COMPOSITIONS AND METHODS FOR IMPROVING PRODUCTION OF RECOMBINANT POLYPEPTIDES

Inventors: Oren Bogin, Sunnyvale, CA (US); Willem P. Stemmer, Los Gatos, CA (US); Volker Schellenberger, Palo Alto, CA (US); Yong Yin, Sunnyvale, CA (US); Chia-wei Wang, Santa Clara, CA (US); Nathan C. Geethin, Santa Clara, CA (US)

Correspondence Address: WILSON, SONSINI, GOODRICH & ROSATI 650 PAGE MILL ROAD PALO ALTO, CA 94304-1050 (US)

Appl. No.: 12/646,566 Filed: Dec. 23, 2009

Related U.S. Application Data

Continuation of application No. 12/228,859, filed on Aug. 15, 2008.

Provisional application No. 60/956,109, filed on Aug. 15, 2007, provisional application No. 60/981,073, filed on Oct. 18, 2007, provisional application No. 60/986,569, filed on Nov. 8, 2007.

Publication Classification

Int. Cl.
A61K 39/395 (2006.01)
C07K 14/00 (2006.01)
C07K 14/575 (2006.01)
C07K 16/00 (2006.01)
C12N 9/00 (2006.01)
C07K 14/54 (2006.01)
C07K 14/56 (2006.01)

(52) U.S. Cl. ..... 424/85.2; 530/350; 530/399; 530/387.1; 435/183; 530/387.3; 530/351; 530/380; 530/381; 514/9.7; 514/21.2; 514/19.3; 514/13.5; 514/2.3; 514/14.9; 514/16.4; 514/6.9; 514/15.1; 514/15.4; 424/130.1; 424/94.1; 424/135.1; 424/85.5; 424/85.6; 424/85.7; 514/7.7; 514/13.7; 536/23.5; 435/320.1; 435/325; 435/69.1

ABSTRACT

The present invention relates to biologically active polypeptides linked to one or more accessory polypeptides. The present invention also provides recombinant polypeptides including vectors encoding the subject proteinaceous entities, as well as host cells comprising the vectors. The subject compositions have a variety of utilities including a range of pharmaceutical applications.
Fig. 1: Modification of polypeptides

accessory polypeptide

Biologically active polypeptide
Fig. 2: Modules

Accessory polypeptide

Biologically active polypeptide

Depot module

Alginate
PLGA
PLGA-PEG

Matrix module

PLA
Fig. 3: Example modules for inclusion in modified polypeptides

- Biologically active polypeptide
- hGH, GLP-1, G-CSF, IFN.
- novel binding modules

rPEG accessory polypeptide

n = 36, 72, 144, 288 amino acids

Glycine-rich
Serine-rich
Glutamate residues for solubility

Depot Module

Lysine-rich
Arginine-rich
Poly-histidine
Protease sensitive
Fig. 4: Examples of modified polypeptides
Fig. 5: A modified polypeptide forming tetramers
Fig. 6: Incorporation of a modified polypeptide into alginate microspheres

Basic peptide

Alginate polymer
Fig. 7: Poly(His)-rPEG modified polypeptide in a chelating hydrogel.
Fig. 8: Protease sensitive multimeric modified polypeptides

+ Protease
Fig. 9: Schematic of plasmid pCW0150
Fig. 10: rPEG(L288)-GFP expression construct

Flag-rPEG_L288-H6-GFP (LCW0169.004)
Fig.11: DNA and amino acid sequence of rPEG(L288)
Fig. 12: hGH-rPEG and GLP-1-rPEG
Fig. 13: Examples of sequences of rPEG-modified polypeptides

- hGH
- hIFNα
- hGCSF
Fig. 14: Sequence Optimization Trends

- Highest solubility
- Non-specific binding to proteins & bone
- High charge can reduce halflife

- Moderate solubility
- Collagen-like aggregation
- Highest entropy (in theory)

- High solubility
- Highest expression
- No aggregation

- Low expression

E: Highest solubility
SE: Moderate solubility
SEG: High solubility
GS: Low expression
Fig.15: Sequence Optimization Trends

- **Gly-based**
  - Poly-G G/G/G/G/G/G/G/G/G/G/G 0% charged low solubility
  - rPEG H GGSGGSGGSGGE/GGSGGSGGSGGE 8% charged moderate solubility
  - rPEG J GGSGGE/GGSGGE/GGSGGE/GGSGGE 16% charged high solubility pH>7
  - rPEG K GEGGEGGEGGEGGEGGEGGGE 33% charged high solubility
  - rPEG M GE/GE/GE/GE/GE/GE/GE/GE 50% charged non-specific binding

- **Ser-based**
  - rPEG O SSSSSE/SSSSSE/SSSSSE/SSSS 16% charged soluble, hi-expressed
  - rPEG L SSSESSESSSSE/SSSESSESSSSE 25% charged soluble, hi-expressed
  - rPEG N SSSESSESSSSESSSSESSSE 33% charged soluble, hi-expressed
Fig. 16: Construction of rPEG-J288 construct
Fig. 17: Sequence of rPEG_J288 accessory polypeptide

GGTGGTTCTGGTGGTGAAGGTTGGTCTGGTGGTGAAGGTTGGTCTGGTGGTGAAGGTTGGTCTGGTGGTGA
GGTGGTTCTGGTGGTGAAGGTTGGTCTGGTGGTGAAGGTTGGTCTGGTGGTGAAGGTTGGTCTGGTGGTGA
GGTGGTTCTGGTGGTGAAGGTTGGTCTGGTGGTGAAGGTTGGTCTGGTGGTGAAGGTTGGTCTGGTGGTGA
GGTGGTTCTGGTGGTGAAGGTTGGTCTGGTGGTGAAGGTTGGTCTGGTGGTGAAGGTTGGTCTGGTGGTGA
GGTGGTTCTGGTGGTGAAGGTTGGTCTGGTGGTGAAGGTTGGTCTGGTGGTGAAGGTTGGTCTGGTGGTGA
Fig. 18: Design of the pCW0051 stuffer vector

**Flag**

| M | D | Y | K | D | D | D | D | G | S | P | G | * | * | P | R | * | * | G | G | S | S | S | S | L | E |

**Bsal**

ATGGATTATAAAACGATCAGTTAAAAAGGGTCTCCAGTTAGGATGGTGCAGTGGCCTTGTCGAG

**KpnI**

GGTACCCATCATCATCACGCTCTGGTACGAAATG

**6x His-tag**

GTHHHEHHHHHHHEHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
Fig. 19: Purification of Flag-rPEG-J288-H6-GFP

Ladder

Flag

IMAC

Lysate

Ladder
Fig. 20: Serum stability of rPEG-J288-modified GFP.

3 Days, 37°C, 50% human serum
No degradation detected
Fig. 21: Interaction of an accessory-modified polypeptide with a cellular target

Accessory polypeptide

Cell surface

Biologically active polypeptide

Target antigen
Fig. 22: Crosslinked accessory polypeptides
Fig. 23: Examples of cross-linking components

BSPEG\textsuperscript{2}
M.W. 332.5
Spacer Arm 21.7 Å

SMPEG\textsuperscript{11}
M.W. 688.71
Spacer Arm 39.2 Å

4-arm PEG succinimidygluturate

Sulfo-TSA\textsuperscript{T}
M.W. 788.49
Fig. 24: Examples of crosslinked accessory polyketides
Fig. 25: Example: Using streptavidin as a multivalent cross-linker
Fig. 26: Examples of combinations of 1-4 Product Modules

A wide variety of product formats with different stoichiometries can be made from a set of 4 modules.
Fig. 27: Product Formats using conjugated accessory polypeptides

Bidirectional linear polymer

Unidirectional linear polymer (one end blocked)

Blocked

Branched multivalent polymers made from two different recombinant proteins

Blocked
Fig. 28: Modified accessory polypeptides
Fig. 29: Accessory Proteins with Protease-cleavable Sites
Fig. 30: Universal accessory polypeptides

+ use Lys-Lys Linkage chemistry from Pierce

E. coli-made Synthetic
Fig. 33: Expression Levels and Solubility of rPEGs

GFP-fused, OD-corrected, Cyttoplasmic

Serine

Glycine

Lysate

Soluble

Relative Fluorescence (RFU/OD)

