleküle im Produktionssystem kann durch Promotor reguliert werden, die zur Regulierung der Transkription der kodierenden Nukleotidsequenzen zu RNA-Molekülen angewendet werden. RNA-Moleküle sind aus einer oder mehreren Kopien der Programm-Nukleotidsequenz zusammengesetzt.

Stärkere Promotor stellen eine größere Anzahl von Kopien der RNA-Moleküle im Produktionssystem bereit und deswegen stellen eine größere Gesamtzahl der verschiedenen Programm-Nukleotidsequenzen in Produktionssystem bereit. Eine größere Gesamtzahl der Programm-Nukleotidsequenzen stellt eine größere Anzahl der funktionellen biosynthetischen Komplexe bereit, die aus immobilisierten chimären Enzymen des biosynthetischen Wegs zusammengesetzt sind.

Die Programm-Nukleotidsequenz wird aus mindestens drei Zielnukleotidsequenzen zusammengesetzt, also binden sich mindestens drei entsprechende Bindungselemente in drei chimären Enzymen des biosynthetischen Wegs auf die Programm-Nukleotidsequenz. Die Programm-Nukleotidsequenz weist eine, zwei oder mehrere Kopien jedes Ziel-Nukleotid-Elements auf. Jedes Ziel-Nukleotid-Element kann also in einer oder mehreren Kopien vorhanden sein. Diese Kopien können unmittelbar nebeneinander sein oder durch einen Nukleotidspacer voneinander getrennt werden. Die Wiederholungen der Zielelemente können auch mit einem oder mehreren anderen Zielelementen, die in einer oder mehreren Kopien vorhanden sein können, getrennt werden. Falls nur drei Ziel-Nukleotid-Elemente in dem Programm-Nukleotidsequenz vorhanden sind, können sie verschieden sein, was eine richtige räumliche Verteilung oder eine hohe örtliche Konzentration von zwei verschiedenen chimären Enzymen des biosynthetischen Wegs oder anderen funktionellen Polypeptiden bereitstellt. Falls nur drei Ziel-Nukleotid-Elemente auf dem Programm-Nukleotidsequenz vorhanden sind, können sie gleichartig sein, was eine richtige räumliche Organisation oder eine hohe örtliche Konzentration für die drei Kopien desselben chimären biosynthetischen Enzym bereitstellt.

Das spezifische Ziel-Nukleotid-Element bestimmt die Nukleotidsequenz vom 5′-Ende gegen 3′-Ende des Nukleinsäuremoleküls. Auf einer doppelsträngigen Programm-Nukleotidsequenz oder einem doppelsträngigen Segment der Programm-Nukleotidsequenz kann die Zielnukleotidsequenz auch ein Reverse-Komplement der Sequenz der ersten Ziel-Nukleotidsequenz sein. Der gleiche Bindungsfaktor kann sich auf das Ziel-Nukleotid-Element und auf sein Reverse-Komplement binden, aber in verschiedener Ausrichtung und an der anderen Seite der Programm-Nukleotidsequenz. Deswegen kann das Reverse-Komplement der Zielnukleotidsequenz anstatt der Zielnukleotidsequenz eingesetzt werden, wenn zwei chimäre Enzyme des biosynthetischen Wegs oder zwei andere funktionelle Polypeptide zu groß sind, um sich auf die Programm-Nukleotidsequenz unmittelbar nebeneinander binden zu können, oder zur Erreichung einer anderen räumlichen Verteilung von zwei Kopien desselben chimären Enzyms des biosynthetischen Wegs oder anderen funktionellen Polypeptids an der Programm-Nukleotidsequenz.

Der Nukleotidspacer kann zwischen zwei Ziel-Nukleotid-Elementen eingesetzt werden, um genug Platz für große chimäre Enzyme des biosynthetischen Wegs oder andere

funktionelle Polypeptide, die in dem biosynthetischem Weg nacheinander wirken, bereitzustellen. In einem doppelsträngigen Programm-Nukleotidsequenz oder in einem doppelsträngigen Segment der Programm-Nukleotidsequenz kann der Nukleotidspacer für eine richtige räumliche Verteilung von zwei benachbarten oder nahe gelegten Enzymen des biosynthetischen Wegs oder anderen funktionellen Polypeptiden sorgen. Die Programm-Nukleotidsequenz kann z. B. eine doppelsträngige DNA-Sequenz sein, die eine Helix bildet. Mit der Veränderung der Länge des Nukleotidspacers wird eine verschiedene räumliche Verteilung der immobilisierten chimären Enzyme des biosynthetischen Wegs oder ein großer Abstand zwischen zwei benachbarten chimären Enzymen des biosynthetischen Wegs erreicht (Figur 2).

Individuelle Programm-Nukleotidsequenz hat eine generelle Formel [((X oder X')n Sp)]m, worin Xn eine Ziel-Nukleotidsequenz bedeutet, Xn´ ist ein Reverse-Komplement der Ziel-Nukleotidsequenz, S ist ein fakultativer Nukleotidspacer einer beliebten Länge, m ist eine Ganzzahl größer als 3 oder gleich 3 und stellt die Anzahl der Ziel-Nukleotidsequenzen mit dem fakultativen Nukleotidspacer in der Programm-Nukleotidsequenz dar, n bedeutet den Typ des Ziel-Nukleotid-Elements (z. B. X1, X2, usw.) und stellt verschiedene einzelne Ziel-Nukleotid-Elemente dar, p bedeutet den Typ des Nukleotidspacers (z. B. S1, S2, usw.) und stellt verschiedene einzelne Nukleotidspacer dar. Einzelne Zielnukleotidsequenzen in der Programm-Nukleotidsequenz können einer beliebigen Länge sein und sind von der entsprechenden Bindungselementen abhängig. Zinkfinger mit drei Fingern können zum Beispiel eingesetzt werden. In diesen Fällen beträgt die Länge der Ziel-Nukleotidsequenz normalerweise, aber nicht immer, 9 oder 10 Nukleotide.

In einigen Ausführungsbeispielen weist die Programm-Nukleotidsequenz die generelle Formel (X1)(S1)(X2)(S2)(X3)(S3)(X4) auf, worin X1 die erste Ziel-Nukleotidsequenz zur Bindung des ersten chimären Enzyms des biosynthetischen Wegs bedeutet; worin X2 die zweite Ziel-Nukleotidsequenz zur Bindung des zweiten chimären Enzyms des biosynthetischen Wegs bedeutet; worin X3 die dritte Ziel-Nukleotidsequenz zur Bindung des dritten chimären Enzyms des biosynthetischen Wegs bedeutet; worin X4 die vierte Zielnukleotidsequenz zur Bindung des vierten chimären Enzyms des biosynthetischen Wegs bedeutet; worin S1, falls vorhanden, den ersten Nukleotidspacer bedeutet, der für eine richtige räumliche Ausrichtung der benachbarten chimären Enzyme des biosynthetischen

Wegs sorgt; worin S2, falls vorhanden, den zweiten Nukleotidspacer bedeutet, der für eine richtige räumliche Ausrichtung der benachbarten chimären Enzyme des biosynthetischen Wegs sorgt; worin S3, falls vorhanden, den dritten Nukleotidspacer bedeutet, der für eine richtige räumliche Ausrichtung der benachbarten chimären Enzyme des biosynthetischen Wegs sorgt.

In einigen anderen Ausführungsbeispielen weist die Programm-Nukleotidsequenz die Formel (X1)(S1)(X2)(S1)(X2)(S1)(X3) auf, worin X1 die erste Zielnukleotidsequenz zur Bindung des ersten chimären Enzyms des biosynthetischen Wegs bedeutet; worin X2 die zweite Zielnukleotidsequenz zur Bindung des zweiten chimären Enzyms des biosynthetischen Wegs bedeutet; worin X3 die dritte Ziel-Nukleotidsequenz zur Bindung des dritten chimären Enzyms des biosynthetischen Wegs bedeutet; worin S1, falls vorhanden, den Nukleotidspacer des gleichen Typs zwischen jedem Paar der Zielnukleotidsequenzen bedeutet. Eine nacheinander wiederholende Zielnukleotidsequenz (in diesem Fall X2) sorgt für eine höhere örtliche Konzentration des chimären Enzyms, das die Geschwindigkeit begrenzende Stufe des biosynthetischen Wegs katalysiert. Daraus ergibt sich eine höhere Effizienz des metabolischen Flusses des Wegs.

Eine nacheinander wiederholende Ziel-Nukleotidsequenz (in diesem Fall X2) sorgt für eine höhere Dimerisationsstufe (oder Multimerisationsstufe, falls sich eine spezifische Ziel-Nukleotidsequenz mehr als zweimal hintereinander wiederholt) dimerer (oder multimerer) chimärer Protein des biosynthetischen Wegs. Die nacheinander wiederholende Ziel-Nukleotidsequenz (in diesem Fall X2) sorgt auch für eine höhere Effizienz des metabolischen Flusses in biosynthetischen Wegen, wo ein Enzym auf das Substrat mehr als einmal nacheinander wirkt.

In einigen anderen Ausführungsbeispielen weist die Programm-Sequenz der Nukleinsäuren die Formel (X1)(S1)(X2)(S2)(X1)(S3)(X3) auf, worin X1 die erste Zielnukleotidsequenz zur Bindung des ersten chimären Enzyms des biosynthetischen Wegs bedeutet und wiederholt sich in der Programm-Nukleotidsequenz, aber nicht unmittelbar nach der ersten Wiederholung; worin X2 die zweite Zielnukleotidsequenz zur Bindung des zweiten chimären Enzyms des biosynthetischen Wegs bedeutet; worin X3 die dritte Ziel-Nukleotidsequenz zur Bindung des dritten chimären Enzyms des biosynthetischen Wegs bedeutet; worin S1, falls vorhanden, den ersten Nukleotidspacer bedeutet, der für eine richtige

räumliche Ausrichtung der benachbarten chimären Enzyme des biosynthetischen Wegs sorgt; worin S2, falls vorhanden, den zweiten Nukleotidspacer bedeutet, der für eine richtige räumliche Ausrichtung der benachbarten chimären Enzyme des biosynthetischen Wegs sorgt; worin S3, falls vorhanden, den dritten Nukleotidspacer bedeutet, der für eine richtige räumliche Ausrichtung der benachbarten chimären Enzyme des biosynthetischen Wegs sorgt. Diese Programm-Nukleotidsequenz kann zur Verbesserung der Effizienz der biosynthetischen Wegen eingesetzt werden, wo ein chimäres Enzym des biosynthetischen Wegs (in diesem Fall das chimäre, sich an die Zielnukleotidsequenz X1 bindende Enzym des biosynthetischen Wegs) in dem biosynthetischen Weg mehr als einmal vorkommt, aber nicht unmittelbar nacheinander.

Gestaltung der Bindungsfaktoren

Die Erfindung bezieht sich auf chimäre Polypeptide, die aus einem Nukleinsäure-Bindungsfaktor (im Folgenden NABF, eng. nucleic acid binding factor) und einem Enzym des biosynthetischen Wegs oder anderen funktionellen Polypeptid zusammengesetzt sind, die miteinander mit einem Aminosäure-Linker verbunden sind. NABF können in der Natur vorhanden sein oder künstlich gestaltet werden. Sie können aus irgendwelchem Organismus isoliert werden. NABF kann jedes Polypeptid, jede Domäne, jedes Protein oder Proteinsegment sein, das/die/der mindestens ein die Nukleinsäure erkennende Motiv umfasst. NABF wirkt mit den Nukleinsäuren derart zusammen, dass er spezifisch die Nukleotidsequenz erkennt. Beispiele von Polypeptiden mit NABF umfassen, aber nicht darauf beschränkt: Helix-Turn-Helix-Motiv, Zinkfinger, Leucin-Reißverschluss, Winged Helix, Winged Helix-Turn-Helix, Helix-Loop-Helix Motiv und HMG-Box.

Die Polypeptide mit dem Grundmotiv Helix-Loop-Helix werden zu Faktoren mit einem Leucin-Reißverschluss, Faktoren mit dem Helix-Loop-Helix-Motiv, Faktoren mit dem Helix-Loop-Helix und Leucin-Reißverschluss, NF-1, F-X und bHSH unterteilt. Sie werden durch die Struktur von zwei mit einer Schleife miteinander verbundenen α-Helices bestimmt. Die diese Domäne umfassende Transkriptionsfaktoren binden sich normalerweise in der dimeren Form an die konsensuelle palindromische Sequenz (CACGTG) auch E-Box genannt. Die Transkriptionsfaktoren bHLH binden sich auf nicht-palindromische Sequenzen, die oft dem E-Box ähnlich sind.

In einigen anderen Ausführungsbeispielen werden NABF aus der Oberklasse der DNA-Bindungsdomänen, die ein Zink-Ion koordinieren, ausgewählt. Diese Oberklasse ist in Unterklassen unterteilt: Zinkfinger Cys4 des Kernrezeptortyps, verschiedene Zinkfinger Cys4, Zinkfinger Cys2His2, Zinkfinger mit dem Kern Cys6-Zn und Zinkfinger einer anderen Zusammensetzung. Einzelne Zinkfinger-Domänen kommen normalerweise in Tandemwiederholungen mit zwei, drei oder mehreren Fingern vor, was eine DNA-Bindungsdomäne eines Proteins zusammensetzt. Diese Tandem-Verteilungen von Zinkfingern können sich in die große Furche des DNA-Moleküls binden, normalerweise jede drei Basenpaare. Die α-Helix jeder Domäne kann mit den Nukleotiden im Nukleinsäuremolekül Wechselwirkungen bilden, die spezifisch für die Nukleotidsequenz sind; die Aminosäureresten der o.g. α-Helix können mit vier oder mehreren Nukleotiden Wechselwirkungen bilden, womit sich ein überlappendes Muster von Wechselwirkungen mit den anliegenden Zinkfingern bildet.

In einigen anderen Ausführungsbeispielen sind NABF aus der Oberklasse mit dem Helix-Loop-Helix-Motiv ausgewählt, die sich in sechs Klassen unterteilt: Homeodomäne, "fork head / winged helix", Hitze-Schock Faktoren, Tryptophan-Cluster und TEA-Domäne. Helix-Loop-Helix ist das wichtigste Strukturmotiv, das sich an DNA binden kann. Beide α-Helices erkennen und binden ein DNA-Molekül, eine Helix befindet sich am N- und die andere am C-Ende des Motives.

Die vierte Oberklasse von NABF besteht aus beta-Gerüst Faktoren, die mit der kleinen Furche zusammenwirken. Diese wird weiter in die folgenden Klassen unterteilt: RHR, STAT, p53, MADS, Transkriptionsfaktoren mit einem beta-Fass und α-Helix, TATA-Bindungsproteine, heteromere CCAAT-Faktoren, "Grainyhead", Cold-Shock-Domain-Faktoren und "Runt".

Die anderen Transkriptionsfaktoren, die sich an spezifische Nukleotidsequenzen binden, wie zum Beispiel mutierte Restriktionsenzyme ohne Restriktionsaktivität, aber mit der Fähigkeit, die Sequenzen zu erkennen, sind ebenso wichtig.

Diese Erfindung bezieht sich, doch nicht beschränkend, auf Polypeptide mit der Zinkfingerdomäne, die eine kleinere, unabhängig gefaltete Domäne bildet, die ein Zink-Ion enthält und eine spezifische Nukleotidsequenz erkennt. Zinkfingerdomänen kommen

normalerweise als Tandemwiederholungen mit mindestens zwei, drei oder mehreren Zinkfingern vor, die zusammen die DNA-Bindungsdomäne des Proteins bilden. Jeder Finger erkennt und bindet sich an die aus drei Basenpaaren zusammengesetzten Stellen. Für eine spezifische Bindung sind die Aminosäureresten an Stellen 1, 2, 3 und 6 in der α-Helix im Inneren der kleineren Zinkfingerdomäne verantwortlich. Mit einem modularen oder kombinatorischen Vorgehen können die wiederholten kleineren Domänen multipliziert werden und eine charakteristische Bindung mit Variationen in einigen Schlüssel-Aminosäureresten erzielt. So werden die Finger mit verschiedener Triplet-Spezifität kombiniert, um eine spezifische Erkennung von längeren Nukleotidsequenzen zu erreichen. Durch Kombinieren von zwei oder mehreren Zinkfingerdomänen kann eine nicht-spezifische Bindung an die im Wirtsorganismus vorhandenen Nukleinsäure minimiert und die Spezifität der Bindung der die Zinkfinger enthaltende Polypeptide an die Programm-Nukleinsäure gesteigert werden.

Zinkfingerdomänen sind stabile Strukturen, derer Konformation bei der Bindung an das Ziel nur selten verändert wird.