11x

14x
Fig. 34: Activity of accessory-modified hGH polypeptide
Fig. 35: rPEG-Facilitated Purification

Cell Lysis ➔ Anion Exchange ➔ Size Exclusion

capture, pH 5.6
pH 8.0

All rPEG fusions have same pI & Size; rPEG is a tag for one-step purification
Fig. 36: Single Band by SDS-PAGE

rPEG 288AA = 21-30kD

Expressed in E. coli cytoplasm
Fig. 37: Single Band by Analytical SEC

- rPEG-GLP1
- rPEG-K
- rPEG-J

Buffer

SEC shows >99% purity of rPEG_K and rPEG_J
Fig. 38: Single Band by Analytical Reverse Phase HPLC
Fig. 39: Single Band by Mass Spectrometry

rPEG_J288

rPEG_J299 is correct size by ion-spray Mass Spectrometry

43093.90
51345.71
10423.54

~21 KDa
Fig. 40: No Binding to Other Proteins

Binding of serum Antibodies to rPEG

Binding of non-Ab serum proteins to rPEG

No serum proteins bind to rPEG
Fig. 41: rPEG Increases Apparent MW

<table>
<thead>
<tr>
<th>Protein</th>
<th>Actual MW</th>
<th>Apparent MW</th>
</tr>
</thead>
<tbody>
<tr>
<td>rPEG288-GFP</td>
<td>59 kDa</td>
<td>433 kDa</td>
</tr>
<tr>
<td>rPEGJ288-GFP</td>
<td>52 kDa</td>
<td>303 kDa</td>
</tr>
<tr>
<td>rPEGK288-GFP</td>
<td>54 kDa</td>
<td>468 kDa</td>
</tr>
</tbody>
</table>

Note: 20kD rPEG behaves as if it is 340kD.
Fig. 42: rPEG is Stable in Rat and Human Serum

*Incubation at 37°C for 7 Days*

*Western Assay*

<table>
<thead>
<tr>
<th>Human</th>
<th>rPEG K</th>
<th>rPEG L</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>0</td>
<td>1d</td>
</tr>
<tr>
<td></td>
<td>3d</td>
<td>7d</td>
</tr>
<tr>
<td>Rat</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

rPEGs are stable for >7 days in human and rat serum at 37°C
Fig. 43: rPEG288 Has 10-20h PK in Rats

Half-life of rPEG288 in rats is about 10-20 hrs
Fig. 44: rPEGs are Non-Immunogenic

IgG 1:12,500

5 rabbits

IgG 1:5,000

3 rats/sample, 3 ID/IP injections, 6 weeks, GFP-fusions

IgM 1:5,000

No evidence that rPEGs J,K,L are immunogenic in rats

Approved 321AA
Bacterial Protein
Fig. 45: Manufacturing Advantage of rPEG

- Periplasmic: Low yield
- Cytoplasmic: Inclusion bodies, Refold

Examples: hGH, GCSF, IFNa

E. coli expressed

Human Protein

rPEG enables cytoplasmic expression of active protein, that otherwise would form inclusion bodies

E. coli expressed

Human Protein + rPEG
Fig. 46: Sustained Release of Modified Polypeptides

Microbead

Skin

Biodegradable Polymeric Gel
Drug + rPEG

Release from Depot
Kidney Clearance

Approved PLGA Depot Products:

- Lupron Peptide Monthly
- Zoladex Peptide Monthly
- Nutropin hGH Monthly

Frequency determined by dose/halflife

Slow Release + rPEG = Monthly Dosing + High Serum Concentration
Fig. 48: GLP1-rPEGK

9.42 ml

9.39 ml

9.28 ml
Fig. 49: SEC: Apparent MW of GLP1-rPEG fusions

J288 = 198kD
K288 = 358kD
L288 = 357kD

\[ y = 0.8437 \ln(x) + 20.153 \]
\[ R^2 = 0.9431 \]
Fig. 50: Purification of rPEG288-K-hGH Fusion Protein

- E. coli expression
- Affinity enrichment
- Protease digestion
- rPEG288-hGH obtained in >95% purity
- Final product is free of tags

Size exclusion
Ion exchange
Affinity depletion
Protease cleavage
Affinity capture
Cell lysate
Marker
Fig. 51: Single Band by RP-HPLC

Detector A Ch1: 220nm
Detector A Ch2: 280nm

BP Concentration (Method)

mAU

RP-HPLC shows >98% purity of rPEG_K
**Fig. 65**

**a.** scFv rPEG50 fusion

- **FLAG tag**
- **His6 tag**

**b.** Anti-Her2 diabody rPEG50 sequence with 3 amino acid linker

```
MEGDIHMEDIQTSEQSPSSLSASVGDRVITTCRASQDVNTAVAVYQQKPGKAKPKLLIYASASFLYSGVSRSRSRGTDPFLTISLQPEDFATYQQQHTTTPQTFQGTKVEIKSEGEEVQLVEISGCLVQPGSGLRLSCAASGFINKDTYIHWVRQAPDGKLEGWARYPTNGYTRYADSVKMGFRTISADTSKNTAYLQMSLRAEDTAVYYCISRSGSDGYPAMDYWQQTLVTVMGGS
```

**c.**
Fig. 68

1. Vaccinated donor
2. Antibody-secreting cells
3. Clone variable regions fused to rPEG
4. Phage panning/Screening
5. Large scale production
Fig. 72a

10 weekly SC injections of 50ug Ex4-PEG50 or ELSPAR in mice (N=10) → IgG → Tox analysis

Fig. 72b

1:500 dilution

Fig. 72c

1:12,500 dilution
Fig. 73

Apparent MW rPEG25 = 500 kDa
Apparent MW rPEG50 = 1,500 kDa

Absorbance (280nm)
<table>
<thead>
<tr>
<th>Input Sequence</th>
<th>Segment Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABCD</td>
<td>1</td>
</tr>
<tr>
<td>BCDA</td>
<td>2</td>
</tr>
<tr>
<td>CDAB</td>
<td>3</td>
</tr>
<tr>
<td>ABCD</td>
<td>4</td>
</tr>
<tr>
<td>DABC</td>
<td>5</td>
</tr>
<tr>
<td>EACD</td>
<td>6</td>
</tr>
<tr>
<td>BCDA</td>
<td>7</td>
</tr>
<tr>
<td>CDAB</td>
<td>8</td>
</tr>
<tr>
<td>DABE</td>
<td>9</td>
</tr>
<tr>
<td>DEEA</td>
<td>10</td>
</tr>
<tr>
<td>EACA</td>
<td>11</td>
</tr>
<tr>
<td>Sum</td>
<td>Max</td>
</tr>
</tbody>
</table>

Fig. 75
Fig. 77

Library of URP segments

Screen for expression
Screen for stability
Screen for length
Screen for amino acid composition
Screen for epitopes

Collection of URP segments

Multimerization/Dimerization
Screening

Collection of nrURP sequences

Screening

In vivo evaluation

Optimized nrURP sequences
Product is an Inactive Pro-Drug
Which is activated by serum protease

Activated Drug has long halflife

Fig. 96a  dAb, scFv  Ligand  receptor
rPEG50

Fig. 96b  Ligand  receptor
rPEG50

time
serum protease

Fig. 96c  dAb, scFv  Ligand  receptor
rPEG50

Fig. 96d  dAb, scFv
rPEG50

time
serum protease

Fig. 96e  receptor  Ligand
rPEG50

Fig. 96f  Ligand
rPEG50

time
serum protease

Fig. 96g  receptor  Ligand
rPEG50

Fig. 96h  receptor
rPEG50

time
serum protease

Options:
- No cleavage sites
- Protease sites (1-N) for in vivo cleavage in patient's blood
- Protease sites for in vitro cleavage (after purification, before injection)
- Plus or minus Association Sequence (ie peptide like SKVILF, RARADADA or Leu-zipper coiled coil, or domains.

Patent Application Publication
Oct. 14, 2010  Sheet 84 of 91
US 2010/0267076 A1
Added to recombinant product (receptor-rPEG) to prevent peak dose toxicity or receptor-mediated clearance.

Fig. 97

( Synthetic) peptide.