Die funktionellen Polypeptide in dieser Erfindung sind mit mindestens einer Linkersequenz verbunden, die kovalent mit jeweils einem Enzym des biosynthetischen Wegs gebunden ist. Die Linkersequenz bestimmt die Lage des Enzyms des biosynthetischen Wegs oder anderen funktionellen Polypeptids in Hinsicht auf die Zinkfingerdomäne und in Hinsicht auf das benachbarte Enzym des biosynthetischen Wegs oder anderen funktionellen Polypeptids. Die Linkersequenzen sind in verschiedenen Längen und Aminosäuresequenzen vorhanden und stellen sicher, dass ein kovalent gebundenes Enzym eine richtige Tertiärstruktur bilden kann und damit seine biologische Funktion aufbewahrt.

Die Anzahl der chimären Enzyme des biosynthetischen Wegs oder anderen funktionellen an NABF verbundenen Polypeptide beträgt, aber nicht darauf beschränkt, drei, vier, fünf oder mehr und kann sich von der Anzahl der einzelnen Ziel-Nukleotidelemente auf der Programm-Nukleotidsequenz unterscheiden. Die Anzahl der chimären Enzyme des biosynthetischen Wegs oder anderen funktionellen mit NABF verbundenen Polypeptiden hängt von der Zahl der Schritte im biosynthetischen Weg ab. Falls der biosynthetische Weg drei, vier, fünf verschiedene Enzyme oder andere funktionelle Polypeptide für die Umwandlung des Substrats zu einem Produkt oder Vorläufer benötigt, dann werden

mindestens drei, vier, fünf verschiedene Bindungselemente in einer gewissen Reihenfolge in die Programm-Nukleotidsequenz eingefügt.

In einigen Ausführungsbeispielen unserer Erfindung kann als Programm-Nukleotidsequenz ein RNA-Molekül eingesetzt werden. In diesem Fall werden NABF, die eine spezifische RNA-Sequenz erkennen, eingesetzt. Diese können werden, aber nicht darauf beschränkt: RNA-Erkennungsmotiv (RRM-, RBD, bzw. RNP-Motiv), heterogene Kern-RNP-K-homologe Domäne (KH-Domäne), Zinkfingerdomäne (am besten charakterisierten Zinkfingerdomänen, die ein RNA-Molekül binden, sind jene, die zum Typ CCHH und CCCH gehören, können aber auch andere Type eingesetzt werden) und Pumilio-Domänen (Puf).

Obwohl die RNA-Bindungsdomänen, die spezifische Sekundärstrukturen auf RNA-Moleküle erkennen, auch wie NABF (zum Beispiel, aber nicht auf S1-Domänen beschränkt) eingesetzt werden können, sind die RNA-Bindungsdomänen, die eine gewisse Sequenz (zum Beispiel, aber nicht auf Zinkfinger-RNA-Bindungsdomänen beschränkt) erkennen, am meisten geeignet. Die Erfinder schlagen die Anwendung der RNA-Bindungsdomänen, die spezifische Sequenzen der doppelsträngigen RNA erkennen, vor, weil sie eine richtige räumliche Verteilung ermöglichen. Die Methoden zur Gestaltung von Zinkfingern, die sich an eine spezifische doppelsträngige RNA-Sequenz binden, sind in der Literatur beschrieben.

Gestaltung von biosynthetischen Wegen.

Diese Erfindung bezieht sich auf die Methode zur Erhöhung der Gewinnung von biosynthetischen Produkten. Diese Methode umfasst: (a) Programm-Nukleotidsequenz; und (ii) chimäre Proteine mit einem Enzym des biosynthetischen Wegs oder anderem funktionellen Polypeptid und NABF. Die beiden benötigten Komponenten (i) und (ii) können (i) *in vitro*, also in einer Lösung in Gegenwart von Substraten und Kofaktoren für die Enzyme des biosynthetischen Wegs gekoppelt werden, oder (ii) können in Wirtszellen *in vivo* eingeführt werden und sich exprimieren.

Der biosynthetische Weg in diesem Ausführungsbeispiel bedeutet irgendwelche Enzymkaskade, die eine neue chemische Bindung auf dem Substrat in einer gewissen Reihenfolge bildet, d. h. wie sie in Natur oder künstlich vorkommt. Ein künstlich gestalteter biosynthetischer Weg bezieht sich auf die Synthese eines in der Natur unbekannten Produkts

oder die Enzyme des biosynthetischen Wegs nicht aus demselben Organismus stammen oder wurden durch Genmanipulierung gewonnen.

Die Erfindung bezieht sich auf einen biosynthetischen Weg, auf irgendwelche gekoppelte Enzymumsetzung, die sequenziell mit mindestens drei oder mehr Enzymen ausgeführt werden muss. Der biosynthetische Weg (Fluss, Effizienz) kann durch Messung der Vorläuferkonzentration und Endprodukte mit geeigneten den Fachmännern bekannten Methoden gemessen werden.

Die Erfindung bezieht sich auf irgendwelchen biosynthetischen Weg, was folgendes, aber nicht beschränkt, umfasst:

- katabolische oder anabolische Wege,
- primäre Wege, wie zum Beispiel, aber nicht beschränkt, Synthese von Aminosäuren, Fettsäuren, Kohlenhydraten, Pyrimidinen und Purinen,
- sekundäre Wege, wie zum Beispiel, aber nicht beschränkt, Synthese von Sekundärmetaboliten mit biologischer Aktivität und pharmakologischen Eigenschaften (Polyketide), von ungewöhnlichen Peptiden, Hormonen, Terpenoiden (Karotenoiden,...), Pigmenten (Violacein, Melanin, ...) usw.

Enzyme des karotenoiden biosynthetischen Wegs

Karotenoide gehören zur Kategorie von Tetraterpenoiden. Astaxanthin, Zeaxanthin und Canthaxanthin sind Derivate des β-Carotins und dessen Synthese wird durch die Enzyme des karotenoiden biosynthetischen Wegs katalysiert: crtE, crtB, crtI, crtY, crtO, crtZ, wobei die Reihenfolge und die Beteiligung von Enzymen zur Synthese von verschiedenen Produkten führen. Der biosynthetische Weg, zum Beispiel, der crtE, crtB, crtI, crtY, crtO, crtZ beinhaltet, führt zur Synthese von Astaxanthin, crtE, crtB, crtI, crtY, crtZ führen zur Synthese von Zeaxantin, und crtE, crtB, crtI, crtY, crtO führen zur Synthese von Canthaxanthin.

Biosynthetischer Weg von Violacein

Die Gene zur Biosynthese des Pigments Violacein befinden sich im Operon, das aus den Genen vioE, vioD, vioC, vioB und vioA gebildet ist. VioA ist FAD-abhängige l-

Tryptophan-Oxidase, die IPA-Imin synthetisiert. Das ist ein Substrat für das VioB-Enzym, Hemoprotein-Oxidase, das IPA-Imin in eine Verbindung mit einer unbekannten chemischen Formel (Verbindung X) umwandelt. VioE ist ein kürzlich entdecktes Protein ohne charakterisierten Homologen und ist ein Schlüsselenzym in der Synthese von Violacein. Es umwandelt die Verbindung X in Zwischenprodukte, die von den Enzymen VioD und VioC als Substrat angewendet können. Beide Enzyme sind FAD-abhängige Monooxigenasen, die diese Verbindungen hydroxylieren und dadurch zur Bildung von Violacein führen.

Biosynthetischer Weg von Resveratrol

Resveratrol (*trans* 3,5,4'-Trihydroxystilben) ist ein pflanzliches Polyphenol. Der biosynthetische Weg von Resveratrol besteht aus vier Enzymen: Phenylalanin/Ammoniak-Liase (PAL) und Zimtsäure-4-Hydroxylase, die durch das Enzym Tyrosin/Ammoniak-Liase (TAL) ersetzt werden können, 4-Coumarat/CoA-Ligase (4CL) und Stilben-Synthase (STS). Die ersten zwei Enzyme, PAL und C4H, überführen Phenylalanin in p-Cumarsäure (4-Cumarsäure). Der dritte Enzym, 4CL, verbindet p-Cumarsäure mit der Pantothengruppe des Koenzyms A (CoA), was das Produkt 4-Coumaroyl-CoA ergibt. Das letzte Enzym im Weg, STS, katalysiert die Kondensation des 4-Coumaroyl-CoA-Moleküls und von drei Malonil-CoA-Moleküle in Resveratrol.

In dieser Erfindung sind die Enzyme aus einem einzelnen biosyntethischen Weg, wie zum Beispiel aus dem biosynthetischen Weg zur Synthese von Violacein, Resveratrol oder Carotenoiden, mit NABF verkoppelt. Mit einer sequenziellen Bindung dieser Enzyme an die Programm-Nukleotidsequenz kann die Sequenz von Enzymreaktionen festgestellt werden. Dies ermöglicht schnellere und wirksamere Reaktionen und bietet die Möglichkeit der Synthese neuer Produkte durch die Umordnung der Reihenfolge von Enzymreaktionen an (z. B. falls die Reihenfolge der Enzyme 1-3-2 ist, ist das Endprodukt anders als ob die Reihenfolge 1-2-3 wäre). Der Vorteil unserer Erfindung liegt in der Möglichkeit künstliche Verbindungen mit gewünschten Eigenschaften zu synthetisieren.

Rekombinante DNA

In der Erfindung wurden herkömmliche molekularbiologische Methoden eingesetzt, die den Fachmännern in diesem Gebiet im Allgemeinen bekannt sind.

Die erfundenen Proteine können sich in einem geeigneten Wirtsorganismus durch Expression der für diese Proteine kodierenden DNA synthetisieren. Die DNA, die für diese Proteine kodiert, wird in einen geeigneten Expressionsvektor eingeführt. Geeignete Vektoren sind, aber nicht darauf beschränkt, Plasmide, Virusvektoren und anderes. Die mit dem Wirtsorganismus kompatiblen Vektoren sind den Fachmännern in diesem Gebiet bekannt und umfassen geeignete Kontrollelemente zur Transkription und Translation einer Nukleotidsequenz.

Ein Expressionsvektor kann für die Expression in prokaryotischen und eukaryotischen Zellen vorbereitet werden. Prokaryotische Zellen sind z. B. Bakterien, in erster Linie *Escherichia coli*. In der Erfindung wurden die prokaryotischen Zellen zur Produktion einer ausreichenden Menge von Nukleinsäuren eingesetzt. Im Allgemeinen beinhalten die Expressionsvektoren Kontrollelemente, die operativ mit der erfindungsgemäßen DNA, die für Proteine kodiert, verbunden sind. Die Kontrollelemente werden so ausgewählt, dass sie eine effektive und gewebespezifische Expression auslösen. Der Promotor kann entweder konstitutiv oder induzierbar sein, je nach gewünschtem Expressionsmuster. Der Promotor kann arteigen oder einer fremden Ursprung sein (ist nicht in den Zellen, in welchen er angewendet ist, befindlich). Der Promotor muss derart ausgewählt werden, dass er in Zielzellen des Wirtsorganismus wirkt. Des Weiteren werden für eine effiziente Translation der chimären Proteine auch Initiirungssignale eingesetzt, was ATG und entsprechende Sequenzen bedeutet. Wenn ein erfindungsgemäßer Vektor zwei oder mehrere Leserahmen beinhaltet, müssen die Leserahmen unabhängig operativ mit den Kontrollelementen verbunden werden und die Kontrollelemente müssen gleich oder verschieden werden, je nach gewünschter Produktion von Proteinen.

Beispiele:

Klonieren und Reinigung von Polypeptiden

Die DNA-Sequenzen der Bindungsdomänen für Nukleinsäuren und biosynthetische Enzyme oder andere oben geschriebene funktionelle Polypeptide wurden auf Basis Aminosäuresequenzen der ausgewählten Proteindomänen anhand des Designer Tools im Programm DNA 2.0 Inc. gestaltet. Das Designer Tool ermöglicht dem Anwender die Gestaltung von DNA-Fragmenten und Optimierung der Expression in gewünschten Wirtsorganismen (z. B. *E. coli*) mit der für das Organismus spezifischen Optimierung von

Kodonen. Die Gene wurden beim Unternehmen GENEART AG (Im Gewebepark B32, D93059 Regensburg) bestellt, mit Restriktionsenzymen ausgeschnitten und zu einem geeigneten Vektor, der geeignete Regulationssequenzen beinhaltet, kloniert. Dabei wurde eine Standardmethode, die den Fachmännern aus diesem Gebiet bekannt ist, verwendet. Die DNA-Sequenzen von gespaltenen Fluoreszenzproteinen wurden mit PCR multipliziert, als Matrize wurden die DNA-Sequenzen der Fluoreszenzproteine eingesetzt. Die eingesetzten Vektoren beinhalten kommerzielle Vektoren pET, pBluescript, pCDNA, pSB1A2 (pSB1A2 http://partsregistry.org/Part:pSB1A2) pSB1C3 (http://partsregistry.org/wiki/index.php?title=Part:pSB1C3), pSB1AK3 (http://partsregistry.org/wiki/index.php?title=Part:pSB1AK3), und zwei Plasmidvektoren in wenigen Kopien pSB4K5 (http://partsregistry.org/Part:pSB4K5) und pSB4C5 (http://partsregistry.org/Part:pSB4C5), die alle nötigen Eigenschaften tragen, wie z. B. Antibiotikaresistenz, Replikationsstelle und Multiple-Cloning-Site.

Molekullarbiologiemethoden (DNA-Fragmentierung mit Restriktionsenzymen, DNA-Multiplizierung mit einer Polymerase-Kettenreaktion – PCR, PCR-Ligation, DNA-Konzentration-Bestimmung, Agarosegel-Elektrophorese, Reinigung von DNA-Fragmenten aus Agarosegelen, Ligation von DNA-Fragmenten zu einem Vektor, Transformierung von chemisch kompetenten Zellen *E.coli* DH5α, Isolierung der Plasmid-DNA mit kommerziel vorhandenen Kitten, Screening und Selektion) wurden für die Zubereitung von DNA-Konstrukten eingesetzt. Alle Methoden wurden unter Sterilbedingungen (aseptische Technik) ausgeführt. DNA-Abschnitte wurden mit Restriktionsanalyse und Sequenzierung analysiert.

Klonierungsverfahren sind den Fachmännern im Gebiet gut bekannt und sind in irgendwelchem Handbuch der Molekularbiologie detailliert beschrieben.

DNA-Konstrukte und die entsprechende chimäre Proteine sind in Tabelle 1 beschrieben. Alle DNA-Konstrukte weisen ein Startkodon (ATG) vor dem Histidin-Tag bzw. der kodierenden Region auf. Die für die chimären Proteine kodierenden Konstrukte wurden in den Vektor pET19b kloniert, um eine hohe Expression zu erreichen, oder in pSB4K5 und pSB4C5, um die Produktion von Carotenoiden und Violacein *in vivo* zu verfolgen. Die Expressionskassette in Richtung 5` gegen 3` beinhaltet einen T7-Promotor, eine Multiple-Cloning-Site für Fusionsproteine und einen T7-Terminator. Diese Regulierungselemente

ermöglichen die Expression des Proteins in einer prokaryotischen Zelllinie von *E.coli*, die den genentischen Code für TN7-RNA-Polymerase trägt.

DNA-Konstrukte wurden mittels Molekularbiologiemethoden vorbereitet, die in irgendwelchem Handbuch der Molekularbiologie beschrieben und den Fachmännern aus diesem Gebiet bekannt sind. Die Plasmide sowie die End- und Zwischenkonstrukte wurden in die Zellen der Bakterie *E. coli* DH5α oder BL21 (DE3) pLysS transformiert.

Mit der Absicht, die Ausführbarkeit der Produktion von den in Tabelle 1 aufgeführten Proteinen zu beweisen, wurden mehrere Konstrukte vorbereitet.

Die für die offenen Leserahmen der Fusionsproteine in Tabelle 1 kodierende Plasmide (SEQ ID No. 52 bis SEQ ID No. 57) wurden mit chemischer Transformation in kompetente Zellen von *E.coli* BL21 (DE3) pLysS transformiert. Die ausgewählten Kolonien von Bakterien, die auf LB-Platten mit dem entsprechenden Antibiotikum (Ampicillin) gewachsen sind, wurden in 10 ml LB-Brühe unter Zugabe eines entsprechenden Antibiotikums inokuliert. Nach mehrstündiger Inkubierung bei 37°C wurden 10-100 µl aus der Kultur in 100 ml des ausgewählten Nährbodens inokuliert und bei 37°C über Nacht geschüttelt. Die Nachtkultur wurde 20-50-mal verdünnt bis OD600 zwischen 0,1 und 0,2. Die Erlenmeyerkolben mit 500 ml der verdünnten Kultur wurden bei 25°C geschüttelt bis OD600 zwischen 0,6 und 0,8, als die Proteinexpression mit der Zugabe von 1 mM von IPTG induziert wurde. Vier Stunden nach der Induktion wurde die Kultur zentrifugiert und die Bakterienzellen in einem Lysepuffer (10 mM Tris pH 8,0, 1M NaCl, 0,1% Desoxycholat, 0,1 mM Zinksulfat und Zugabe eines Protease-Inhibitor-Cocktails) resuspendiert und über Nacht bei -80°C gefroren. Die aufgetauten Zellsuspensionen wurden dann mit Beschallung lysiert und zentrifugiert. Der Niederschlag (Zellmembrane, Inklusionskörperchen) und der Überstand wurden an Expression von Konstrukten mit SDS-PAGE und Western-Blot getestet, als primäre Antikörper wurden anti-His-Tag-Antikörper eingesetzt, wenn nötig. Die Fusionsproteine wurden meistens im unlöslichen Teil (Inklusionskörperchen), der > 80 % des ausgewählten Proteins beinhaltete, zu finden. Die Inklusionskörperchen wurden zweimal mit dem Lysepuffer, zweimal mit 1 M Urea und 10 mM Tris pH 8,0 und zweimal mit 2M Urea und 10 mM Tris pH 8,0 abgewaschen. Normalerweise resultierte ein solches Verfahren in >95% Reinheitsgrad des Proteins. Die gereinigten Inklusionskörperchen wurden in 8 M Urea und 50 mM Tris pH 8,0 gelöst und über eine Ni-NTA-Säule gereinigt. Reinigung unter

Denaturierungsbedingungen wurde nach Herstelleranleitung ausgeführt. Die Proteine wurden mit 8 M Urea, 50 mM Tris pH 8,0 und 250 mM Imidazol abgewaschen. Die Neuproteinfaltung erfolgte mit einer Dialyse gegen dem Puffer mit 50 mM Tris pH 8,0, 500 mM NaF, 500 µM $ZnSO_4$, 5 mM DTT, 0,005% Tween und 10% Glycerol. Dialyse wurde mindestens zweimal für 4 Stunden bei 4°C ohne Schütteln ausgeführt. Nach der Beendung des Verfahrens wurde das Dialysat vorsichtig aus den Dialysenröhrchen in Zentrifugengläser übertragen.

Die Proteine, die sich nicht erneut gefaltet haben, haben beim Zentrifugieren bei 10000 U/m ausgefällt und so von der erneut gefalteten Proteinen in der löslichen Phase abgeschieden. Die Konzentration an Proteinen wurde mit Spektrophotometrie und mit dem Bradford-Test gemessen, die Proteingröße wurde mit SDS-PAGE/Commasie-Färbung und Western-Blot mit anti-His-Tag-Antikörpern bestimmt.

Tabelle 1: Zusammensetzung von Plasmiden, die zur Darstellung der Erfindung vorbereitet wurden

Seq ID No. DNA	**Seq ID No. Protein**	**Konstruktzusammensetzung**	**Plasmid-Rückgrat**
		Bindungsdomänen	
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 der biosynthetischen Wege oder andere funktionelle Polypeptide	
17	18	vioA	pSB1A2
19	20	vioB	pSB1A2
39	40	vioC	pSB1A2
41	42	vioD	pSB1A2
43	44	vioE	pSB1A2
45	46	nCFP (N-Terminus mCerulean)	pSBAK3
47	48	nYFP (N-Terminus mCitrine)	pSBAK3
49	50	cCFP (C-Terminus mCerulean)	pSBAK3
51	52	cYFP (C-Terminus mCitrine)	pSBAK3
		Hilfs-Polypeptide	
53	54	Linkerpeptid	pSBAK3

55		Histidin-Tag	pSBAK3
		Programm-Sequenzen von Nukleinsäuren	
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-Linkerpeptid-vioA	pSB1A2
63	64	Znf_Zif268-Linkerpeptid-vioB	pSB1A2
65	66	Znf_PBSII-Linkerpeptid-vioE	pSB1A2
67	68	Znf_HivC-Linkerpeptid-vioD	pSB1A2
69	70	Znf_Gli1-Linkerpeptid-vioC	pSB1A2
71	72	His_tag-Znf_Gli1-Linkerpeptid-nCFP	pSBAK8
73	74	His_tag-Znf_Blues-Linkerpeptid-nCFP	pSBAK8
75	76	His_tag-Znf_PBSII-Linkerpeptid-nYFP	pSBAK8
77	78	cCFP-Linkerpeptid-Znf_HIVC-His_tag	pSBAK8
79	80	cCFP-Linkerpeptid-Znf_Jazz-Linkerpeptid-His_tag	pSBAK8
81	82	cYFP-Linkerpeptid-Znf_Zif268-Linkerpeptid -His_ tag	pSBAK8

Bindung von Bindungsdomänen an ein Zielelement der Nukleinsäure

Methode der Messung der Aktivität des β-Galaktosidase-Enzyms

Bakterienkulturen mit den in Tabelle 1 aufgeführten Plasmidkonstrukten wurden in 10 ml LB-Kulturmedium mit der Zugabe von 5 µl 1M IPTG (Isopropyl-β-D-thio-Galactosid) und steigenden Konzentrationen (0%, 0,0025%, 0,1% und 1%) von L-Arabinose inokuliert und im Schüttler bei 180 U/m und 37°C für 18 Stunden inkubiert. Die Bakteriendichte wurde mit Messung der optischen Dichte (OD600) bestimmt. Die Messungen der β-Galaktosidaseaktivität wurden im auf 28°C vorgewärmten ELISA-Leser ausgeführt. 5 µl jeder Kultur wurden in die Näpfchen der 96-teiligen Mikrotiterplatte mit einem Klarboden übertragen, dann wurden 100 µl von Z-Puffer mit Chloroform (Z-Puffer: 0,06 M Na2HPO4 x 7H20, 0,04 M NaH2PO4 x H20, 0,1 M KCl, 0,001 M MgSO4x7H2O, pH 7; Z-Puffer mit Chloroform: Z-Puffer, 1% β-Mercaptoethanol, 10% Chloroform) zugefügt. Bakterienzellen wurden mit einer Zugabe von 50 µl Z-Puffer mit SDS (Z-Puffer, 1.6 SDS) lysiert, dann erfolgte eine 10-Minuten-Inkubation bei 28°C. 50 µl der 0,4% ONPG-Lösung in Z-Puffer wurden zu jedem Näpfchen zugegeben und die Unterschiede in OD405 wurden in 30-Sekunden-Intervale für 20 Minuten gemessen.

Ergebnisse

Um es herauszufinden, ob sich die Transkriptionsfaktor (NABF) (Zinkfinger Zif 268, Blues; PBSII, HivC) und His-Nictal an entsprechende DNA-Ziele auch *in vivo* binden, haben wir ein Reportersystem gestaltet. Der Reporter beinhaltet ein Plasmid, zusammengebaut aus a) lacZ unter einem synthetischen Promotor, der eine DNA-Bindungsstelle für ein gewisses DNA-Protein enthält und b) einem entsprechenden DNA-Bindungsprotein unter dem Arabinosepromotor. Eine erfolgreiche NABF-Bindung an den synthetischen Promotor verhindert die Transkription von lacZ, was in einer niedrigeren β-Galactosidaseaktivität resultiert. Die Kulturen von *E. coli* mit Plasmiden mit verschiedenen NABF sind über Nacht im LB-Medium mit steigernden Konzentrationen von Arabinose gewachsen. Die β-Galactosidaseaktivität wurde wie im Text beschrieben gemessen. Es hat sich für alle getesteten Konstrukte erwiesen, dass sich die β-Galactosidaseaktivität verhältnismäßig mit der steigernden Konzentration von Arabinose vermindert hat. Daraus geht hervor, dass sich NABF (Tabelle 1) an das Ziel-DNA-Element im inneren des synthetischen Promotors in vivo bindet (Figur 3A). Des Weiteren, die β-Galactosidaseaktivität blieb in Gegenwart eines nicht-geeigneten NABF und DNA-Ziels (z. B. Blues Zielelement Blues_O wurde durch PBSII Zielelement PBSII_O, und HivC_O Zielelement durch Gli1_O Zielelement ersetzt) ungeändert, womit die Spezifität unseres NABF-Testsystems nachgewiesen wurde (Figur 3B).

Oberflächenplasmonenresonanz (SPR)

Die SPR-Analyse wurde auf Biacore T100 (Ge Healthcare) ausgeführt. Series S-Sensor-Chip SA (Ge Healthcare) wurde mit einer Biotinylierung-einsträngigen-DNA-Probe bei einer Konzentration von 100 nM immobilisiert, womit die Endkonzentration von etwa 300 RU erreicht wurde. Hybridisierung der komplementären DNA (Programm-DNA-Sequenz mit Bindungsstellen für Zinkfinger HivC, Zif268, Jazz, Blues, PBSII, Gli1) wurde mit Einspritzen von 0,5-2 µM in Laufpuffer (10 mM HEPES, 150 mM NaCl, 0,1mM EDTA, 0,005% P20, pH 7,4) für zu 300 s ausgeführt, womit die Endantwort von etwa 300 RU erreicht wurde. Die Bindung des Analyts wurde mit Einspritzen von verschiedenen Analytkonzentrationen für 1 Minute verfolgt, danach folgte eine Dissoziation von 5 Minuten. Die Regeneration der Chipoberfläche wurde mit zwei 30 s Einspritzungen von 50 mM NaOH und einer 24 s

Einspritzung von 0,5 % SDS erreicht. Dann wurde die neue DNA eingespritzt, womit eine neue Bindungsoberfläche besorgt wurde.

Series S-Sensorchip CM5 (Ge Healthcare) wurde mit 3000 RU von Avidin durch die Bindung an den Aminosäurerest nach Herstellerprotokoll immobilisiert. Die carboxymethylierte Oberfläche des CM5-Chips wurde mit einem 7 min Einspritzungsimpuls 1:1 NHS:EDC aktiviert. Avidin auf Na-Acetat bei pH 5,5 wurde dann in mehreren kurzen Impulsen bis zur Endantwort von 3000 RU eingespritzt. Avidin wurde nur in die zweite Durchflusszelle gekoppelt, da die erste als Referenzoberfläche zur Abzählung einer unspezifischen Bindung der Analyte an Dextranmatrix des Sensorchips diente. Die nichtreagierten Stellen an der Sensoroberfläche wurden mit einem 7 Min Einspritzungsimpul von 1 M Ethanolamin (pH 8,5) blockiert.

Ergebnisse:

In vitro-Bindung von chimären Proteine, von NABF-gebundenen biosynthetischen Enzymen auf Programm-DNA (Figur 4). Figur 4A stellt die Bindung der Programm-DNA, dann des chimären Gli1 NABF-enthaltenden Proteins und Regeneration des Substrats-Chips dar. DNA-Hybridisierung wurde mit bis 300 s Einspritzung von 0,5-2 µM bis zur Endantwort 300 RU ausgeführt. Die Bindung des Analyts wurde mit Einspritzen von verschiedenen Analytkonzentrationen für 1 Minute verfolgt, danach folgte eine Dissoziation von 5 Minuten. Die Regeneration wurde mit zwei 30 s Einspritzungen von 50 mM NaOH und einer 24 s Einspritzung von 0,5 % SDS erreicht.

Die Bindung des Analyts auf die Programm-DNA und die Bindung einer frischen DNA für jedes Protein sind auf Figur 4B dargestellt. Analyte von oben: Gli1, Zif268, Blues, ZNF_HivC, PBSII, Jazz.

Eine sequentielle NABF-Bindung auf die Programm-DNA ist in Figur 4C dargestellt. Die DNA wurde am Cyclusanfang erfasst und die Proteine für 1 Min hintereinander in zwei verschiedenen Reihenfolgen eingespritzt. Durchgezogene Linie: Znf_Jazz, Znf Blues, Znf_Zif268, Znf_PBSII, Znf_HivC und Znf_Gli1; unterbrochene Linie: Znf_Gli1, Znf_PBSII, Znf_HivC, Znf_Jazz, Znf_Blues und Znf_Zif268. Die Bindung von allen Analyten in der Sequenz hat erwiesen, dass die Analytenbindung am schwächsten war, ob Glil1 als erster eingespritzt wurde.

EMSA – Electrophoretic Mobility Shift Assay

Spezifische DNA-Bindung von synthetischen Domänen aus Zinkfingern an Programm-Nukleinsäure wurde mit dem Electrophoretic Mobility Shift Assay getestet. 1 µg von gereinigten Proteinen wurde mit verschiedenen Mengen der Programm-Nukleinsäure – 500 ng, 750 ng, 1000 ng – für 3 Stunden inkubiert. Die Proben, die mit Laborwasser hohen Reinheitsgrades bis zu 20 µl verdünnt wurden, wurden auf 2,0% Agarosegel aufgetragen, das im Voraus mit Ethidium-Chlorid gefärbt wurde. Die Elektrophorese der Proben erfolgte für 40 Minuten bei konstanter Spannung von 70 V. Die Nukleinsäure-Protein-Komplexe wurden unter UV-Licht sichtbar.

Ergebnisse:

Die Migration der mit Ethidium-Chlorid gefärbten Nukleinsäure hieng von der Gegenwart des NABFs ab (d. h. DNA-Bindungsdomäne aus Zinkfingern), was eine langsamere Migration im Vergleich mit nur Programm-Nukleinsäure bedeutete (Figur 5). Diese Angaben stellen die NABF-Bindung an spezifische Zielsequenzen im DNA-Programm dar (Figur 5).

Konfokale Mikroskopie

Zur Ermittlung von chimären split-Fluoreszenzproteine (d. h. nCFP-gebundenes NABF, cCFP-gebundenes NABF) und zur Rekonstitution von FRET auf der Programm-Nukleinsäure *in vivo* wurde das Leica TSC SP5 Konfokalmikroskop eingesetzt. Die transfektierten HEK293-Zellen, die über Nacht in einem Mikroskopnäpfchen bei 37°C kultiviert wurden, wurden auf das Objektiv gesetzt und ein Tropfen Immersionsöls wurde zugefügt. Die Zellen wurden mit 433 nm für CFP und mit 515 nm für YFP angeregt. FRET wurde anhand der Methodologie des Mikroskopherstellers bestimmt (FRET AB und FRET SE Assistent in Leica LAS AF-Programmpaket).

Split, FRET-Experimente

Die HEK293-Zellen wurden mit einem Säugetierexpressionsvektor mit Genen für Split-Fluoreszenzfusionsproteine unter dem CMV-Promotor transfektiert. Transfektion erfolgte mit dem jetPEI™-Transfeksionsreagent und die Zellen wurden schließlich mit 4%-igem Paraformaldehyd fixiert. Rekonstitution von Split-Fluoreszenzproteinen, sowie das FRET-Effekt, wurden unter dem Konfokalmikroskop (unter „Konfokalne Mikroskopie“ beschrieben)

beobachtet. CFP (Znf_Gli1-Linkerpeptid-nCFP und cCFP-Linkerpeptid- Znf_HivC) und YFP (Znf_PBSII-Linkerpeptid-nYFP und cYFP-Linkerpeptid-Znf_Zif268) wurden nur in Gegenwart einer spezifischen, doch nicht zufälligen, Programm-Nukleinsäure rekonstituiert.

Ergebnisse:

Split-Fluoreszenzproteine binden sich an die benachbarten Zielnukleotidsequenzen, wenn die Plasmiden mit den Genen für Split-Fluoreszenzproteine gleichzeitig mit dem Plasmid, das die Ziel-Programmnukleinsäure trägt, transfektiert sind. Da die Split-Fluoreszenzproteine nicht in der Lage sind, einen funktionellen Chromophor *in vivo* zufällig zu bilden, ist das Fluoreszenzsignal eine Folge der Bindung des chimären Proteins an die DNA-Programm-Nukleotidsequenz.

Rekonstituierung des Split-GFP. Figur 6 zeigt *in vivo* Rekonstituierung von zwei nicht-funktionellen Luminophors (Split-GFPs), die an NABF nur in Gegenwart der Programm-Nukleotidsequenz gebunden sind (Figur 6a).

Die zwei Plasmide, je 75 ng, mit den Genen für Znf_Gli1-Linkerpeptid-nCFP und cCFP-Linkerpeptid- Znf_HIVC, wurden mit 350 ng der Programm-DNA in HEK293-Zellen ko-transfektiert. Das Emissionsignal des türkisen Fluoreszenzproteins wurde durch die Anregung der Zellen mit 433 nm Laserlicht beobachtet (Figur 6b).

Die zwei Plasmide, je 75 ng, mit den Genen für Znf_PBSII-Linkerpeptid-nYFP und cYFP-Linkerpeptid Znf_Zif268) wurden mit 350 ng der Programm-DNA in HEK293-Zellen kotransfektiert. Die YFP-Emission mit dem Peak bei 529 wurde unter dem Konfokalmikroskop beobachtet (Figur 6C).