Fig. 98

rPEG product with multiple peptides with the same activity.

rPEG product with multiple peptides having different activities, with one or more copies of the peptide per activity.
Fig. 99

The Pro-Drug-PEG Format Yields a Constant Effective Concentration, Preventing Peak Dose Toxicity.
Fig. 100

Receptor-mediated clearance reduces half-life of a drug, but can be decreased with a drug-binding site that prevents receptor binding.
COMPOSITIONS AND METHODS FOR IMPROVING PRODUCTION OF RECOMBINANT POLYPEPTIDES

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the priority benefit of U.S. Provisional Application Ser. Nos. 60/956,109 filed on Aug. 15, 2007; 60/981,073, filed Oct. 18, 2007 and 60/986,569, filed Nov. 8, 2007, pending, which are hereby incorporated herein by reference in their entirety.

[0002] The instant application contains a Sequence Listing which has been submitted via EFS-Web and is hereby incorporated by reference in its entirety. Said ASCII copy, created on Apr. 21, 2010, is named 32808301.txt, and is 280,493 bytes in size.

BACKGROUND OF THE INVENTION

[0003] Recombinant proteins have become very attractive candidates for the development of novel therapeutics. However, production of protein pharmaceuticals requires significant optimization of processes to obtain sufficient yields of specific biologically active polypeptides. It is well established that the expression of recombinant proteins in the cytoplasm of Escherichia coli, in particular mammalian recombinant proteins, frequently results in the formation of insoluble aggregates known as inclusion bodies. High cell density fermentation and purification of the recombinant protein from inclusion bodies of E. coli are two major bottlenecks for the cost effective production of therapeutically useful proteins (Panda, A. K. 2003, Adv. Biochem. Eng. Biotechnol., 85, 43).

Similarly, for research purposes, where hundreds of proteins may need to be screened for various activities, the expression of soluble, active protein is desirable, thereby avoiding the step of first purifying inclusion bodies and then having to denature and refold protein each separately.


[0005] Human proteins typically fold using a hydrophobic core comprising a large number of hydrophobic amino acids. Research has shown that proteins can aggregate and form inclusion bodies, especially when genes from one organism are expressed in another expression host, such that the protein’s native binding partners are absent, so that folding help is unavailable and hydrophobic patches remain exposed. This is especially true when large evolutionary distances are crossed: a cDNA isolated from a eukaryote for example, when expressed as a recombinant gene in a prokaryote, has a high risk of aggregating and forming an inclusion body. While the cDNA may properly code for a translatable mRNA, the protein that results will emerge in a foreign microenvironment. This often results in misfolded, inactive protein that generally accumulates as aggregates if the concentration is high enough. Other effectors, such as the internal microenvironment of a prokaryotic cell (pH, osmolarity) may differ from that of the original source of the gene and affect protein folding. Mechanisms for folding a protein may also be host-dependent and thus be absent in a heterologous host, and hydrophobic residues that normally would remain buried as part of the hydrophobic core instead remain exposed and available for interaction with hydrophobic sites on other proteins.

Processing systems for the cleavage and removal of internal peptides of the expressed protein may also be absent in bacteria. In addition, the fine controls that may keep the concentration of a protein low will also be missing in a prokaryotic cell, and over-expression can result in filling a cell with protein that, even if it were properly folded, would precipitate by saturating its environment.

[0006] The recovery of biologically active products from the aggregated state found in inclusion bodies is typically accomplished by unfolding with chaotropic agents or acids, followed by dilution or dialysis into optimized refolding buffers. However, many polypeptides (especially structurally complex oligomeric proteins and those containing multiple disulfide bonds) do not easily adopt an active conformation following chemical denaturation.

[0007] Small changes in primary structure can affect solubility, presumably by altering folding pathways (Mitra, A. et al. 1989) Bio/Technology 7, 690; Banex, F. et. al. 2004 Nat Biotechnol., 22, 1399; Ventura, S. 2005 Microb Cell Fact, 4,11). In order to reduce the formation of insoluble aggregates during high-density fermentation, some groups have linked heterologous fusion proteins to the protein of interest. Examples of such fusion sequences are Glutathione-S-Transferase (GST), Protein Disulfide Isomerase (PDI), Thioredoxin (TRX), Maltose Binding Protein (MBP), His6 tag (SEQ ID NO:1), Clathrin Binding Domain (CBD) and Cellulose Binding Domain (CBD) (Sahadev, S. et al. 2007, Mol. Cell. Biochem.; Dysom, M. R. et al. 2004, BMC Biotechnol, 14, 52). In summary, these approaches were found to be protein-specific, as they do not work for all proteins.

[0008] While various fusion proteins have been designed to improve folding, chemical PEGaloylation of proteins has also been reported to enhance protein solubility, reduce aggregation, reduce immunogenicity, and reduce proteolysis. Nonetheless, the proper folding of overproduced polypeptides remains problematic within the highly concentrated and viscous environment of the cell cytoplasm, where aggregation occurs in a concentration-dependent manner. Another approach for the expression of mammalian proteins in bacterial hosts avoids leader peptides and expresses the active protein directly in the cytoplasm of the host. However, this process tends to result in aggregation and inclusion body formation.

[0009] One widely used approach for the expression of mammalian proteins in active form in bacteria is to direct the protein into the non-reducing environment of the periplasmic space of bacterial hosts such as E. coli, typically using signal or leader peptides to direct secretion. Secretion into the periplasm (and rarely into the media) appears to mimic the native
The present invention involves the steps of (a) providing a polynucleotide sequence coding for a modified polypeptide comprising the biologically active polypeptide linked with an accessory polypeptide such that expression of the modified polypeptide in a host cell yields a higher quantity of soluble form of biologically active polypeptide, as compared to expression of the biologically active polypeptide by itself (e.g., free from said accessory polypeptide); and (b) causing the modified polypeptide to be expressed in said host cell, thereby producing the biologically active polypeptide. In one embodiment, the expression of the soluble, active form of a biologically active polypeptide is about 1%, 5%, 25%, 50%, 75%, 95% or 99% of the total of that protein. In one embodiment, the expression of the modified polypeptide in a host cell yields at least about 2-fold more soluble form of biologically active polypeptide as compared to expression of the biologically active polypeptide by itself. In another embodiment, the biologically active polypeptide is linked to the accessory polypeptide via a proteinase cleavage site. Where desired, the cleavage site can be selected from the group consisting of TEV protease, enterokinase, Factor Xa, thrombin, PreScission™ protease, 3C protease, sortase A, and granzyme B. In some embodiments, the expression of the modified polypeptide in a host cell yields at least about 2-fold, 5-fold, 10-fold, 30-fold, or 100-fold, or more soluble form of biologically active polypeptide.

The present invention also provides a host cell for expressing the modified polynucleotide sequence. The host cell is typically prokaryotic including but not limited to E. Coli, and it may also be eukaryotic such as yeast cells and also mammalian cells (e.g. CHO cells).

The present invention also provides a genetic vehicle comprising the subject polynucleotide sequence that encodes a biologically active polypeptide linked with or without an accessory polypeptide.

Further provided by the present invention is a composition comprising soluble form of a biologically active polypeptide linked with an accessory polypeptide, wherein said accessory polypeptide when linked with the biologically active polypeptide increases solubility of the biologically active polypeptide in a cytosolic fraction of a host cell in which the linked biologically active polypeptide is expressed. Where desired, the biologically active polypeptide is linked via a protease cleavage site to the accessory polypeptide. The cleavage site can be selected from the group consisting of TEV protease, enterokinase, Factor Xa, thrombin, PreScission™ protease, 3C protease, sortase A, and granzyme B.

The accessory polypeptide used in the subject methods or compositions can be characterized in whole or in part by the following. In one embodiment, the subject accessory polypeptide provides an average net positive charge density of the modified biologically active polypeptide of about +0.25, +0.3, +0.05, +0.075, +0.01, +0.2, +0.4, +0.5, +0.7, +0.8, +0.9 or even +1.0 charges per amino acid residue. In another embodiment, the subject accessory polypeptide provides an average net negative charge density of the modified biologically active polypeptide of about -0.25, -0.5, -0.75, -0.1, -0.2, -0.3, -0.4, -0.5, -0.6, -0.7, -0.8, -0.9 or even -1.0 average net charges per amino acid residue. In one embodiment, the subject accessory polypeptide provides a net positive charge of the modified biologically active polypeptide of about +3, +4, +5, +6, +7, +8, +9, +10, +12, +14, +16, +18, +20, +25, +30, +35, +40, +50 or more. In another embodiment, the subject accessory polypeptide provides a net negative charge of the modified biologically active polypeptide of about -3, -4, -5, -6, -7, -8, -9, -10, -12, -14, -16, -18, -20, -25, -30, -35, -40, -50 or more.

In yet another embodiment, the accessory polypeptides of the invention may comprise more than about 10, 30, 50 or 100 amino acids. In one embodiment, the accessory polypeptide comprises at least 40 contiguous amino acids and is substantially incapable of non-specific binding to a serum protein. In some embodiments, the sum of glycine (G), aspartate (D), alanine (A), serine (S), threonine (T), glutamate (E) and proline (P) and lysole (K) residues contained in the accessory polypeptide, constitutes more than about 80% of the total amino acids of the accessory polypeptide; and/or at least 50% of the amino acids in the accessory polypeptide are devoid of secondary structure as determined by the Chou-Fasman algorithm. In a related embodiment, the accessory polypeptide comprises at least 40 contiguous amino acids and the accessory polypeptide has an in vitro serum half-life greater than about 4 hours, 5 hours, 10 hours, 15 hours or 24 hours. Further wherein (a) the sum of glycine (G), aspartate (D), alanine (A), serine (S), threonine (T), glutamate (E) and proline (P) and lysole (K) residues contained in the accessory polypeptide, constitutes more than about 80% of the total amino acids of the accessory polypeptide; and/or (b) at least 50% of the amino acids in the accessory polypeptide are devoid of secondary structure as determined by Chou-Fasman algorithm. In some embodiments the set of amino acids from which the
80% (or 50, 60, 70 or 90%) of the total amino acids are chosen is chosen is G/S/E/D, G/S/K/R, G/S/E/D/K/R, or G/A/S/T/Q.

In some embodiments, an accessory polypeptide comprises at least 50% glycine residues (i.e., 50% of all residues are glycine). Alternatively, an accessory polypeptide may comprise less than 50% glycine residues. In some embodiments, accessory polypeptides comprise at least 50% serine residues. Other embodiments provide for accessory polypeptides comprising at least 50% serine and glycine residues. Further embodiments provide for accessory polypeptides which comprise at least 50% glutamic acid, or alternatively at least 10, 20 or 30% glutamic acid.

In one embodiment, an accessory polypeptide may also be characterized in that (a) it consists of three types of amino acids, and each type being selected from a group consisting of alanine (A), aspartic acid (D), glutamic acid (E), glycine (G), histidine (H), lysine (K), asparagine (N), proline (P), glutamine (Q), arginine (R), serine (S), threonine (T) and tyrosine (Y); and (b) it comprises 10, 25, 50, 100 or more amino acids. In a related embodiment, the accessory polypeptide consists of three types of amino acids, each type being selected from the group consisting of D, E, G, K, P, R, S, and T. The accessory polypeptide may also consist of three types of amino acids, each type being selected from the group consisting of E, G, and S.

The invention also provides for an accessory polypeptide characterized in that: (i) it consists of three types of amino acids, two of which are serine (S) and glycine (G) and the other type being selected from the group consisting of aspartic acid (D), glutamic acid (E), lysine (K), proline (P), arginine (R), glycine (G), threonine (T), alanine (A), histidine (H), asparagine (N), tyrosine (Y), leucine (L), valine (V), tryptophan (W), methionine (M), phenylalanine (F), isoleucine (I), and cysteine (C); and (ii) it comprises ten or more amino acid residues, of which 50% or more are serine or glycine.

In another embodiment, the accessory polypeptide is characterized in that: (a) it consists of two types of amino acids, one of which is glycine (G) and the other type is selected from the group consisting of aspartic acid (D), glutamic acid (E), lysine (K), proline (P), arginine (R), serine (S), threonine (T), alanine (A), histidine (H), asparagine (N), tyrosine (Y), leucine (L), valine (V), tryptophan (W), methionine (M), phenylalanine (F), isoleucine (I), and cysteine (C); and (b) it comprises ten or more amino acid residues, of which 50% or less are glycine.

Alternatively, the accessory polypeptide consists of two types of amino acids, wherein 50% or less of the total amino acids are selected from the group consisting of A, S, T, D, E, K and H.

In still another embodiment, the accessory polypeptide is characterized in that: (a) it comprises 50 or more amino acids; (b) it consists of two types of amino acids, and (c) 50% or less of the total amino acids are selected from the group consisting of A, S, T, D, E, K and H.

Accessory polypeptides may comprise 1, 2, 5 or 10 or more repeating motifs, each of which may comprise two to five hundred amino acids. In some cases, repeating motifs consist of two or more different types of amino acids. Multiple accessory polypeptides may be used. Accessory polypeptide may also comprise charged amino acids.

In some embodiments, the accessory polypeptide comprises an amino acid sequence (GGGES)n (SEQ ID NO: 2), wherein n is an integer of 2, 3, 4, 5, 6, 7, 8, 9, 10 or greater. In other embodiments, the accessory polypeptide comprises an amino acid sequence (GES)n, wherein n is an integer of 2, 3, 4, 5, 6, 7, 8, 9, 10 or greater. Alternatively, the accessory polypeptide comprises an amino acid sequence (GEGGEGE)n, wherein G, E, and S are in any order and n is an integer of 2, 3, 4, 5, 6, 7, 8, 9, 10 or greater. In yet another embodiment, the accessory polypeptide comprises an amino acid sequence (GEGGEGE)n, wherein G, E, and S can be in any order and n is an integer of 2, 3, 4, 5, 6, 7, 8, 9, 10 or greater. In yet another embodiment, the accessory sequence comprises minor acid sequence (GE)n, wherein G and E can be in any order and n is an integer of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or greater.

In some embodiments, the accessory polypeptide comprises an amino acid sequence (S)n (SEQ ID NO: 4), wherein n is an integer of 10, 15, 20, 50 or greater. In other embodiments, the accessory polypeptide comprises an amino acid sequence (SSSSSS)n, wherein E and S can be in any order and n is an integer of 2, 3, 4, 5, 6, 7, 8, 9, 10 or greater. In other embodiments, the accessory polypeptide comprises an amino acid sequence (SSSSSS)n, wherein E and S can be in any order and n is an integer of 3 or greater. In other embodiments, the accessory polypeptide comprises an amino acid sequence (SSSSSSSSSSSSS)n, wherein E and S can be in any order and n is an integer of 3 or greater. In still other embodiments, the accessory polypeptide comprises an amino acid sequence (SSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS
polypeptide. In one embodiment, the apparent molecular weight factor of the modified polypeptide is greater than 3. In another embodiment, the apparent molecular weight factor of the modified polypeptide is greater than 5. In yet another embodiment, the apparent molecular weight factor of the modified polypeptide is greater than 7. In still another embodiment, the apparent molecular weight factor of the modified polypeptide is greater than 9.

[0031] The accessory polypeptide can increase the serum half-life of a biologically active polypeptide. Alternatively, accessory polypeptides can increase the protease resistance of a biologically active polypeptide. In other cases, accessory polypeptides can increase the solubility of a biologically active polypeptide. In other cases, accessory polypeptides can decrease the immunogenicity of a biologically active polypeptide. The accessory polypeptides of the invention may comprise more than about 10, 30, 50 or 100 amino acids. In some embodiments, the biologically active polypeptide can be human growth hormone (hGH), glucagon-like peptide-1 (GLP-1), exenatide, pramlintide, uricase, granulocyte-colony stimulating factor (G-CSF), interferon-alpha, interferon-beta, interferon-gamma, insulin, interleukin receptor antagonist (IL-1RA), erythropoietin or tumor necrosis factor-alpha (TNF-alpha).