Rekonstitierung von Split-GFPs in vitro. HEK293-Zellen wurden kotransfektiert mit Znf_Gli1-Linkerpeptid-nCFP und cCFP-Linkerpeptid-Znf_HIVC sowie Znf_PBSII-Linkerpeptid-nYFP und cYFP-Linkerpeptid Znf_Zif268. Nach 72 Stunden der Kultivierung wurden die Zellen in 150 µl von Promega Passivpuffer zur Zelllyse lysiert, dann wurden 50 µg der Programm-Nukleinsäure zu den Lysaten zugegeben. Der funktionelle Luminophor wurde in Gegenwart des DNA-Programms nach 18 Stunden der Kultivierung bei 4°C merkbar. Das wurde mit Hilfe von Lumineszenz-Spektrometer PerkinElmer LS55 detektiert.

Effizienz von FRET. Die Effizienz von FRET für vier chimäre Proteine, Split-GFPs mit NABF verbunden, in Gegenwart der Programm-DNA ist in Tabelle 2 dargestellt. FRET wurde anhand von FRET AB-Assistenten im LAS AF Programm gemessen, das Mikroskopieren auf dem Laserkonfokalmikroskop Leica TCS SP5 ermöglicht. Die Emissionssignale im CFP- und YPF-Kanal wurden eingestellt. Der Akzeptor wurde mit 515 nm Laser zerstört und die FRET-Effizient im zerstörten Bereich daraus berechnet. Die HEK293-Zellen wurden mit der Programm-DNA und Znf_Gli1-Linkerpeptid-nCFP, cCFP-Linkerpeptid-Znf_HIVC, Znf_PBSII-Linker-nYFP, Znf_cYFP-Linkerpeptid-Znf_Zif268 transfektiert. Für die negative Kontrolle wurden die Zellen mit Zytosol-CFP und –YFP ohne DNA-Bindungsfaktoren transfektiert und für die positive Kontrolle mit CFP, das mit dem YFP mit dem Linkerpeptid verbunden wurde.

Tabelle 2: Effizienz von FRET von vier Split-GFPs, jeweils mit einem anderen DNA-Bindungsfaktor verbunden.

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

Synthese von Violacein

Die Übernacht-Kulturen von *E. coli* mit den für die chimären Proteine kodierenden Plasmiden (s. Tabelle 1, Kombinationen von Konstrukten Znf_Blues-Linkerpeptid-vioA, Znf_Zif268-Linkerpeptid-vioB, Znf_PBSII-Linkerpeptid-vioE, Znf_HivC-Linkerpeptid-vioD, Znf_Gli1-Linkerpeptid-vioC) mit oder ohne Plasmid, das das DNA-Programm kodiert (s. Tabelle 1, Biosynthese-Programm (123456) oder Biosynthese-Programm (341256)) wurden zum frischen Luria Bertani Kulturmedium bis OD600 0,05 verdünnt und in Gegenwart von geeigneten Antibiotika bei 37°C kultiviert. Proben der Fermentationsbrühe wurden in verschiedenen Zeiten während des Wachstums entnommen (insgesamt 3 ml). Der repräsentative Zeitpunkt (17,5 h) ist aus Figur 9 ersichtlich. Die Bakterien wurden mit der Zugabe des gleichen Volumens (1,5 ml) von 10% SDS lysiert und Violacein wurde mit Ethyl-Acetat (1:1 vol. vol.) extrahiert. Nach einem kurzen Vortexieren wurde die organische Phase gesammelt und das Absorptionsspektrum bei 575 nm gemessen. Die Proben wurden auf TLC analysiert und die quantitative Analyse erfolgte mit Densitometrie bei 575 nm nach einer unmittelbaren Spektrumanalyse mit HPTLC. Violacein wurde ferner auch mit

Massenspektrometrie identifiziert. Die Menge der Violacein-Produktion mit oder ohne Programm-DNA oder in Gegenwart eines gemischten DNA-Programms wurde verglichen (Figur 9).

Obwohl die vorliegende Erfindung in Bezug auf spezifische erfindungsgemäßen Ausführungsformen dargestellt wurde, versteht sich jedoch, dass verschiedenartige Modifikationen vorgenommen und Äquivalente substituiert werden können, ohne von dem Sinne und der Umfang der Erfindung abzuweichen. Ebenso können viele Modifikationen vorgenommen werden mit der Absicht der Anpassung an gewisse Situation, Material, Materialzusammensetzung, Prozess, Prozessstufen oder Schritte zum Sinne und Umfang der Erfindung.

SEQUENZAUFLISTUNG

<110> KEMIJSKI INSTITUT
EN-FIST CENTER ODLICNOSTI

<120> Verbesserte Synthese eines biosynthetischen Produkts mit einer angeordneten Zusammensetzung der biosyntethischen 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 Gln 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
```

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

<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

<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

<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

```
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
```

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                  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 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
```

```
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 ggt ggc ggt 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 ggt      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 ggt      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 ggt aaa cag gcg 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 ggt ggt 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 ggt 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 ggt 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 ggt 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 ggt ggg ggt 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
```

```
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
```

```
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
```

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

<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       48
Met Lys His Ser Ser Asp Ile Cys Ile Val Gly Ala Gly Ile Ser Gly
1               5                  10                  15

ttg acg tgc gca agc cat ctg ctg gac agc ccg gca tgc cgt ggt ctg       96
Leu Thr Cys Ala Ser His Leu Leu Asp Ser Pro Ala Cys Arg Gly Leu
            20                  25                  30

agc ctg cgt atc ttt gac atg cag caa gaa gcc ggt ggc cgt atc cgc      144
Ser Leu Arg Ile Phe Asp Met Gln Gln Glu Ala Gly Gly Arg Ile Arg
        35                  40                  45

agc aaa atg ctg gat ggt aag gca agc att gaa ctg ggc gca ggt cgc      192
Ser Lys Met Leu Asp Gly Lys Ala Ser Ile Glu Leu Gly Ala Gly Arg
    50                  55                  60

tac tcc cct cag ttg cac ccg cat ttc caa agc gca atg cag cac tat      240
Tyr Ser Pro Gln Leu His Pro His Phe Gln Ser Ala Met Gln His Tyr
65                  70                  75                  80

agc caa aag agc gaa gtc tat ccg ttc acc cag ttg aag ttc aaa tct      288
Ser Gln Lys Ser Glu Val Tyr Pro Phe Thr Gln Leu Lys Phe Lys Ser
                85                  90                  95

cac gtg cag caa aag ctg aag cgc gcc atg aat gaa ctg tcc ccg cgt      336
His Val Gln Gln Lys Leu Lys Arg Ala Met Asn Glu Leu Ser Pro Arg
            100                 105                 110

ctg aaa gag cat ggt aaa gag agc ttt ttg cag ttt gtc agc cgt tat      384
Leu Lys Glu His Gly Lys Glu Ser Phe Leu Gln Phe Val Ser Arg Tyr
        115                 120                 125

caa ggt cac gat agc gcg gtt ggt atg atc cgc tct atg ggt tac gac      432
Gln Gly His Asp Ser Ala Val Gly Met Ile Arg Ser Met Gly Tyr Asp
    130                 135                 140

gca ctg ttc ctg ccg gat atc agc gca gaa atg gcc tac gac att gtg      480
Ala Leu Phe Leu Pro Asp Ile Ser Ala Glu Met Ala Tyr Asp Ile Val
145                 150                 155                 160

ggt aag cac ccg gag atc cag agc gtg acg gac aac gac gcg aac caa      528
Gly Lys His Pro Glu Ile Gln Ser Val Thr Asp Asn Asp Ala Asn Gln
                165                 170                 175

tgg ttt gca gcg gaa acg ggc ttt gct ggt ctg att cag ggc atc aag      576
Trp Phe Ala Ala Glu Thr Gly Phe Ala Gly Leu Ile Gln Gly Ile Lys
            180                 185                 190

gct aag gtt aag gcg gca ggt gcg cgt ttt agc ctg ggt tat cgt ctg      624
Ala Lys Val Lys Ala Ala Gly Ala Arg Phe Ser Leu Gly Tyr Arg Leu
        195                 200                 205

ctg agc gtc cgt acc gac ggt gac ggc tac ctg ctg caa ctg gca ggt      672
Leu Ser Val Arg Thr Asp Gly Asp Gly Tyr Leu Leu Gln Leu Ala Gly
    210                 215                 220

gac gac ggc tgg aaa ctg gag cac cgt acc cgc cat ctg att ctg gcg      720
Asp Asp Gly Trp Lys Leu Glu His Arg Thr Arg His Leu Ile Leu Ala
225                 230                 235                 240

att ccg ccg agc gcg atg gcg ggt ttg aat gtt gat ttt cca gaa gcc      768
Ile Pro Pro Ser Ala Met Ala Gly Leu Asn Val Asp Phe Pro Glu Ala
```

```
                245                 250                 255

tgg tcc ggt gcg cgc tat ggc agc ctg ccg ctg ttt aag ggc ttt ctg      816
Trp Ser Gly Ala Arg Tyr Gly Ser Leu Pro Leu Phe Lys Gly Phe Leu
            260                 265                 270

acg tac ggt gag ccg tgg tgg ttg gac tac aaa ctg gac gat cag gtg      864
Thr Tyr Gly Glu Pro Trp Trp Leu Asp Tyr Lys Leu Asp Asp Gln Val
        275                 280                 285

ctg att gtt gac aac ccg ctg cgc aaa atc tat ttc aaa ggc gat aag      912
Leu Ile Val Asp Asn Pro Leu Arg Lys Ile Tyr Phe Lys Gly Asp Lys
    290                 295                 300

tac ctg ttc ttc tat acc gat agc gag atg gcg aat tac tgg cgc ggt      960
Tyr Leu Phe Phe Tyr Thr Asp Ser Glu Met Ala Asn Tyr Trp Arg Gly
305                 310                 315                 320

tgt gtc gcg gag ggc gag gac ggt tac ctg gag caa att cgc acc cat     1008
Cys Val Ala Glu Gly Glu Asp Gly Tyr Leu Glu Gln Ile Arg Thr His
                325                 330                 335

ttg gct agc gca ctg ggt atc gtc cgt gaa cgt atc ccg caa ccg ctg     1056
Leu Ala Ser Ala Leu Gly Ile Val Arg Glu Arg Ile Pro Gln Pro Leu
            340                 345                 350

gca cac gtt cac aag tat tgg gcg cac ggc gtt gag ttt tgc cgt gat     1104
Ala His Val His Lys Tyr Trp Ala His Gly Val Glu Phe Cys Arg Asp
        355                 360                 365

tct gat att gac cac ccg agc gca ctg tct cat cgc gac agc ggt atc     1152
Ser Asp Ile Asp His Pro Ser Ala Leu Ser His Arg Asp Ser Gly Ile
    370                 375                 380

atc gcg tgc tcc gat gcg tac acg gag cat tgt ggt tgg atg gag ggc     1200
Ile Ala Cys Ser Asp Ala Tyr Thr Glu His Cys Gly Trp Met Glu Gly
385                 390                 395                 400

ggt ctg ctg agc gcc cgt gag gca agc cgt ctg ctg ttg cag cgt atc     1248
Gly Leu Leu Ser Ala Arg Glu Ala Ser Arg Leu Leu Leu Gln Arg Ile
                405                 410                 415

gcc gcg tga                                                        1257
Ala Ala
```

<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 Ile 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
```

325 330 335

Leu Ala Ser Ala Leu Gly Ile Val Arg Glu Arg Ile Pro Gln Pro Leu
340 345 350

Ala His Val His Lys Tyr Trp Ala His Gly Val Glu Phe Cys Arg Asp
355 360 365

Ser Asp Ile Asp His Pro Ser Ala Leu Ser His Arg Asp Ser Gly Ile
370 375 380

Ile Ala Cys Ser Asp Ala Tyr Thr Glu His Cys Gly Trp Met Glu Gly
385 390 395 400

Gly Leu Leu Ser Ala Arg Glu Ala Ser Arg Leu Leu Leu Gln Arg Ile
405 410 415

Ala Ala

<210> 19
<211> 2994
<212> DNA
<213> künstliche Sequenz

<220>
<223> VioB

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

<400> 19

atg agc att ctg gat ttc ccg cgt atc cac ttc cgt ggc tgg gcc cgt 48
Met Ser Ile Leu Asp Phe Pro Arg Ile His Phe Arg Gly Trp Ala Arg
1 5 10 15

gtc aat gcg ccg acc gcg aac cgc gat ccg cac ggc cac atc gat atg 96
Val Asn Ala Pro Thr Ala Asn Arg Asp Pro His Gly His Ile Asp Met
20 25 30

gcc agc aat acc gtg gcg atg gcg ggt gag ccg ttc gac ctg gca cgc 144
Ala Ser Asn Thr Val Ala Met Ala Gly Glu Pro Phe Asp Leu Ala Arg
35 40 45

cat cct acg gag ttc cac cgt cac ctg cgc tcc ctg ggt ccg cgc ttc 192
His Pro Thr Glu Phe His Arg His Leu Arg Ser Leu Gly Pro Arg Phe
50 55 60

ggc ttg gat ggt cgt gct gac ccg gaa ggc ccg ttc agc ctg gcc gag 240
Gly Leu Asp Gly Arg Ala Asp Pro Glu Gly Pro Phe Ser Leu Ala Glu
65 70 75 80

ggc tac aac gct gcc ggt aac aac cac ttt tcg tgg gag agc gca acc 288
Gly Tyr Asn Ala Ala Gly Asn Asn His Phe Ser Trp Glu Ser Ala Thr
85 90 95