[0032] In one embodiment, the accessory polypeptide comprises at least 40 contiguous amino acids and is substantially incapable of non-specific binding to a serum protein. In some embodiments, the sum of glycine (G), aspartate (D), alanine (A), serine (S), threonine (T), glutamate (E) and proline (P) residues contained in the accessory polypeptide, constitutes more than about 80% of the total amino acids of the accessory polypeptide; and/or at least 50% of the amino acids in the accessory polypeptide are devoid of secondary structure as determined by the Chou-Fasman algorithm. In a related embodiment, the accessory polypeptide comprises at least 40 contiguous amino acids and the accessory polypeptide has an in vitro serum half-life greater than about 4 hours, 5 hours, 10 hours, 15 hours or 24 hours. Further wherein (a) the sum of glycine (G), aspartate (D), alanine (A), serine (S), threonine (T), glutamate (E) and proline (P) residues contained in the accessory polypeptide, constitutes more than about 80% of the total amino acids of the accessory polypeptide; and/or (b) at least 50% of the amino acids in the accessory polypeptide are devoid of secondary structure as determined by the Chou-Fasman algorithm.

[0033] In some embodiments, an accessory polypeptide comprises at least 50% glycine residues (i.e., 50% of all residues are glycine). Alternatively, an accessory polypeptide may comprise less than 50% glycine residues. In some embodiments, accessory polypeptides comprise at least 50% serine residues. Other embodiments provide for accessory polypeptides comprising at least 50% serine and glycine residues. Further embodiments provide for accessory polypeptides which comprise at least 5% glutamic acid, or alternatively at least 10, 20 or 30% glutamic acid.

[0034] In one embodiment, an accessory polypeptide may also be characterized in that (a) it consists of three types of amino acids, and each type being selected from a group consisting of alanine (A), aspartic acid (D), glutamic acid (E), glycine (G), histidine (H), lysine (K), asparagine (N), proline (P), glutamine (Q), arginine (R), serine (S), threonine (T) and tyrosine (Y); and (b) it comprises at least 10, 25, 50, 100 or more amino acids. In a related embodiment, the accessory polypeptide consists of three types of amino acids, each type being selected from the group consisting of D, E, G, K, P, R, S, and T. The accessory polypeptide may also consist of three types of amino acids, each type being selected from the group consisting of E, G, and S.

[0035] The invention also provides for an accessory polypeptide characterized in that: (i) it consists of three types of amino acids, two of which are serine (S) and glycine (G) and the other type being selected from the group consisting of aspartic acid (D), glutamic acid (E), lysine (K), proline (P), arginine (R), glycine (G), threonine (T), alanine (A), histidine (H), asparagine (N), tyrosine (Y), leucine (L), valine (V), tryptophan (W), methionine (M), phenylalanine (F), isoleucine (I), and cysteine (C); and (ii) it comprises ten or more amino acid residues, of which 50% or more are serine or glycine.

[0036] In another embodiment, the accessory polypeptide is characterized in that: (a) it consists of two types of amino acids, one of which is glycine (G) and the other type is selected from the group consisting of aspartic acid (D), glutamic acid (E), lysine (K), proline (P), arginine (R), serine (S), threonine (T), alanine (A), histidine (H), asparagine (N), tyrosine (Y), leucine (L), valine (V), tryptophan (W), methionine (M), phenylalanine (F), isoleucine (I), and cysteine (C); and (b) it comprises at least 50% or more of the total amino acid residues, of which 50% or less are glycine.

[0037] Alternatively, the accessory polypeptide consists of two types of amino acids, wherein 50% or less of the total amino acids are selected from the group consisting of A, S, T, D, E, and K.

[0038] In still another embodiment, the accessory polypeptide is characterized in that: (a) it comprises 50 or more amino acids; (b) it consists of two types of amino acids, and (c) 50% or less of the total amino acids are selected from the group consisting of A, S, T, D, E, and K.

[0039] Accessory polypeptides may comprise 1, 2, 5 or 10 or more repeating motifs, each of which may comprise two to five hundred amino acids. In some cases, repeating motifs consist of two or three or more different types of amino acids. Multiple accessory polypeptides may be used. Accessory polypeptide may also comprise charged amino acids.

[0040] In some embodiments, the accessory polypeptide comprises an amino acid sequence (GGE(GGS)n) (SEQ ID NO: 2), wherein n is an integer of 2, 3, 4, 5, 6, 7, 8, 9, 10 or greater. In other embodiments, the accessory polypeptide comprises an amino acid sequence (GES)n, wherein G, E, and S can be in any order and n is an integer of 2, 3, 4, 5, 6, 7, 8, 9, 10 or greater. Alternatively, the accessory polypeptide comprises an amino acid sequence (GGSGG)n, wherein G, E, and S can be in any order and n is an integer of 2, 3, 4, 5, 6, 7, 8, 9, 10 or greater. In yet another embodiment, the accessory polypeptide comprises an amino acid sequence (GEGG(GGE)G(n) (SEQ ID NO: 3), wherein n is an integer of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or greater. In yet another embodiment, the accessory polypeptide comprises an amino acid sequence (GE)n, wherein G and E can be in any order and n is an integer of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or greater.

[0041] In some embodiments, the accessory polypeptide comprises an amino acid sequence (S)n (SEQ ID NO: 4), wherein n is an integer of 10, 15, 20, 50 or greater. In other embodiments, the accessory polypeptide comprises an amino acid sequence (SSSSSSSSSS)n, wherein E and S can be in any order and n is an integer of 2, 3, 4, 5, 6, 7, 8, 9, 10 or greater. In other embodiments, the accessory polypeptide comprises an amino acid sequence (SSSSSSSSSS)n, wherein E and S can be in
any order and \( n \) is an integer of 2, 3, 4, 5, 6, 7, 8, 9, 10 or greater. In yet other embodiments, the accessory polypeptide comprises an amino acid sequence \( (S\underbrace{E}_{E}S\underbrace{E}_{E}S\underbrace{E}_{E}S)_{n} \), wherein \( E \) and \( S \) can be in any order and \( n \) is an integer of 2, 3, 4, 5, 6, 7, 8, 9, 10 or greater. In some embodiments, the accessory polypeptide comprises an amino acid sequence \( (S\underbrace{E}_{E}S\underbrace{S}_{S}S\underbrace{E}_{E}S\underbrace{S}_{S}S\underbrace{E}_{E}S\underbrace{S}_{S})_{n} \), wherein \( E \) and \( S \) can be in any order and \( n \) is an integer of 3 or greater. In other embodiments, the accessory polypeptide comprises an amino acid sequence \( (S\underbrace{E}_{E}S\underbrace{S}_{S}S\underbrace{S}_{S}S\underbrace{S}_{S}S\underbrace{E}_{E})_{n} \), wherein \( E \) and \( S \) can be in any order and \( n \) is an integer of 3 or greater. In still other embodiments, the accessory polypeptide comprises an amino acid sequence \((S\underbrace{E}_{E}S\underbrace{S}_{S}S\underbrace{S}_{S}S\underbrace{S}_{S}S\underbrace{S}_{S}S\underbrace{E}_{E})_{n}\), wherein \( E \) and \( S \) can be in any order and \( n \) is an integer of 3 or greater.

In some embodiments the accessory polypeptide is not composed of repeating units of a peptide motif of 3, 4, 5, 6 or 7 amino acids, or is not composed of repeating units of any single polypeptide motif. In some embodiments the accessory polypeptide is composed of more than 2, 5, 10, or 20 different repeating motifs of a fixed length. In some embodiments the accessory polypeptide is composed of more than 2, 5, 10, or 20 different repeating motifs of any length.

A slow release agent may include a polymeric matrix. In some embodiments, the polymeric matrix is charged. In specific embodiments, the polymeric matrix may be poly-D-1-lactide (PLA), poly-D,L-lactide-co-glycolide (PLGA), PLGA-PEG copolymers, alginates, dextran and/or chitosan. A slow release agent may also be packaged including a transdermal patch.

The present invention also provides a method of producing modified polypeptides, comprising: a) providing a nucleotide sequence encoding the modified polypeptide; b) causing said modified polypeptide to be expressed in a host cell, thereby producing said modified polypeptide. A genetic vehicle comprising a nucleic acid sequence encoding the modified polypeptide is also provided, as well as host cells expressing the modified polypeptides of the invention.

Additionally, the invention describes a method of making a pharmaceutical composition, comprising: (a) providing a modified polypeptide; (b) mixing said modified polypeptide with a polymer matrix.

Pharmaceutical compositions of the invention may comprise a) a slow release agent, b) a modified polypeptide comprising a biologically active polypeptide linked to a PEG group of greater than 5 KD in size.