gtt agc cac gtg caa tgg gat ggc ggt gag gcg gat cgt ggt gac ggt 336
Val Ser His Val Gln Trp Asp Gly Gly Glu Ala Asp Arg Gly Asp Gly
100 105 110

```
ctg gtc ggt gct cgt ttg gca ctg tgg ggt cac tac aat gat tat ctg      384
Leu Val Gly Ala Arg Leu Ala Leu Trp Gly His Tyr Asn Asp Tyr Leu
        115                 120                 125

cgt acc acc ttc aat cgt gct cgt tgg gtc gac agc gac ccg acg cgc      432
Arg Thr Thr Phe Asn Arg Ala Arg Trp Val Asp Ser Asp Pro Thr Arg
    130                 135                 140

cgt gac gct gca caa atc tat gcg ggc caa ttc acc att agc ccg gct      480
Arg Asp Ala Ala Gln Ile Tyr Ala Gly Gln Phe Thr Ile Ser Pro Ala
145                 150                 155                 160

ggt gcc ggt ccg ggt acg ccg tgg ctg ttt acg gca gac att gat gat      528
Gly Ala Gly Pro Gly Thr Pro Trp Leu Phe Thr Ala Asp Ile Asp Asp
                165                 170                 175

agc cat ggt gca cgt tgg acg cgt ggc ggc cac att gca gag cgt ggc      576
Ser His Gly Ala Arg Trp Thr Arg Gly Gly His Ile Ala Glu Arg Gly
            180                 185                 190

ggc cac ttc ttg gat gaa gag ttt ggt ctg gca cgc ctg ttt cag ttc      624
Gly His Phe Leu Asp Glu Glu Phe Gly Leu Ala Arg Leu Phe Gln Phe
        195                 200                 205

tct gtg ccg aaa gat cac cca cat ttt ctg ttt cac ccg ggt ccg ttt      672
Ser Val Pro Lys Asp His Pro His Phe Leu Phe His Pro Gly Pro Phe
    210                 215                 220

gat tcc gag gcc tgg cgt cgt ctg caa ttg gct ctg gag gat gac gac      720
Asp Ser Glu Ala Trp Arg Arg Leu Gln Leu Ala Leu Glu Asp Asp Asp
225                 230                 235                 240

gtt ctg ggt ctg acc gtg caa tat gcg ttg ttc aat atg agc acc ccg      768
Val Leu Gly Leu Thr Val Gln Tyr Ala Leu Phe Asn Met Ser Thr Pro
                245                 250                 255

cct cag ccg aac agc ccg gtt ttt cac gat atg gtc ggt gtt gtc ggt      816
Pro Gln Pro Asn Ser Pro Val Phe His Asp Met Val Gly Val Val Gly
            260                 265                 270

ctg tgg cgt cgt ggt gaa ctg gcg agc tac ccg gct ggt cgt ctg ctg      864
Leu Trp Arg Arg Gly Glu Leu Ala Ser Tyr Pro Ala Gly Arg Leu Leu
        275                 280                 285

cgt ccg cgt caa ccg ggt ctg ggt gac ctg acc ctg cgc gtc aac ggt      912
Arg Pro Arg Gln Pro Gly Leu Gly Asp Leu Thr Leu Arg Val Asn Gly
    290                 295                 300

ggt cgc gtt gcg ctg aat ttg gcg tgt gcc att ccg ttc agc act cgt      960
Gly Arg Val Ala Leu Asn Leu Ala Cys Ala Ile Pro Phe Ser Thr Arg
305                 310                 315                 320

gcc gcg cag cca agc gca ccg gac cgc cga ccc cgg acc tgg gtg cca     1008
Ala Ala Gln Pro Ser Ala Pro Asp Arg Arg Pro Arg Thr Trp Val Pro
                325                 330                 335

aac tgc cgc tgg gcg atc tgc tgc tgc gtg atg agg acg gcg cac tgt     1056
Asn Cys Arg Trp Ala Ile Cys Cys Cys Val Met Arg Thr Ala His Cys
            340                 345                 350

tgg cac gtg tgc cgc agg ctc tgt acc aag act att gga cga atc acg     1104
Trp His Val Cys Arg Arg Leu Cys Thr Lys Thr Ile Gly Arg Ile Thr
        355                 360                 365

gta ttg tgg acc tgc cgc tgc tgc gcg aac cgc gtg gta gct tga ccc     1152
Val Leu Trp Thr Cys Arg Cys Cys Ala Asn Arg Val Val Ala     Pro
    370                 375                 380
```

```
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 Met Ala Val Phe Val Ser Thr Ala Gly Asn Cys Ser     Arg Asn
650                 655                 660

gat tca tta cct ggt cgt taa caa cct gct gat ggc cct ggg cga gcc      2064
Asp Ser Leu Pro Gly Arg     Gln Pro Ala Asp Gly Pro Gly Arg Ala
665                 670                 675

gtt cta cgc ggg tgt ccc gct gat ggg cga agc ggc acg tca ggc gtt      2112
Val Leu Arg Gly Cys Pro Ala Asp Gly Arg Ser Gly Thr Ser Gly Val
680                 685                 690                 695

tgg cct gga cac cga gtt cgc tct gga acc gtt tag cga aag cac gct      2160
Trp Pro Gly His Arg Val Arg Ser Gly Thr Val     Arg Lys His Ala
                700                 705                 710

ggc acg ttt tgt tcg tct gga atg gcc gca ctt tat ccc agc acc ggg      2208
Gly Thr Phe Cys Ser Ser Gly Met Ala Ala Leu Tyr Pro Ser Thr Gly
                715                 720                 725

caa atc cat cgc gga ctg cta tgc cgc cat tcg tca ggc gtt ttt gga      2256
Gln Ile His Arg Gly Leu Leu Cys Arg His Ser Ser Gly Val Phe Gly
            730                 735                 740

tct gcc gga ctt gtt tgg tgg cga ggc agg taa gcg tgg cgg tga aca      2304
Ser Ala Gly Leu Val Trp Trp Arg Gly Arg     Ala Trp Arg     Thr
        745                 750                 755

cca cct gtt cct gaa tga gct gac caa ccg tgc gca tcc ggg tta tca      2352
Pro Pro Val Pro Glu     Ala Asp Gln Pro Cys Ala Ser Gly Leu Ser
            760                 765                 770

act gga agt ttt cga tcg cga ctc ggc gct gtt tgg tat tgc att tgt      2400
Thr Gly Ser Phe Arg Ser Arg Leu Gly Ala Val Trp Tyr Cys Ile Cys
            775                 780                 785

gac cga tca ggg cga agg tgg cgc tct gga cag ccc gca cta cga aca      2448
Asp Arg Ser Gly Arg Arg Trp Arg Ser Gly Gln Pro Ala Leu Arg Thr
        790                 795                 800

tag cca ttt tca acg tct gcg tga aat gag cgc gcg tat cat ggc tca      2496
    Pro Phe Ser Thr Ser Ala     Asn Glu Arg Ala Tyr His Gly Ser
        805                 810                 815

aag cgc acc gtt cga acc ggc gct gcc ggc gtt gcg taa tcc ggt tct      2544
Lys Arg Thr Val Arg Thr Gly Ala Ala Gly Val Ala     Ser Gly Ser
        820                 825                 830

gga tga gag ccc ggg ttg cca acg tgt cgc aga cgg tcg tgc gcg tgc      2592
Gly     Glu Pro Gly Leu Pro Thr Cys Arg Arg Arg Ser Cys Ala Cys
            835                 840                 845

gct gat ggc att gta cca agg cgt tta tga gct gat gtt tgc gat gat      2640
Ala Asp Gly Ile Val Pro Arg Arg Leu     Ala Asp Val Cys Asp Asp
        850                 855                 860

ggc gca gca ctt cgc cgt gaa acc gct ggg tag ctt gcg tcg cag ccg      2688
Gly Ala Ala Leu Arg Arg Glu Thr Ala Gly     Leu Ala Ser Gln Pro
        865                 870                 875

cct gat gaa cgc agc aat cga tct gat gac cgg tct gtt gcg tcc gct      2736
Pro Asp Glu Arg Ser Asn Arg Ser Asp Asp Arg Ser Val Ala Ser Ala
        880                 885                 890

gag ctg cgc gct gat gaa cct gcc aag cgg cat cgc cgg tcg cac ggc      2784
Glu Leu Arg Ala Asp Glu Pro Ala Lys Arg His Arg Arg Ser His Gly
    895                 900                 905

cgg tcc gcc gct gcc ggg tcc ggt tga cac ccg tag cta tga cga cta      2832
Arg Ser Ala Ala Ala Gly Ser Gly     His Pro     Leu     Arg Leu
```

910 915 920

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

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

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

<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

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

<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

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
```

```
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
```

```
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
```

```
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 Arg Leu Asn Pro Gln Phe Leu Glu Asp Phe Lys Leu Val His His
65                  70                  75                  80

aat gag ccg tcc ttg atg tcc acg ggc gtt ttg ttg tgc ggc gtg gag      288
Asn Glu Pro Ser Leu Met Ser Thr Gly Val Leu Leu Cys Gly Val Glu
                85                  90                  95

cgt cgc ggt ctg gtt cac gcg ctg cgc gat aag tgc cgc agc caa ggc      336
Arg Arg Gly Leu Val His Ala Leu Arg Asp Lys Cys Arg Ser Gln Gly
            100                 105                 110

att gct att cgt ttc gaa agc ccg ttg ctg gaa cac ggt gag ctg ccg      384
Ile Ala Ile Arg Phe Glu Ser Pro Leu Leu Glu His Gly Glu Leu Pro
        115                 120                 125

ctg gcg gac tat gat ctg gtg gtc ctg gct aat ggt gtt aat cac aaa      432
Leu Ala Asp Tyr Asp Leu Val Val Leu Ala Asn Gly Val Asn His Lys
    130                 135                 140

acc gcg cat ttc acc gag gct ctg gtc ccg cag gtg gac tac ggc cgc      480
Thr Ala His Phe Thr Glu Ala Leu Val Pro Gln Val Asp Tyr Gly Arg
145                 150                 155                 160

aat aag tac att tgg tat ggc act agc cag ctg ttc gat cag atg aat      528
Asn Lys Tyr Ile Trp Tyr Gly Thr Ser Gln Leu Phe Asp Gln Met Asn
                165                 170                 175

ctg gtt ttt cgt acc cat ggt aaa gat atc ttt atc gcg cat gcc tat      576
Leu Val Phe Arg Thr His Gly Lys Asp Ile Phe Ile Ala His Ala Tyr
            180                 185                 190

aag tat agc gat acc atg agc acg ttc att gtc gaa tgt agc gaa gag      624
Lys Tyr Ser Asp Thr Met Ser Thr Phe Ile Val Glu Cys Ser Glu Glu
        195                 200                 205

act tac gca cgc gca cgc ctg ggc gaa atg tcc gaa gag gcg agc gca      672
Thr Tyr Ala Arg Ala Arg Leu Gly Glu Met Ser Glu Glu Ala Ser Ala
    210                 215                 220

gaa tac gtt gct aag gtg ttc cag gcc gag ctg ggt ggt cac ggc ctg      720
Glu Tyr Val Ala Lys Val Phe Gln Ala Glu Leu Gly Gly His Gly Leu
225                 230                 235                 240

gtg agc cag ccg ggt ctg ggt tgg cgt aac ttc atg acg ttg tct cat      768
Val Ser Gln Pro Gly Leu Gly Trp Arg Asn Phe Met Thr Leu Ser His
                245                 250                 255

gac cgt tgt cat gat ggt aag ttg gtt ctg ctg ggt gac gcg ctg caa      816
Asp Arg Cys His Asp Gly Lys Leu Val Leu Leu Gly Asp Ala Leu Gln
            260                 265                 270

agc ggt cac ttt agc atc ggc cac ggc acc acg atg gcc gtg gtg gtg      864
Ser Gly His Phe Ser Ile Gly His Gly Thr Thr Met Ala Val Val Val
        275                 280                 285

gcg cag ctg ctg gtt aaa gcg ctg tgt acc gaa gat ggt gtg cct gcc      912
Ala Gln Leu Leu Val Lys Ala Leu Cys Thr Glu Asp Gly Val Pro Ala
    290                 295                 300

gcg ctg aaa cgt ttc gaa gag cgt gcc ctg ccg ctg gtg cag ttg ttc      960
Ala Leu Lys Arg Phe Glu Glu Arg Ala Leu Pro Leu Val Gln Leu Phe
305                 310                 315                 320

cgt ggc cac gca gac aac agc cgc gtt tgg ttc gaa acc gtc gaa gag     1008
Arg Gly His Ala Asp Asn Ser Arg Val Trp Phe Glu Thr Val Glu Glu
                325                 330                 335

cgc atg cac ctg tcc tcg gcg gaa ttt gtg caa agc ttc gac gca cgc     1056
Arg Met His Leu Ser Ser Ala Glu Phe Val Gln Ser Phe Asp Ala Arg
```

340 345 350

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 Val 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 ggt 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

tcg gag gtt ggt 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 ggt cgt gag cgt acc gcc ctg ggt 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 ggt 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 ggt cgc gaa gcg gac ggt tgg cgt tac cag cgc cca ggt aaa ggt      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 tcg 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 Arg 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 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

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
145 150 155 160

atc aag gcc aac ttc aag atc cgc cac aac atc taa 516
Ile Lys Ala Asn Phe Lys Ile Arg His Asn Ile
165 170

<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

<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

```
                                                    40
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
```

<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

<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>

<223> Linkerpeptid

<220>
<221> CDS
<222> (1)..(30)
<223> Linkerpeptid

<400> 53
ggt ggc tct ggc ggt ggg agt ggg gga agc 30
Gly Gly Ser Gly Gly Gly Ser Gly Gly Ser
1 5 10

<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

gcggggtgtg gaaattgatg ctgcattgac cacccaagac 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

<213> künstliche Sequenz

<220>
<223> Programm Biosynthese (123456)

<400> 58
aattcgcggc cgctctagag ctgctgcggt gtttggatgg agcgtgggcg gggtgtggaa 60

attgatgctg cattgaccac ccaagacgaa 90

<210> 59
<211> 79
<212> DNA
<213> künstliche Sequenz

<220>
<223> Programm Biosynthese (12346)

<400> 59
aattcgcggc cgctctagag ctgctgcggt gtttggatgg agcgtgggcg gggtgtggaa 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

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

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

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

agt act aga atg aaa cat tct tcc gat atc tgc att gtt ggt gct ggt      384
Ser Thr Arg Met Lys His Ser Ser Asp Ile Cys Ile Val Gly Ala Gly
        115                 120                 125

att tct ggt ttg acg tgc gca agc cat ctg ctg gac agc ccg gca tgc      432
Ile Ser Gly Leu Thr Cys Ala Ser His Leu Leu Asp Ser Pro Ala Cys
    130                 135                 140

cgt ggt ctg agc ctg cgt atc ttt gac atg cag caa gaa gcc ggt ggc      480
Arg Gly Leu Ser Leu Arg Ile Phe Asp Met Gln Gln Glu Ala Gly Gly
145                 150                 155                 160

cgt atc cgc agc aaa atg ctg gat ggt aag gca agc att gaa ctg ggc      528
Arg Ile Arg Ser Lys Met Leu Asp Gly Lys Ala Ser Ile Glu Leu Gly
                165                 170                 175

gca ggt cgc tac tcc cct cag ttg cac ccg cat ttc caa agc gca atg      576
Ala Gly Arg Tyr Ser Pro Gln Leu His Pro His Phe Gln Ser Ala Met
            180                 185                 190

cag cac tat agc caa aag agc gaa gtc tat ccg ttc acc cag ttg aag      624
Gln His Tyr Ser Gln Lys Ser Glu Val Tyr Pro Phe Thr Gln Leu Lys
        195                 200                 205

ttc aaa tct cac gtg cag caa aag ctg aag cgc gcc atg aat gaa ctg      672
Phe Lys Ser His Val Gln Gln Lys Leu Lys Arg Ala Met Asn Glu Leu
    210                 215                 220

tcc ccg cgt ctg aaa gag cat ggt aaa gag agc ttt ttg cag ttt gtc      720
Ser Pro Arg Leu Lys Glu His Gly Lys Glu Ser Phe Leu Gln Phe Val
225                 230                 235                 240

agc cgt tat caa ggt cac gat agc gcg gtt ggt atg atc cgc tct atg      768
Ser Arg Tyr Gln Gly His Asp Ser Ala Val Gly Met Ile Arg Ser Met
                245                 250                 255

ggt tac gac gca ctg ttc ctg ccg gat atc agc gca gaa atg gcc tac      816
Gly Tyr Asp Ala Leu Phe Leu Pro Asp Ile Ser Ala Glu Met Ala Tyr
            260                 265                 270

gac att gtg ggt aag cac ccg gag atc cag agc gtg acg gac aac gac      864
Asp Ile Val Gly Lys His Pro Glu Ile Gln Ser Val Thr Asp Asn Asp
        275                 280                 285

gcg aac caa tgg ttt gca gcg gaa acg ggc ttt gct ggt ctg att cag      912
Ala Asn Gln Trp Phe Ala Ala Glu Thr Gly Phe Ala Gly Leu Ile Gln
    290                 295                 300

ggc atc aag gct aag gtt aag gcg gca ggt gcg cgt ttt agc ctg ggt      960
Gly Ile Lys Ala Lys Val Lys Ala Ala Gly Ala Arg Phe Ser Leu Gly
305                 310                 315                 320

tat cgt ctg ctg agc gtc cgt acc gac ggt gac ggc tac ctg ctg caa     1008
Tyr Arg Leu Leu Ser Val Arg Thr Asp Gly Asp Gly Tyr Leu Leu Gln
                325                 330                 335
```

```
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                 350

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                 365

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                 380

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                 395                 400

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                 415

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                 430

tgg 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                 445

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                 460

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                 475                 480

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                 495

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                 510

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                 525

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
```

```
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
```

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

<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

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 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

ggt 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 ggc 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
```