In yet other embodiments the accessory polypeptide substantially lacks secondary structure. In still other embodiments, the accessory polypeptide exhibits a two-fold longer serum half-life as compared to a corresponding polypeptide lacking the accessory polypeptide. The biologically active polypeptide and the accessory polypeptide may be linked via a peptide bond.

In some embodiments, the modified polypeptide further comprises at least one depot module. The depot module is at least 10 amino acids in length, preferably at least 100 amino acids in length. Positively charged depot modules (e.g., lysine rich or arginine rich polypeptides) may be useful in conjunction with a negatively charged polymer. Negatively charged depot modules may be useful in conjunction with a positively charged polymer. A depot module including poly-His sequences may be used in conjunction with a chelating hydrogel. In some cases, the depot module can be protease sensitive, e.g., and without limitation, sensitive to serum proteases or other proteases. Multiple and/or different depot modules may be employed. Any combination of depot module, biologically active polypeptide and accessory polypeptides may be potentially used to produce a sustained-release therapeutic. In a particular embodiment, the slow release agent is a depot module linked to the modified polypeptide.

Additionally, a genetic vehicle comprising a nucleic acid sequence encoding an API of the invention is provided. In another embodiment, a host cell is described expressing the polypeptides.

The present invention relates to accessory polypeptides that may be used to modify the properties of biologically active polypeptides. In one embodiment, the invention provides for an isolated polypeptide comprising a biologically active polypeptide and an accessory polypeptide, wherein the accessory polypeptide is characterized in that it (i) consists of three types of amino acids, and each type being selected from a group consisting of alanine (A), aspartic acid (D), glutamic acid (E), glycine (G), histidine (H), lysine (K), asparagine (N), proline (P), glutamine (Q), arginine (R), serine (S), threonine (T) and tyrosine (Y); and (ii) comprises ten or more amino acids. In a related embodiment, the accessory polypeptide consists of three types of amino acids, and each type being selected from a group consisting of D, E, G, K, P, R, S, and T. In another related embodiment, the accessory polypeptide consists of three types of amino acids, and each type being selected from a group consisting of E, S, G, R, and A. In another related embodiment, the accessory polypeptide consists of three types of amino acids, and each type being selected from a group consisting of E, S, G, R, and A. In yet another embodiment, the accessory polypeptide consists of three types of amino acids, and each type being selected from a group consisting of E, G, and S. The isolated polypeptide may be a therapeutic polypeptide.

The invention also provides for isolated polypeptides comprising a biologically active polypeptide and an accessory polypeptide, wherein the accessory polypeptide is characterized in that: (i) it is poly-serine, and (ii) it comprises ten or more amino acids. In a related embodiment, the isolated polypeptide (i) consists of two types of amino acids, the majority of which are serine, and (ii) it comprises ten or more amino acids.

The accessory polypeptide consists of two types of amino acids, one of which is glycine (G) and the other type is selected from the group consisting of aspartic acid (D), glutamic acid (E), lysine (K), proline (P), arginine (R), serine (S), threonine (T), alanine (A), histidine (H), asparagine (N), tyrosine (Y), leucine (L), valine (V), tryptophan (W), methionine (M), phenylalanine (F), isoleucine (I), and cysteine (C); and (ii) it comprises ten or more amino acid residues, of which 50% or less are glycine.

The invention also provides for isolated polypeptides comprising a biologically active polypeptide and an accessory polypeptide, wherein the accessory polypeptide is characterized in that: (i) it consists of two types of amino acids, one of which is serine (S) and the other type is selected from the group consisting of aspartic acid (D), glutamic acid (E), lysine (K), proline (P), arginine (R), glycine (G), threonine (T), alanine (A), histidine (H), asparagine (N), tyrosine (Y), leucine (L), valine (V), tryptophan (W), methionine (M), phenylalanine (F), isoleucine (I), and cysteine (C); and (ii) it comprises ten or more amino acid residues, of which 50% or more are serine.

Alternatively, the invention describes an isolated polypeptide comprising a biologically active polypeptide and
an accessory polypeptide, wherein the accessory polypeptide is characterized in that: (i) it comprises ten or more amino acids; (ii) it consists of two types of amino acids, wherein 50% or less of the total amino acids are selected from the group consisting of A, S, T, D, E, and H. 0055 In yet another embodiment, the invention describes an isolated polypeptide comprising a biologically active polypeptide and an accessory polypeptide, wherein the accessory polypeptide is characterized in that: (i) it comprises ten or more amino acids; (ii) it consists of two types of amino acids, 50% or less of the total amino acids are selected from the group consisting of A, G, T, D, E, and H.

0056 In some embodiments, an isolated polypeptide is provided comprising a biologically active polypeptide and an accessory polypeptide, wherein the accessory polypeptide is characterized in that: (i) it consists of two types of amino acids, one of which is selected from the group consisting of P, R, L, V, Y, W, M, F, I, K, and C; and (ii) it comprises ten or more amino acids.

0057 In other embodiments, an isolated polypeptide is provided comprising a biologically active polypeptide and an accessory polypeptide, wherein the accessory polypeptide comprises at least 10 amino acids in length and consists of two different types of amino acids represented in equal numbers. Alternatively, the two different types of amino acids are represented in 1:2, 2:3, or 3:4 ratio. The accessory polypeptide may additionally comprise four or more repeating motifs, each of which comprises two to five hundred amino acids and is made of two different types of amino acids. The repeating motif may comprise more than 8 amino acids, and in some embodiments four or more of the repeating motifs are identical. The four or more repeating motifs may comprise different amino acid sequences. In a related embodiment, the accessory polypeptide comprises at least ten repeating motifs.

0058 Yet other embodiments provide biologically active polypeptides modified with accessory polypeptides which substantially lack secondary structure. Alternatively, the apparent molecular weight of the isolated polypeptides is greater than that of a corresponding polypeptide lacking the accessory polypeptide. In a particular embodiment, the apparent molecular weight of the accessory polypeptide is at least 3 times greater than its actual molecular weight. In still other embodiments, the accessory polypeptide exhibits a two-fold longer serum half-life as compared to a corresponding polypeptide lacking the accessory polypeptide. The biologically active polypeptide and the accessory polypeptide may be linked via a peptide bond.

0059 In some embodiments, the accessory polypeptide comprises an amino acid sequence (GEGGEG)n (SEQ ID NO: 5), wherein n is an integer of 3 or greater. In other embodiments, the accessory polypeptide comprises an amino acid sequence (GES)n, wherein G, E, and S can be in any order and n is an integer of 3 or greater. Alternatively, the accessory polypeptide comprises an amino acid sequence (GSGGGE)n, wherein G, E, and S can be in any order and n is an integer of 3 or greater. In yet another embodiment, the accessory polypeptide comprises an amino acid sequence (GEGGE)n, wherein G and E can be in any order.

0060 In some embodiments, the accessory polypeptide comprises an amino acid sequence (S)n (SEQ ID NO: 7), wherein n is an integer of 10 or greater. In other embodiments, the accessory polypeptide comprises an amino acid sequence (SSSSSEN)n, wherein E and S can be in any order and n is an integer of 2 or greater. In yet other embodiments, the accessory polypeptide comprises an amino acid sequence (SESSESSE)n, wherein E and S can be in any order and n is an integer of 3 or greater. In some embodiments, the accessory polypeptide comprises an amino acid sequence (SSESSESSSE)n, wherein E and S can be in any order and n is an integer of 3 or greater. In still other embodiments, the accessory polypeptide comprises an amino acid sequence (SSSSSSSSSESSSS)n, wherein E and S can be in any order and n is an integer of 3 or greater.

0061 The present invention also provides a method of producing an isolated polypeptide, comprising: a) providing a polynucleotide sequence encoding the isolated polypeptide of any one of claim 1, 6, 7, 8, or 9; b) causing said polypeptide to be expressed in a host cell, thereby producing said polypeptide.

0062 Additionally, a genetic vehicle comprising a nucleic acid sequence encoding the isolated polypeptides of the invention is provided. In another embodiment, a host cell is described expressing the subject polypeptides. Libraries of subject polypeptides are also envisioned. In a particular embodiment, libraries of polypeptides are displayed on phage particles.