```
tac ttt cgc ggt gaa gcg cgt gcg cgt ccg gat atc ccg cat cgt atc      1632
Tyr Phe Arg Gly Glu Ala Arg Ala Arg Pro Asp Ile Pro His Arg Ile
    530                 535                 540

gag ggc atg ggc ctg gtc ggc gtc gaa tct cgt cag gat ggc gac gct      1680
Glu Gly Met Gly Leu Val Gly Val Glu Ser Arg Gln Asp Gly Asp Ala
545                 550                 555                 560

gcg gaa tgg cgt ctg acg ggt ctg cgt ccg ggt ccg gca cgc att gtt      1728
Ala Glu Trp Arg Leu Thr Gly Leu Arg Pro Gly Pro Ala Arg Ile Val
                565                 570                 575

ctg gac gat ggt gcc gag gcg atc cct ctg cgt gtt ctg cct gac gat      1776
Leu Asp Asp Gly Ala Glu Ala Ile Pro Leu Arg Val Leu Pro Asp Asp
            580                 585                 590

tgg gcg ctg gat gac gcg acc gtc gaa gaa gtg gat tac gcc ttt ttg      1824
Trp Ala Leu Asp Asp Ala Thr Val Glu Glu Val Asp Tyr Ala Phe Leu
        595                 600                 605

tac cgc cac gtt atg gcg tat tac gag ctg gtg tat cca ttc atg agc      1872
Tyr Arg His Val Met Ala Tyr Tyr Glu Leu Val Tyr Pro Phe Met Ser
    610                 615                 620

gac aag gtg ttt tcc ctg gct gat cgt tgc aaa tgt gaa acg tac gca      1920
Asp Lys Val Phe Ser Leu Ala Asp Arg Cys Lys Cys Glu Thr Tyr Ala
625                 630                 635                 640

cgt ctg atg tgg cag atg tgt gat ccg cag aac cgc aac aag tcc tat      1968
Arg Leu Met Trp Gln Met Cys Asp Pro Gln Asn Arg Asn Lys Ser Tyr
                645                 650                 655

tac atg ccg agc acc cgc gaa ctg tcg gca ccg aaa gct cgt ttg ttc      2016
Tyr Met Pro Ser Thr Arg Glu Leu Ser Ala Pro Lys Ala Arg Leu Phe
            660                 665                 670

ttg aag tat ctg gcc cac gtg gaa ggc cag gca cgc ctg caa gca cct      2064
Leu Lys Tyr Leu Ala His Val Glu Gly Gln Ala Arg Leu Gln Ala Pro
        675                 680                 685

ccg cca gcg ggt ccg gca cgc att gaa tct aaa gcc cag ttg gcg gca      2112
Pro Pro Ala Gly Pro Ala Arg Ile Glu Ser Lys Ala Gln Leu Ala Ala
    690                 695                 700

gag ctg cgt aaa gcc gtc gac ctg gag ctg tct gtg atg ctg caa tac      2160
Glu Leu Arg Lys Ala Val Asp Leu Glu Leu Ser Val Met Leu Gln Tyr
705                 710                 715                 720

ctg tac gcg gcg tat agc att ccg aac tat gca cag ggc caa caa cgt      2208
Leu Tyr Ala Ala Tyr Ser Ile Pro Asn Tyr Ala Gln Gly Gln Gln Arg
                725                 730                 735

gtt cgt gac ggt gcg tgg acc gcc gag cag ctg caa ctg gcg tgc ggt      2256
Val Arg Asp Gly Ala Trp Thr Ala Glu Gln Leu Gln Leu Ala Cys Gly
            740                 745                 750

agc ggt gac cgt cgc cgt gat ggc ggt att cgt gca gca ctg ctg gaa      2304
Ser Gly Asp Arg Arg Arg Asp Gly Gly Ile Arg Ala Ala Leu Leu Glu
        755                 760                 765

att gct cat gaa gaa atg att cat tac ctg gtc gtt aac aac ctg ctg      2352
Ile Ala His Glu Glu Met Ile His Tyr Leu Val Val Asn Asn Leu Leu
    770                 775                 780

atg gcc ctg ggc gag ccg ttc tac gcg ggt gtc ccg ctg atg ggc gaa      2400
Met Ala Leu Gly Glu Pro Phe Tyr Ala Gly Val Pro Leu Met Gly Glu
785                 790                 795                 800
```

```
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 tcg 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 Glu 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
Ala

<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

```
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
```

740 745 750

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

```
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
```

```
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
```

```
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

<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 ggg      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
```

```
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 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

<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 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 Gln 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

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

act aga ggc ggg tca ggc ggg ggc tca ggc ggg tca act aga atg aaa       576
Thr Arg Gly Gly Ser Gly Gly Gly Ser Gly Gly Ser Thr Arg Met Lys
            180                 185                 190

cgt gcg att atc gtt ggt ggc ggc ctg gcg ggt ggc ctg acc gcg atc       624
Arg Ala Ile Ile Val Gly Gly Gly Leu Ala Gly Gly Leu Thr Ala Ile
        195                 200                 205

tac ctg gcg aag cgt ggc tac gaa gtg cac gtc gtg gag aag cgt ggt       672
Tyr Leu Ala Lys Arg Gly Tyr Glu Val His Val Val Glu Lys Arg Gly
    210                 215                 220

gat cct ctg cgc gat ctg agc tct tac gtg gac gtt gtt agc agc cgt       720
Asp Pro Leu Arg Asp Leu Ser Ser Tyr Val Asp Val Val Ser Ser Arg
225                 230                 235                 240

gcg atc ggc gtg agc atg acc gtt cgt ggt atc aag agc gtt ttg gct       768
Ala Ile Gly Val Ser Met Thr Val Arg Gly Ile Lys Ser Val Leu Ala
                245                 250                 255

gcg ggc att ccg cgt gca gag ctg gat gcg tgt ggc gaa ccg atc gtg       816
Ala Gly Ile Pro Arg Ala Glu Leu Asp Ala Cys Gly Glu Pro Ile Val
            260                 265                 270
```

```
gca atg gct ttc tcc gtg ggt ggt cag tat cgc atg cgc gaa ctg aag      864
Ala Met Ala Phe Ser Val Gly Gly Gln Tyr Arg Met Arg Glu Leu Lys
        275                 280                 285

ccg ttg gag gat ttc cgt ccg ctg agc ttg aac cgt gcg gcg ttt caa      912
Pro Leu Glu Asp Phe Arg Pro Leu Ser Leu Asn Arg Ala Ala Phe Gln
    290                 295                 300

aag ctg ctg aac aaa tac gcg aac ctg gca ggc gtt cgt tac tac ttt      960
Lys Leu Leu Asn Lys Tyr Ala Asn Leu Ala Gly Val Arg Tyr Tyr Phe
305                 310                 315                 320

gag cat aag tgc ctg gat gtt gac ctg gat ggt aag agc gtg ttg att     1008
Glu His Lys Cys Leu Asp Val Asp Leu Asp Gly Lys Ser Val Leu Ile
                325                 330                 335

cag ggc aaa gat ggt cag ccg cag cgt ctg caa ggt gac atg att atc     1056
Gln Gly Lys Asp Gly Gln Pro Gln Arg Leu Gln Gly Asp Met Ile Ile
            340                 345                 350

ggt gcg gat ggc gcc cac agc gcc gtc cgt cag gcg atg cag agc ggc     1104
Gly Ala Asp Gly Ala His Ser Ala Val Arg Gln Ala Met Gln Ser Gly
        355                 360                 365

ctg cgt cgt ttc gag ttc cag caa acg ttc ttc cgc cat ggc tac aaa     1152
Leu Arg Arg Phe Glu Phe Gln Gln Thr Phe Phe Arg His Gly Tyr Lys
    370                 375                 380

acc ctg gtt ttg ccg gac gcg caa gca ctg ggt tac cgt aaa gac acg     1200
Thr Leu Val Leu Pro Asp Ala Gln Ala Leu Gly Tyr Arg Lys Asp Thr
385                 390                 395                 400

ctg tac ttt ttc ggc atg gat tcc ggt ggc ctg ttc gcg ggt cgt gcg     1248
Leu Tyr Phe Phe Gly Met Asp Ser Gly Gly Leu Phe Ala Gly Arg Ala
                405                 410                 415

gct acg atc cca gat ggt agc gtc agc atc gcc gtt tgc ctg ccg tac     1296
Ala Thr Ile Pro Asp Gly Ser Val Ser Ile Ala Val Cys Leu Pro Tyr
            420                 425                 430

tcg ggt agc cct tcc ctg acg acc acc gac gaa ccg acg atg cgt gcg     1344
Ser Gly Ser Pro Ser Leu Thr Thr Thr Asp Glu Pro Thr Met Arg Ala
        435                 440                 445

ttc ttc gat cgt tac ttc ggt ggc ctg ccg cgt gac gcg cgt gac gaa     1392
Phe Phe Asp Arg Tyr Phe Gly Gly Leu Pro Arg Asp Ala Arg Asp Glu
    450                 455                 460

atg ctg cgt cag ttt ctg gcg aag ccg agc aac gac ctg att aac gtg     1440
Met Leu Arg Gln Phe Leu Ala Lys Pro Ser Asn Asp Leu Ile Asn Val
465                 470                 475                 480

cgc tct agc acc ttt cac tat aag ggt aat gtg ctg ttg ctg ggt gat     1488
Arg Ser Ser Thr Phe His Tyr Lys Gly Asn Val Leu Leu Leu Gly Asp
                485                 490                 495

gct gcg cat gcg act gcg ccg ttc ctg ggt cag ggt atg aac atg gcg     1536
Ala Ala His Ala Thr Ala Pro Phe Leu Gly Gln Gly Met Asn Met Ala
            500                 505                 510

ctg gag gac gcc cgc acg ttt gtc gag ctg ctg gac cgc cac cag ggc     1584
Leu Glu Asp Ala Arg Thr Phe Val Glu Leu Leu Asp Arg His Gln Gly
        515                 520                 525

gac caa gac aaa gcc ttt ccg gag ttc acg gag ctg cgc aaa gtc cag     1632
Asp Gln Asp Lys Ala Phe Pro Glu Phe Thr Glu Leu Arg Lys Val Gln
    530                 535                 540
```

```
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
545                 550                 555                 560

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
                565                 570                 575

cat tcc aag ttt ccg ggc 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
            580                 585                 590

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
        595                 600                 605

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
```

```
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 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 Glu
450 455 460

Met Leu Arg Gln Phe Leu Ala Lys Pro Ser Asn Asp Leu Ile Asn Val
465 470 475 480

Arg Ser Ser Thr Phe His Tyr Lys Gly Asn Val Leu Leu Leu Gly Asp
485 490 495

Ala Ala His Ala Thr Ala Pro Phe Leu Gly Gln Gly Met Asn Met Ala
500 505 510

Leu Glu Asp Ala Arg Thr Phe Val Glu Leu Leu Asp Arg His Gln Gly
515 520 525

Asp Gln Asp Lys Ala Phe Pro Glu Phe Thr Glu Leu Arg Lys Val Gln
530 535 540

Ala Asp Ala Met Gln Asp Met Ala Arg Ala Asn Tyr Asp Val Leu Ser
545 550 555 560

Cys Ser Asn Pro Ile Phe Phe Met Arg Ala Arg Tyr Thr Arg Tyr Met
565 570 575

His Ser Lys Phe Pro Gly Leu Tyr Pro Pro Asp Met Ala Glu Lys Leu
580 585 590

Tyr Phe Thr Ser Glu Pro Tyr Asp Arg Leu Gln Gln Ile Gln Arg Lys
595 600 605

Gln Asn Val Trp Tyr Lys Ile Gly Arg Val Asn
610 615

<210> 71
<211> 1113
<212> DNA
<213> künstliche Sequenz

<220>
<223> His_tag-Znf_Gli1-Linkerpeptid-nCFP

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

<400> 71
atg tat cat cac cat cac cat cac act aga aag cgt gag cct gaa tct 48
Met Tyr His His His His His His Thr Arg Lys Arg Glu Pro Glu Ser
1 5 10 15

gtg tat gaa act gac tgc cgt tgg gat ggc tgt agc cag gaa ttt gac 96

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

tcc caa gag cag ctg gtg cac cac atc aac agc gag cac atc cac ggg      144
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      192
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      240
Arg Pro Phe Lys Ala Gln 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      288
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      336
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      384
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      432
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      480
Lys Pro Tyr Val Cys Lys Leu Pro Gly Cys Thr Lys Arg Tyr Thr Asp
145                 150                 155                 160

cct agc tcg ctg cga aaa cat gtc aag aca gtg cat ggt cct gac gcc      528
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      576
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      624
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      672
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      720
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      768
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      816
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      864
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      912
Glu Gly Tyr Val Gln Glu Arg Thr Ile Phe Phe Lys Asp Asp Gly Asn
```

290 295 300

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

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

ggt 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

ggt 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
```

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

195 200 205

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_Linkerpeptid-nYFP

<220>
<221> CDS
<222> (1)..(840)
<223> His_tag-Znf_PBSII_Linkerpeptid-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
65 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

```
gac gta aac ggc cac aag ttc agc gtg tcc ggc gag ggc gag ggc gat      432
Asp Val Asn Gly His Lys Phe Ser Val Ser Gly Glu Gly Glu Gly Asp
    130                 135                 140

gcc acc tac ggc aag ctg acc ctg aag ttc atc tgc acc acc ggc aag      480
Ala Thr Tyr Gly Lys Leu Thr Leu Lys Phe Ile Cys Thr Thr Gly Lys
145                 150                 155                 160

ctg ccc gtg ccc tgg ccc acc ctc gtg acc acc ttc ggc tac ggc ctg      528
Leu Pro Val Pro Trp Pro Thr Leu Val Thr Thr Phe Gly Tyr Gly Leu
                165                 170                 175

atg tgc ttc gcc cgc tac ccc gac cac atg aag cag cac gac ttc ttc      576
Met Cys Phe Ala Arg Tyr Pro Asp His Met Lys Gln His Asp Phe Phe
            180                 185                 190

aag tcc gcc atg ccc gaa ggc tac gtc cag gag cgc acc atc ttc ttc      624
Lys Ser Ala Met Pro Glu Gly Tyr Val Gln Glu Arg Thr Ile Phe Phe
        195                 200                 205

aag gac gac ggc aac tac aag acc cgc gcc gag gtg aag ttc gag ggc      672
Lys Asp Asp Gly Asn Tyr Lys Thr Arg Ala Glu Val Lys Phe Glu Gly
    210                 215                 220

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

gac ggc aac atc ctg ggg cac aag ctg gag tac aac tac aac agc cac      768
Asp Gly Asn Ile Leu Gly His Lys Leu Glu Tyr Asn Tyr Asn Ser His
                245                 250                 255

aac gtc tat atc atg gcc gac aag cag aag aac ggc atc aag gtg aac      816
Asn Val Tyr Ile Met Ala Asp Lys Gln Lys Asn Gly Ile Lys Val Asn
            260                 265                 270

ttc aag atc cgc cac aac atc taa                                     840
Phe Lys Ile Arg His Asn Ile
        275
```

<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
```

```
                                                75
65                  70                  75                  80

Phe Ser Arg Ser Asp Val Leu Val Arg His Gln Arg Thr His Thr Gly
                85                  90                  95

Gly Gly Ser Gly Gly Gly Ser Gly Gly Ser Thr Arg Val Ser Lys Gly
            100                 105                 110

Glu Glu Leu Phe Thr Gly Val Val Pro Ile Leu Val Glu Leu Asp Gly
        115                 120                 125

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

Ala Thr Tyr Gly Lys Leu Thr Leu Lys Phe Ile Cys Thr Thr Gly Lys
145                 150                 155                 160

Leu Pro Val Pro Trp Pro Thr Leu Val Thr Thr Phe Gly Tyr Gly Leu
                165                 170                 175

Met Cys Phe Ala Arg Tyr Pro Asp His Met Lys Gln His Asp Phe Phe
            180                 185                 190

Lys Ser Ala Met Pro Glu Gly Tyr Val Gln Glu Arg Thr Ile Phe Phe
        195                 200                 205

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

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

Asp Gly Asn Ile Leu Gly His Lys Leu Glu Tyr Asn Tyr Asn Ser His
                245                 250                 255

Asn Val Tyr Ile Met Ala Asp Lys Gln Lys Asn Gly Ile Lys Val Asn
            260                 265                 270

Phe Lys Ile Arg His Asn Ile
        275
```

<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

<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 tcg 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 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 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 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

ggt 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
```

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

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

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

Ser Ser Arg Asp Val Leu Arg Arg His Asn Arg Thr His Thr Gly Glu
145 150 155 160

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

Val Leu Arg Arg His Asn Arg Ile His Leu Arg Gln Asn Asp Leu Glu
180 185 190

Gly Gly Ser Gly Gly Gly Ser Gly Gly Ser Thr Arg His His His His
195 200 205

His His
210

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

<220>
<223> cYFP-Linkerpeptid-Znf_Zif268-Linkerpeptid -His_ tag

<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