INTEGRATION BY REFERENCE

0063 All publications and patent applications mentioned in this specification are herein incorporated by reference to the same extent as if each individual publication or patent application was specifically and individually indicated to be incorporated by reference.

BRIEF DESCRIPTION OF THE DRAWINGS

0064 The novel features of the invention are set forth with particularity in the appended claims. A better understanding of the features and advantages of the present invention will be obtained by reference to the following detailed description that sets forth illustrative embodiments, in which the principles of the invention are utilized, and the accompanying drawings of which:

0065 FIG. 1 is an illustrative representation of an accessory polypeptide modifying a biologically active protein.

0066 FIGS. 2 and 3 show possible modules for inclusion in modified polypeptides of the invention: accessory polypeptide(s), biologically active polypeptide(s), optional depot module(s) and optional polymeric matrix or matrices.

0067 FIG. 4 shows examples of various product configurations. Modules may be used several times in the same product, for example to increase affinity of the biologically active protein for its target, to increase half-life by extending the rPEG module, or to modify the properties of the depot formulation.

0068 FIG. 5 presents a specific example of a tetrameric modified polypeptide comprising a depot module that allows for site-specific biotinylation. The addition of streptavidin induces the formation of highly stable, yet non-covalent, modified polypeptide tetramers. Multivalent polypeptides can also be created by combining multiple modules into a
single protein chain or by chemically linking multiple protein chains containing a specific module.

[0069] FIG. 6 illustrates a lysine- or arginine-rich depot module (depicted as rectangles) which may be incorporated into the polymer matrix of an alginate microsphere. The matrix module is depicted as larger circles. The lysine- or arginine-rich depot will carry a net positive charge at physiological pH and this property can be exploited to bind the modified polypeptide to the negatively charged alginate polymer. Binding may occur in a multivalent fashion.

[0070] FIG. 7 illustrates a divalent cation chelating hydrogel (matrix module) exemplified by the divalent cation Cu\(^{2+}\) bound to the polymer. The polyhistidine depot module (rectangular module) binds with high affinity to the Cu\(^{2+}\) cations.

[0071] FIG. 8 depicts a protease sensitive multimeric modified polypeptide. The depot module (depicted by a rectangle) connects individual modified polypeptide units in an extended polymer. The depot module is designed such that it is specifically sensitive to serum proteases. Protease cleavage of the depot module releases individual active modified polypeptides.

[0072] FIG. 9 shows the design of the expression vector pCW0150. FIG. 9 discloses the "6×His-tag" and "H6" sequence as SEQ ID NO: 1.

[0073] FIG. 10 shows the design and construction of the accessory polypeptide rPEG(1,288) fused to GFP. FIG. 10 discloses the "110" sequence as SEQ ID NO: 1.

[0074] FIG. 11 shows the amino acid (SEQ ID NO: 485) and nucleotide sequence (SEQ ID NO: 484) of the rPEG_1,288 polypeptide.

[0075] FIG. 12 shows the design of hGH-rPEG(1,288) and GLP-1-rPEG(1,288) constructs.

[0076] FIG. 13 shows examples biologically active proteins conjugated to accessory polypeptides (SEQ ID NOS 486-489, respectively, in order of appearance).

[0077] FIGS. 14 and 15 describe exemplary guidelines for sequence optimization of accessory polypeptides. FIG. 15 discloses SEQ ID NOS 490-498, respectively, in order or appearance.

[0078] FIG. 16 describes the construction of a vector comprising the rPEG_J288 accessory polypeptide sequence fused to GFP. FIG. 16 discloses the "H6" sequences as SEQ ID NO: 1.

[0079] FIG. 17 shows the amino acid (SEQ ID NO: 500) and nucleotide sequence (SEQ ID NO: 499) of the rPEG_J288 polypeptide.

[0080] FIG. 18 shows the design of a stuffer vector suitable for use in the present invention. FIG. 18 discloses the amino acid sequences as SEQ ID NOS 502 and 503, respectively, the nucleotide sequence as SEQ ID NO: 501 and the "6×His-tag" as SEQ ID NO: 1.

[0081] FIG. 19 shows the purification of rPEG_J288-modified GFP.

[0082] FIG. 20 shows the determination of serum stability of rPEG_J288-modified GFP.

[0083] FIG. 21 shows the interaction of an accessory-modified polypeptide with a cellular target.

[0084] FIG. 22 illustrates the concept of crosslinked accessory polypeptides.

[0085] FIG. 23 describes examples of crosslinking components.

[0086] FIG. 24 lists several examples of crosslinked accessory polypeptides.

[0087] FIG. 25 shows an example wherein streptavidin is used as a linker.

[0088] FIG. 26 describes different modalities of constructing crosslinked accessory polypeptides.

[0089] FIG. 27 identifies illustrates several possible formats of crosslinked accessory polypeptides.

[0090] FIG. 28 describes accessory polypeptides additionally modified with binding domains or other groups.

[0091] FIG. 29 illustrates the concept of slow-release accessory polypeptides.

[0092] FIG. 30 shows universal accessory polypeptides. FIG. 30 discloses the "KKKKKK" sequences as SEQ ID NO: 504.

[0093] FIG. 31 shows an antibody Fc fragment from human IgG1, but this could also be from IgG2, IgG3, IgG4, IgA, IgD or IgE. This Fc can have a native hinge from IgG1, IgG2, IgG3, IgG4, IgA, IgD or IgE. There is natural diversity in the number of hinge disulfides, but this can also be created by mutation, deletion, or truncation of the hinge, especially the cysteine residues. The variants that are useful have either three disulfides (not shown), two disulfides, one disulfide (choice of first or second natural one of IgG1) or no disulfides.

[0094] FIG. 32 illustrates various configurations of modified polypeptides comprising affinity tags, solubility tags and/or protease cleavage sites.

[0095] FIG. 33 illustrates improved expression levels of modified polypeptides using specific accessory polypeptides.

[0096] FIG. 34 illustrates shows activity of an accessory-modified hGH polypeptide relative to unmodified hGH.

[0097] FIG. 35 shows purification of accessory-modified polypeptides by anion exchange and size exclusion chromatography.

[0098] FIG. 36 shows pure product obtained by purification of rPEG-modified GFP as confirmed by SDS-PAGE.

[0099] FIG. 37 shows the purity of rPEG-linked GLP1 as ascertained by analytical size exclusion chromatography.

[0100] FIG. 38 shows the purity of rPEG_J288-GFP modified polypeptide as observed by analytical reverse-phase HPLC.

[0101] FIG. 39 Mass spectrometry of rPEG J288-GFP.

[0102] FIG. 40 demonstrates that little nonspecific binding is observed between modified polypeptides and serum proteins.

[0103] FIG. 41 describes the increase in apparent molecular weight observed upon linking a biologically active polypeptide to an accessory polypeptide.

[0104] FIG. 42 shows the stability of modified polypeptides in rat and human serum.

[0105] FIG. 43 illustrates a PK profile of rPEG_K288-GFP polypeptide in rat serum.

[0106] FIG. 44 describes the relative lack of immunogenicity of rPEG polypeptides as determined in animal experiments for rPEG J288-GFP, rPEG_K288-GFP and rPEG_J,288-GFP.

[0107] FIG. 45 illustrates the advantage of expressing biologically active polypeptides linked to accessory polypeptides.

[0108] FIG. 46 illustrates sustained release of accessory-modified polypeptides.

[0109] FIG. 47 shows the purity of rPEG J288-GLP1 polypeptide as determined by size exclusion chromatography (multiple injections per run).
FIG. 48 shows the purity of rPEG_K288-GLP1 polypeptide as determined by size exclusion chromatography (multiple injections per run).

FIG. 49 describes the increase in apparent molecular weight observed upon linking a biologically active polypeptide (GLP1) to rPEG(J288, rPEG_K288, and rPEG_L288 accessory polypeptides.)

FIG. 50 shows the products obtained through protease cleavage of a polypeptide comprising an affinity tag, an accessory polypeptide and hGH as a biologically active polypeptide (rPEG_K288-hGH). The protease removes the Tag, while leaving a final product which is hGH linked to the rPEG_K288 accessory polypeptide.

FIG. 51 shows the purity of rPEG_K288-hGH after protease cleavage and further purification.