```
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 ccg ggt gaa aaa ccg tat gca tgt ccg gtt gaa agc tgt gat      336
Thr Arg Pro Gly Glu Lys Pro Tyr Ala Cys Pro Val Glu Ser Cys Asp
            100                 105                 110

cgt cgt ttt agc cgt agt gat gaa ctg acc cgt cat att cgt att cat      384
Arg Arg Phe Ser Arg Ser Asp Glu Leu Thr Arg His Ile Arg Ile His
        115                 120                 125

aca ggt cag aaa ccg ttt cag tgt cgt att tgc atg cgt aat ttt agc      432
Thr Gly Gln Lys Pro Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser
    130                 135                 140

cgt tca gat cat ctg acc acc cat att cgt acc cat act ggc gaa aaa      480
Arg Ser Asp His Leu Thr Thr His Ile Arg Thr His Thr Gly Glu Lys
145                 150                 155                 160

ccg ttt gcc tgt gat att tgt ggt cgt aaa ttt gca cgt tcc gat gaa      528
Pro Phe Ala Cys Asp Ile Cys Gly Arg Lys Phe Ala Arg Ser Asp Glu
                165                 170                 175

cgt aaa cgc cat acc aaa att cat acg ggt ggt ggt agc ggt ggt ggt      576
Arg Lys Arg His Thr Lys Ile His Thr Gly Gly Gly Ser Gly Gly Gly
            180                 185                 190

tca ggt ggt agt act aga cac cac cac cac cac cac tag                 615
Ser Gly Gly Ser Thr Arg His His His His His His
        195                 200
```

<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 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

Ansprüche

1. Methode zur Synthese einer Verbindung durch einen biosynthetischen Pfad durch Kultivierung einer genetisch modifizierten Wirtszelle, wobei die Zelle mit einer oder mehreren Nukleinsäuren, die folgendes umfassen, modifiziert wurde:
   a) Nukleotidsequenz, die mindestens drei bis etwa 100 chimäre Proteine kodiert, wobei jedes chimäre Protein aus mindestens einem Enzym des biosynthetischen Wegs oder funktionellen Polypeptids und mindestens einem Nukleinsäurebindendungsfaktor zusammengesetzt ist, wobei das genannte Enzym des biosynthetischen Wegs oder das funktionelle Polypeptid und der genannte Nukleinsäurebindungsfaktor, mit einem Linkerpeptid verbunden sind;

   b) Programm-Nukleotidsequenz, die mindestens drei bis etwa 100 Zielnukleinsäureelemente umfasst, von denen jedes durch einen Nukleinsäurebindungsfaktor vom chimären Polypeptid erkannt wird, und die genannte Programm-Nukleotidsequenz die Anordnung der Enzyme des biosynthetischen Wegs oder andere funktionelle Polypeptide entlang der Programm-Nukleotidsequenz liefert;

   c) Substrat für den biosynthetischen Pfad, das in der Zelle befindlich ist oder zur Wirtszelle extrazellulär zugefügt ist, und
   die genannte Kultivierung zur Synthese von den genannten chimären Proteinen in genetisch modifizierten Wirtszellen dient, was in der Produktion der Verbindung resultiert.

2. Methode zur Synthese eines biosynthetischen Produkts mit der Zugabe eines oder mehreren Bestandteilen umfassend:
   a) drei bis etwa 100 chimäre Proteine, wobei jedes chimäre Protein aus mindestens einem Enzym des biosynthetischen Wegs oder funktionellen Polypeptids und mindestens einem sich an das Nukleinsäuremotiv bindenden Faktor zusammengesetzt ist, wobei das genannte Enzym des biosynthetischen Wegs oder der funktionelle Polypeptid und der genannte Faktor, der sich an die Nukleinsäure bindet, mit einem Linkerpeptid verbunden sind;
   b) Programm-Nukleotidsequenz, die mindestens drei bis etwa 100 Zielelemente von Nukleinsäuren umfasst, von denen jedes durch einen gewissen Faktor des

chimären sich an das Nukleinsäuremotiv bindenden Polypeptids erkannt wird, und die genannte Programm-Nukleotidsequenz die Enzyme des biosynthetischen Wegs oder andere funktionelle Polypeptide entlang der Program-Sequenz anordnet;

c) Substrat für den biosynthetischen Weg und Kofaktoren für Enzyme werden dem Gemisch zugefügt oder entstehen im Gemisch, wobei sich die genannte Nukleinsäuresequenz im Gemisch befindet oder wird auf eine andere Verbindung oder Festphase immobilisiert.

3. Methode gemäß irgendwelchem der Ansprüche von 1 bis 2, dadurch gekennzeichnet, dass die Programm-Nukleinsäuresequenz aus Zielelementen der Nukleinsäure und aus einem Spacersequenz, die die Nukleinsäureelemente trennt, zusammengesetzt ist, und die genannten Nukleinsäureelemente von definierten Nukleotidsequenzen durch erkennbare Motive der die Nukleinsäure bindenden Faktoren bestimmt sind; und die Spacersequenz eine beliebige Länge, vorzugsweise von 1 bis 50 Nukleotide, aufweist.

4. Methode gemäß irgendwelchem der Ansprüche von 1 bis 3, dadurch gekennzeichnet, dass die Nukleotidsequenz jedes Zielelements der Nukleinsäure ausgewählt ist je nach einem gewissen Motiv, das durch den sich an das Nukleinsäuremotiv bindenden Faktor erkannt wird und das in der Methode als der Träger des Enzyms des biosynthetischen Wegs oder anderen funktionellen Peptids angewendet ist; und
jedes Zielelement der Nukleinsäure des Nukleotidmotivs einmal oder mehrmals angewendet ist, normalerweise von 1- bis 16-mal; und
drei oder mehrere verschiedene Elemente der Nukleinsäure des Nukleotidmotivs sind auf der Programm-Sequenz der Nukleinsäure angewendet, typisch 3 bis 16, vorzugsweise 3 bis 7, und dies hängt von der Zustand der Oligomerisation und der Anzahl von katalytischen Schritten im biosynthetischem Weg ab.

5. Methode gemäß irgendwelchem der Ansprüche von 1 bis 4, dadurch gekennzeichnet, dass das Sequenzmotiv des Zielelements der Nukleinsäure in irgendwelcher Größe, die 4 Nukleotide überschreitet, vorkommt, gegebenenfalls aus 4 bis etwa 30 Nukleotiden zusammengesetzt ist, und der die Nukleinsäure bindender Faktor erkennt das Zielelement der Nukleinsäure an jeder Stelle des Zielelements ein Nukleotid oder jede Kombination von zwei, drei oder vier Nukleotiden und jeder Zielelement der

Nukleinsäure ist auf der Programm-Sequenz der Nukleinsäure in einer oder mehr Kopien anwesend und in der Sequenz, die die örtliche Verteilung der chimären Proteine unter denen bestimmt.

6. Methode nach irgendwelchem der Ansprüche von 1 bis 5, dadurch gekennzeichnet, dass das Enzym des biosynthetischen Wegs oder irgendwelcher funktioneller Polypeptid kovalent über einen Linkerpeptid einer Länge von 1 bis etwa 100 Aminosäuren pro die Nukleinsäure bindenden Faktor gebunden ist, und dass der Faktor in Natur vorhanden oder künstlich gestaltet ist und die Eigenschaft für eine spezifische Bindung an Nukleinsäureelemente aufweist; und
der Faktor, der sich an das Nukleinsäuremotiv bindet, ist aus der folgenden Gruppe ausgewählt, doch nicht darauf beschränkt: Nukleinbindungsdomäne Helix-Loop-Helix, Nukleinbindungsdomäne Zinkfinger, Leucin-Reißverschluss als Bindungsdomäne, Nukleinbindungsdomäne Winged Helix, Winged Helix-Turn-Helix, Bindungsdomäne Helix-Loop-Helix, Bindungsdomäne HMG (HMG-Box), inaktive Restriktions-Endonuklease und Transkriptionsfaktoren.

7. Methode gemäß irgendwelchem der Ansprüche von 1 bis 6, dadurch gekennzeichnet, dass die Programmsequenz der Nukleinsäure als solche verwendet wird oder kloniert in einem Vektor, der für die Wirtszelle geeignet ist, und der Vektor repliziert sich außerhalb der Kromosome.

8. Nukleinsäure, die das Polypeptid kodiert, der aus dem Enzym des biosynthetischen Wegs oder funktionellen Polypeptids und einem Bindungsfaktor zusammengesetzt ist und in der Methode gemäß irgendwelchen der Ansprüche von 1 bis 6 angewendet ist, und die genannte Nukleinsäure in einen geeigneten Vektor unter der Kontrolle der regulatorischen Sequenz eingeführt ist, wobei die regulatorische Sequenz die Expression der Polypeptide moduliert und der Vektor für die Expression in der Wirtszelle geeignet ist.

9. Genetisch modifizierte Wirtszellen, die eine oder mehrere Programm-Nukleinsäuresequenzen gemäß irgendwelchem der Ansprüche von 3 bis 5 beinhalten und die Wirtszelle unter eukaryotischen oder prokaryotischen Zellen ausgewählt ist.

10. Genetisch modifizierte Wirtszellen, die ein oder mehrere Enzyme des biosynthetischen Wegs oder anderen funktionellen Peptide gemäß Anspruch 6 beinhalten und die Wirtszelle unter eukaryotischen oder prokaryotischen Zellen ausgewählt ist.

11. Genetisch modifizierte Wirtszellen, die eine oder mehrere Programm-Nukleinsäuresequenzen gemäß irgendwelchem der Ansprüche 3 bis 5 und ein oder mehr funktionellen Polypeptide gemäß Anspruch 6 beinhalten und die Wirtszelle unter eukaryotischen oder prokaryotischen Zellen ausgewählt ist.

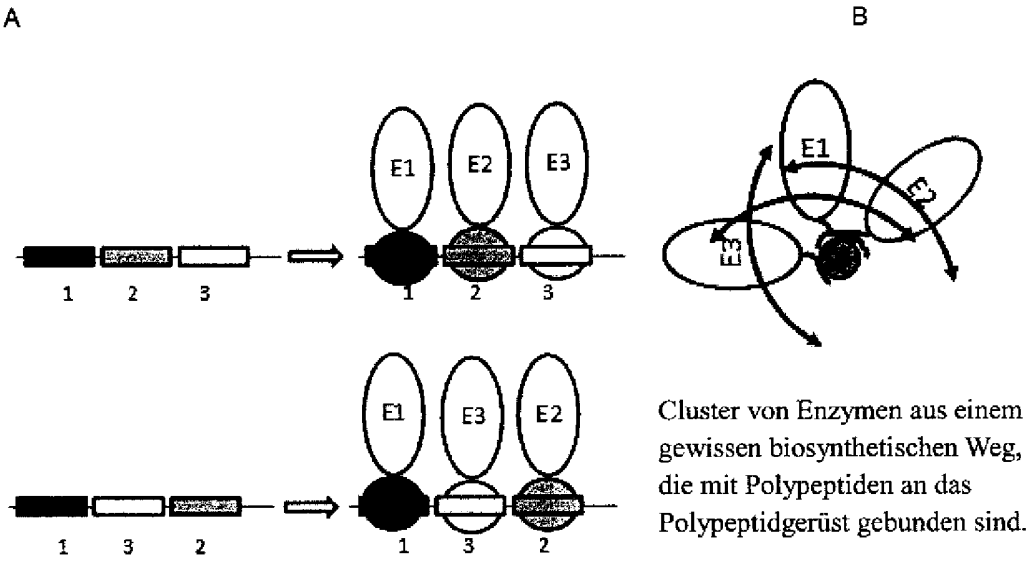
12. Metode gemäß irdendwelchem der Ansprüche 1 bis 2, dadurch gekennzeichnet, dass der biosynthetische Weg entweder katabolisch oder anabolisch ist und die Synthese von primären Metaboliten, wie z. B., doch nicht darauf beschränkt, mit: Aminosäuren, Fettsäuren, Kohlenhydraten, Pyrimidinen, Purinen, Zitronsäure, Itakonsäure, Ethanol, Glycerol, Methanol, Butanol, Propanol, Isoprene oder höheren Alkoholen.

13. Methode gemäß irgendwelchem der Ansprüche 1 bis 2, dadurch gekennzeichnet, dass der biosynthetische Weg eine Synthese von Sekundärmetaboliten ist, z. B., doch nicht darauf beschränkt: Polyketide, nichtribosomale Peptide, Hormone, Terpene, Antioxidanten, Pigmente.

14. Methode gemäß irgendwelchem der Ansprüche 1 bis 2, dadurch gekennzeichnet, dass sich bei dem biosynthetischen Weg um eine Synthese von Karotenoiden handelt.

15. Methode gemäß irgendwelchem der Ansprüche 1 bis 2, dadurch gekennzeichnet, dass sich bei dem biosynthetischen Weg um eine Synthese von Violacein handelt.

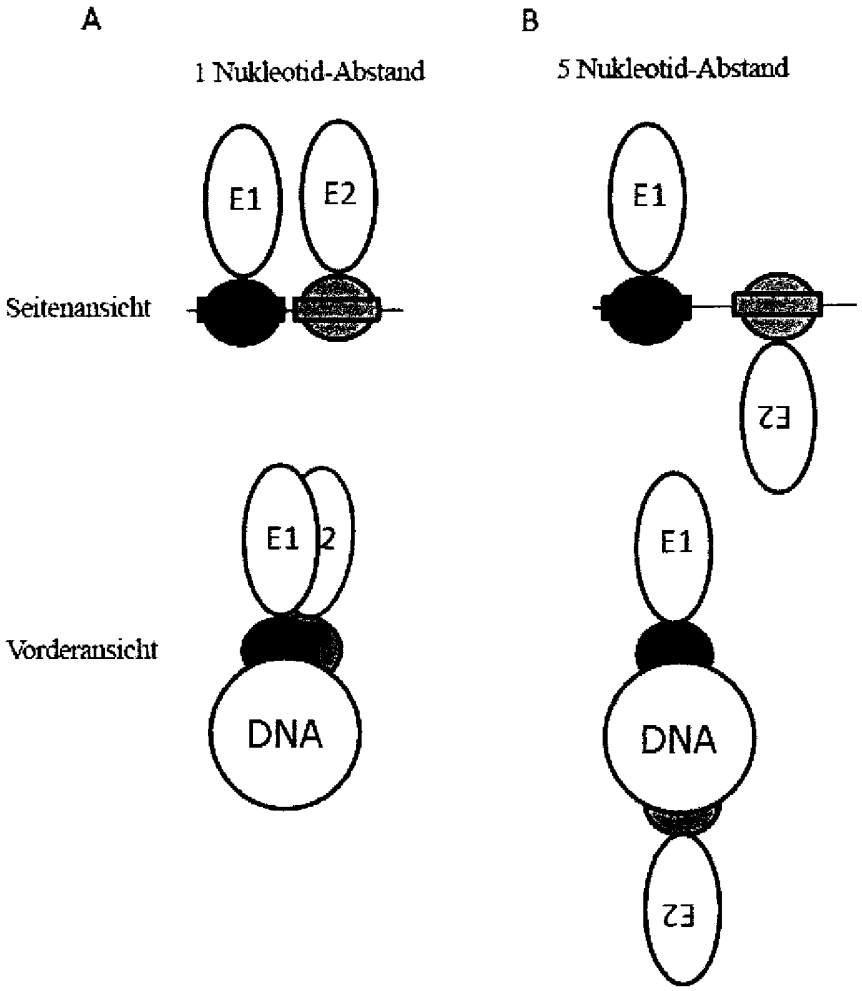
16. Methode gemäß irgendwelchem der Ansprüche 1 bis 2, dadurch gekennzeichnet, dass sich bei dem biosynthetischen Weg um eine Synthese von Resveratrol oder methylirtem Resveratrol handelt.



Nukleotidsequenz diktiert die Anordnung von Enzymen in einem gewissen biosynthetischen Weg mit Hilfe von Interaktionen zwischen der Nukleotidsequenz und den Nukleotidbindungsdomänen.

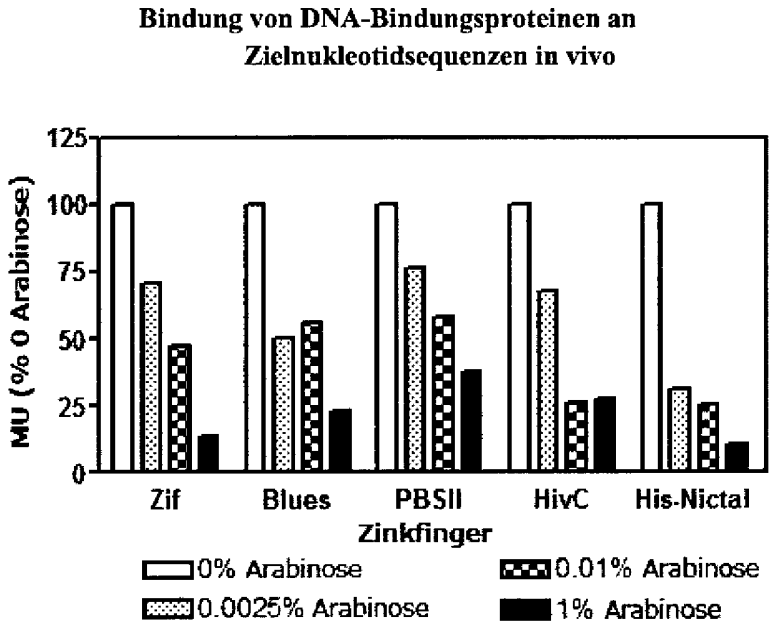
Cluster von Enzymen aus einem gewissen biosynthetischen Weg, die mit Polypeptiden an das Polypeptidgerüst gebunden sind.

Figur 1.

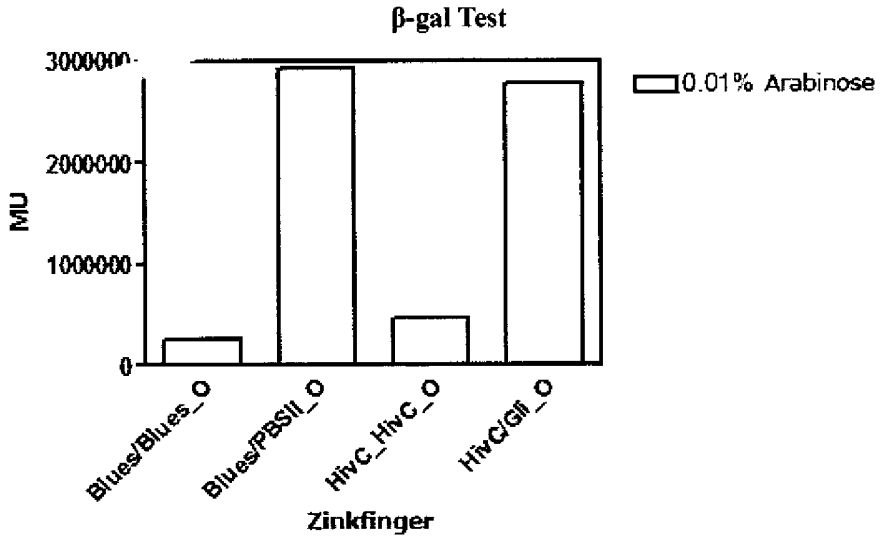

A
B
1 Nukleotid-Abstand
5 Nukleotid-Abstand
E1
E2
Seitenansicht
E1
E2
E1
2
Vorderansicht
DNA
E1
DNA
E2

Figur 2.

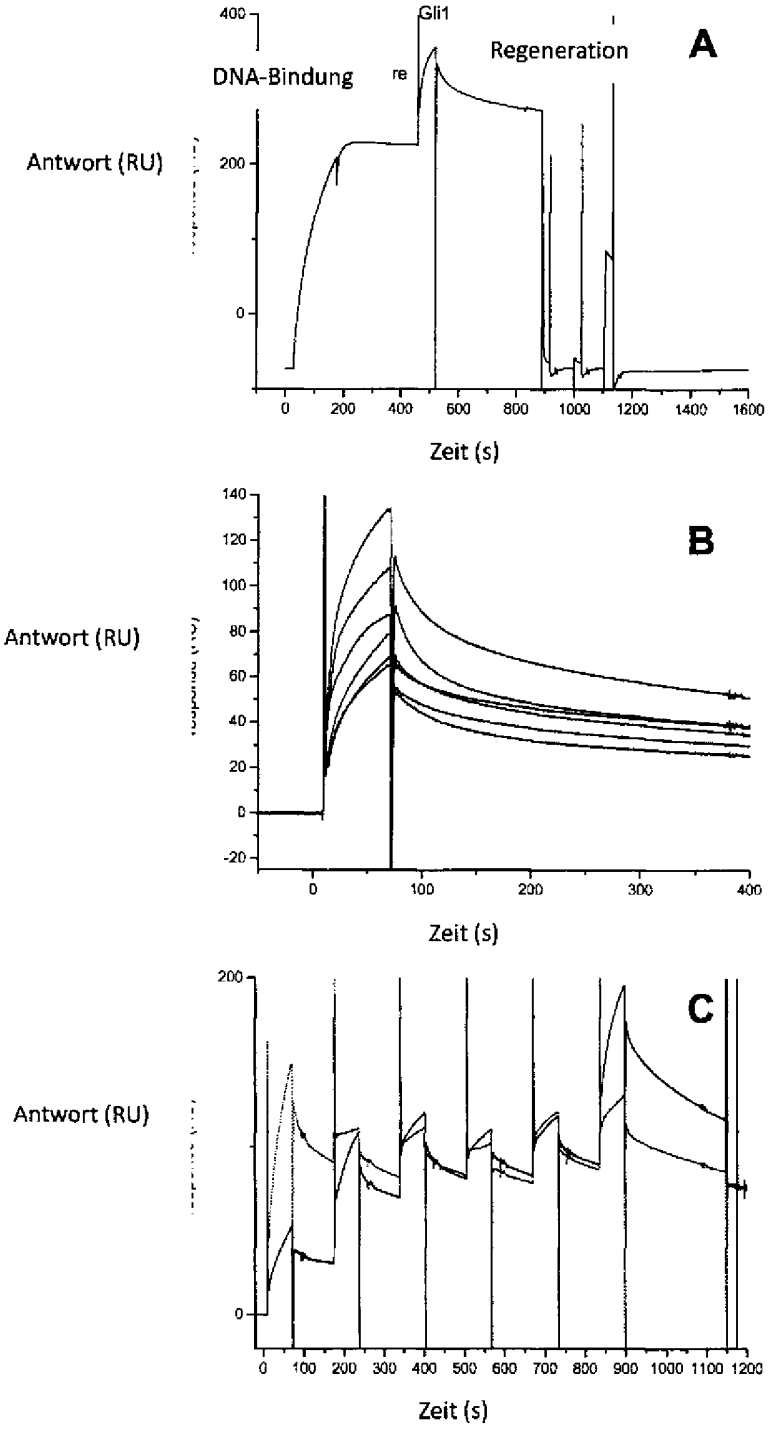
A


Bindung von DNA-Bindungsproteinen an
Zielnukleotidsequenzen in vivo
MU (% 0 Arabinose)
125
100
75
50
25
0
Zif
Blues
PBSII
HivC
His-Nictal
Zinkfinger
0% Arabinose
0.01% Arabinose
0.0025% Arabinose
1% Arabinose

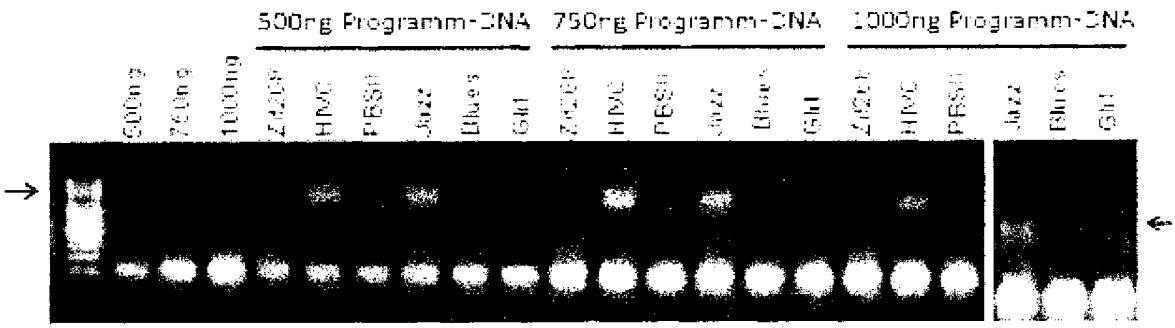
B


β-gal Test
MU
3000000
2000000
1000000
0
0.01% Arabinose
Blues/Blues_O
Blues/PBSII_O
HivC_HivC_O
HivC/Gli_O
Zinkfinger

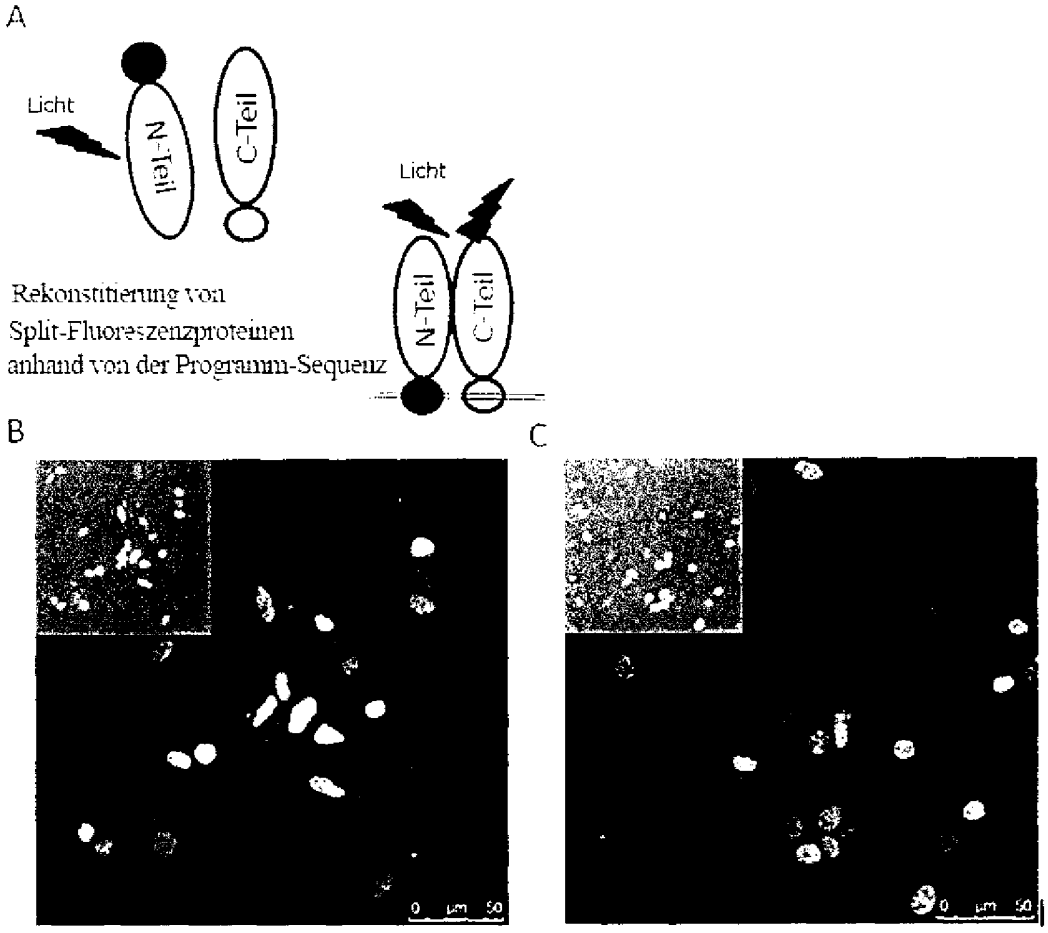
Figur 3.


A
Gli1
DNA-Bindung
Regeneration
Antwort (RU)
Zeit (s)
B
Antwort (RU)
Zeit (s)
C
Antwort (RU)
Zeit (s)

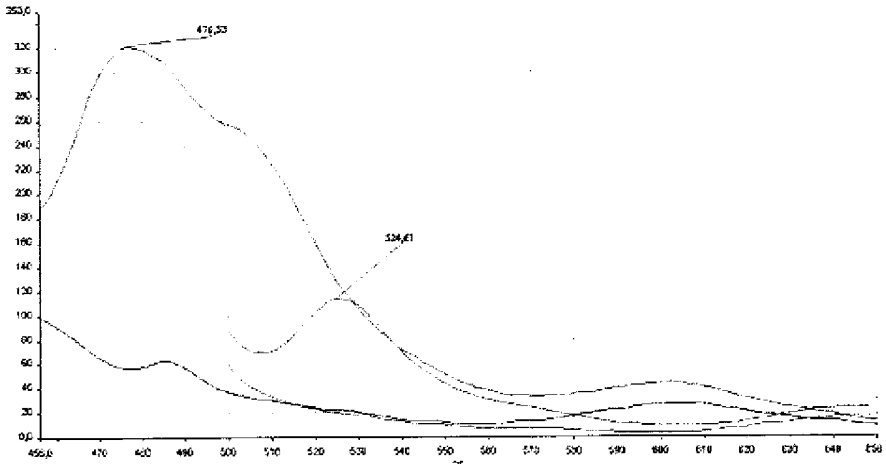
Figur 4.


500ng Programm-DNA
750ng Programm-DNA
1000ng Programm-DNA
500ng
750ng
1000ng
Z420P
HMC
PBSH
Jazz
Blues
Gld
Z420P
HMC
PBSH
Jazz
Blues
Gld
Z420P
HMC
PBSH
Jazz
Blues
Gld

Figur 5.

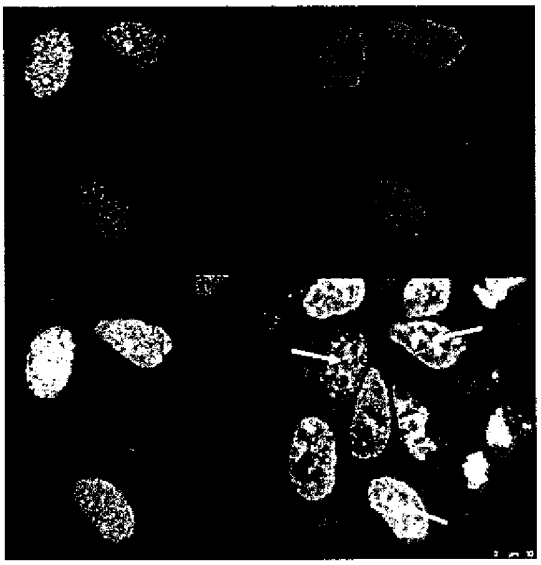

A
Licht
N-Teil
C-Teil
Licht
N-Teil
C-Teil
Rekonstitierung von
Split-Fluoreszenzproteinen
anhand von der Programm-Sequenz
B
C
0 µm 50
0 µm 50

Figur 6.

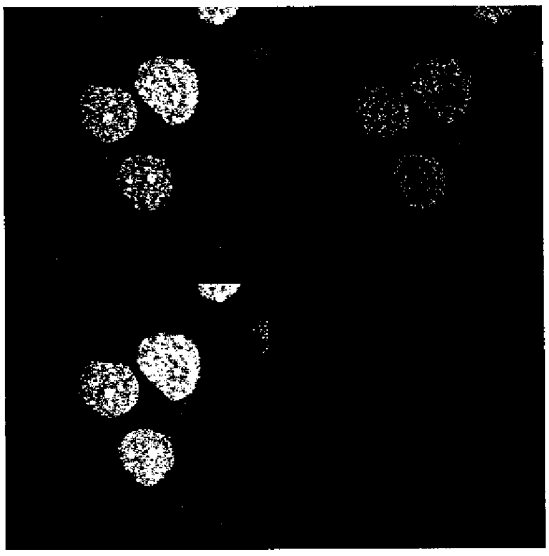


Figur 7.

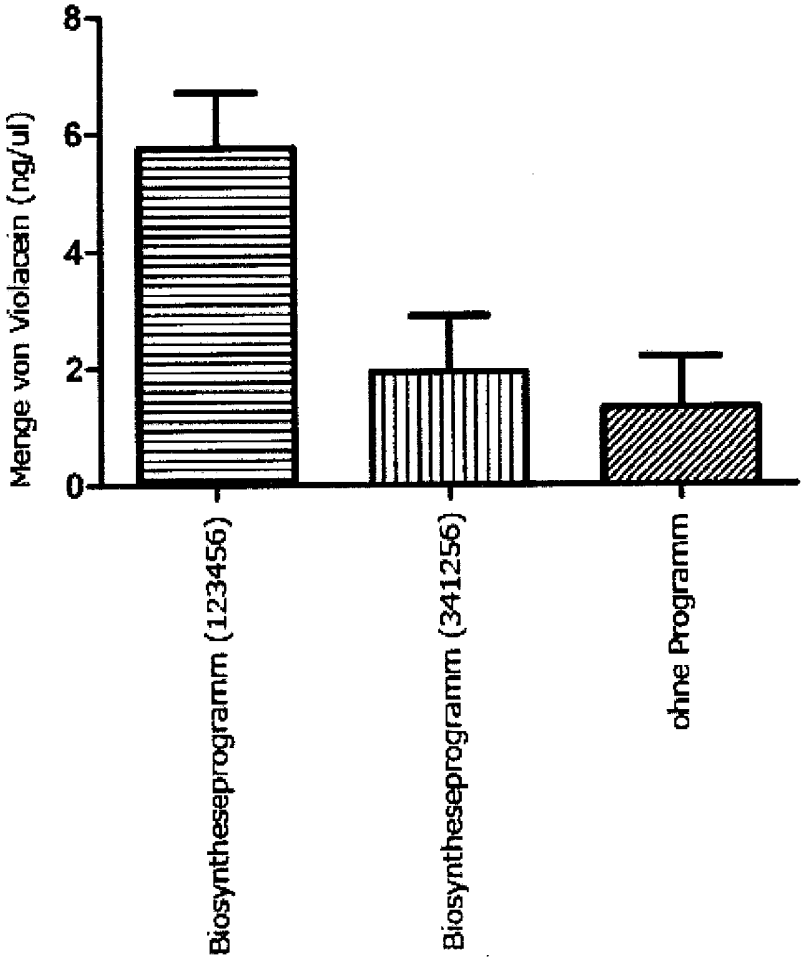
A)



B)



Figur 8.


Menge von Violacein (ng/ul)
8
6
4
2
0
Biosynthesepogramm (123456)
Biosynthesepogramm (341256)
ohne Programm

Figur 9.