FIG. 52 shows the structure of a whole IgG1, but IgG2, IgG3, IgG4, IgE, IgD, IgA and IgM can similarly be used as starting points. A dAb-dAb-Fc fusion protein is also useful because of its tetravalency; it is not shown.

FIG. 53 Constructs are shown with rPEG separating the Fc and antigen binding domains, and the Fc at the C-terminal: (dAb/sFcV)-rPEG-Fc and (dAb/sFcV)-dAb/rPEG-Fc. However, formats with a different order of the same elements are also useful, like rPEG-Fc-(dAb/sFcV), rPEG-Fc-(dAb/sFcV)-dAb, rPEG-(dAb/sFcV)-dAb/rPEG-Fc, Fc-rPEG-(dAb/sFcV)-dAb, Fc-rPEG-(dAb/sFcV)-dAb/rPEG-Fc, Fc-(dAb/sFcV)-rPEG-Fc, and (dAb/sFcV)-rPEG-Fc-rPEG. One can also mix scFv and dAbs, like dAb-sFcV or scFv-dAb or combine two scFvs or two dAbs of different target specificities: scFv1-sFcV2 or dAb1-dAb2.

FIG. 53a shows a scFv-Fc fusion protein. FIG. 53b shows a dAb-Fc fusion protein. FIG. 53c shows a scFv-sFcV-Fc fusion protein, which is tetravalent.

FIG. 54 shows a dimer of a scFv fragment. Both heterodimers and homodimers can be constructed.

FIG. 55 single chain diabody

FIG. 56 shows an example of a single chain Fc fragment. Optionally, biologically active proteins can be fused to either terminus of this construct.

FIG. 56 shows a scFv fragment. FIG. 57 shows a single chain diabody

FIG. 58: Structure of AFBTs. 58a: Monovalent AFBT; 58b: Structure of a bispecific AFBT

FIG. 59: Multivalent binding of an AFBT to a target antigen

FIG. 60a: Multivalent AFBT containing antibody fragments derived from two parent antibodies;

FIG. 60b: Structure of an AFBT comprising a diabody and a payload

FIG. 61: Preparation of a semisynthetic AFBT

FIG. 62: Purification, characterization and binding activity of an anti Her-2 scFv fused to rPEG50. 62a: binding activity. Filled diamonds: binding to coated Her-2; open diamonds: binding to coated IgG.

FIG. 62b: Size exclusion chromatography; 62c: Detection of free SH groups.

FIG. 63: Purification, characterization and binding activity of an anti Her-2 diabody, alHer203-rPEG50. 63a: binding activity. Filled diamonds: binding to coated Her-2; open diamonds: binding to coated IgG.

FIG. 63b: Size exclusion chromatography of diabody alHer203-rPEG50 and scFv alHer230-rPEG50, 63c: SEC of alHer203-rPEG50 over time shows no increase in higher multimers.

FIG. 64: Construction, sequence, and expression of scFv-rPEG50 fusion proteins. 64a: Cartoon of the protein architecture (FIG. 64a discloses the “His6 tag” sequence as SEQ ID NO: 1); 64b: sequence (SEQ ID NO: 505) of an AFBT with specificity for Her-2; 64c: SDS/PAGE showing the expression of scFv-rPEG50 fusion proteins; 64d: sequence (SEQ ID NO: 506) of an AFBT with specificity for EGF.

FIG. 65: Construction, sequence, and expression of a diabody-rPEG50 fusion proteins, alHer203-rPEG50. 65a: Cartoon of the protein architecture (FIG. 65a discloses the “His6 tag” as SEQ ID NO: 1); 65b: protein sequence (SEQ ID NO: 507); 65c: SDS/PAGE demonstrating the expression of fusion protein in the cytosol of E. coli.

FIG. 66: Codon optimization of an Fc domain for bacterial expression. 66a: Illustration of the process and oligonucleotide design. The sequence encoding the human Fc was assembled from semi-random oligonucleotides and cloned in front of rPEG25 and GFP that served as reported.

FIG. 67: Cartoon illustrating expression constructs for Fab-rPEG fusion proteins.

FIG. 68: Flow chart of the discovery process for AFBTs from antisera.

FIG. 69: Amino acid sequence (SEQ ID NO: 510) of GFP-rPEG50. The sequence of GFP is underlined.

FIG. 70: Pharmacokinetics of GFP-rPEG50 and Ex4-rPEG50 in cynomologus monkeys.

FIG. 71a: Amino acid sequence (SEQ ID NO: 511) of the CDB-Ex4-rPEG50 fusion protein. FIG. 71b: Illustration of the process used to liberate Ex4-rPEG50 from the fusion sequence shown in FIG. 14a.

FIG. 72: Immunogenicity of Ex4-rPEG50 in mice. FIG. 72a illustrates the time course of injections and blood sample analyses. FIG. 72b shows ELISA analyses of blood samples at 1,500 dilution. FIG. 72c: ELISA analyses of blood samples at 1,12,500 dilution.

FIG. 73: Size exclusion chromatography of GFP-rPEG25 and GFP-rPEG50. Grey line indicates molecular weigh standard using globular proteins.

FIG. 74: Comparison of the interaction of repetitive and non-repetitive URP-B cells. FIG. 74a shows a repetitive URP that is composed of multiple identical sequence repeats. Such a repetitive URP can form multivalent contacts with B cells that recognize the repeating sequence, which can trigger B cell proliferation. FIG. 74b shows a non-repetitive URP that is composed of multiple different subsequences. Each subsequence can be recognized by a particular subset of B-cells with cognate specificity. However, an individual molecule of a non-repetitive URP can only form one or few interactions with any particular B cell, which is unlikely to trigger proliferation.

FIG. 75: Algorithm to assess the repetitiveness of an amino acid sequence.

FIG. 76: Computer algorithm to design nrURPs with very low repetitiveness.

FIG. 77: Construction of nrURPs from libraries of URP segments.

FIG. 78: Amino acid sequences (SEQ ID NOS 15, 18, 16, 19, 17, 20, 11, 13, 12 and 14, respectively, in order of appearance) that were used to construct rPEG_Y. The figure
also indicates the relative concentrations of oligonucleotides that were used to construct the segment libraries.

FIG. 79: Assembly of URP segments from synthetic oligonucleotides. FIG. 79a shows the ligation reaction. Repeating segments are encoded by partially overlapping oligonucleotides that are phosphorylated. A second pair of annealed oligonucleotides is added to terminate chain elongation. One of these capping oligonucleotides is not phosphorylated, which prevents ligation at one end. FIG. 79b shows an agarose gel of a ligation reaction.

FIG. 80: Examples of URP_Y144 sequences (SEQ ID NOs 512-521, respectively, in order or appearance).

FIG. 81: Amino acid sequence (SEQ ID NO: 522) encoded by plasmid pCW0279. The open reading frame encodes a fusion protein of Flag-URP_Y576-GFP. The amino acid sequence of URP_Y576 is underlined.

FIG. 82 shows general ways of making 'rPEG linked binding pairs', which have the advantage of initial activity and therefore no burst release effect (increasing the dose that can be administered without toxicity) and reduced initial receptor-mediated clearance. The general binding pairs can be receptor-ligand, antibody-ligand, or generally binding protein 1—binding protein 2. The construct can have a cleavage site, which can be cleaved before injection, after injection (in serum by proteases) and can be located such that the rPEG stays with the therapeutically active protein (active protein), which can be either the ligand, the receptor or the antibody.

FIG. 83a shows a construct with a drug module at the N-terminus, followed by rPEG, fused to an antibody fragment, including the hinge. The Fc fragments provides long half-life and the rPEG allows the Fc fragment to be expressed in the E. coli cytoplasm in soluble and active form.

FIG. 83b shows a construct with a drug module at the N-terminus, followed by rPEG, fused to an antibody Fc fragment, but without the hinge. The Fc fragments provides long half-life and the rPEG allows the Fc fragment to be expressed in the E. coli cytoplasm in soluble and active form.

FIG. 84a: A Diabody is formed when the single chain linker between the VH and VL domain is shorter than about 10-20AA, preventing the formation of a single chain Fv fragment. A diabody has two protein chains and can have an rPEG at one or both C-terminal ends, and/or at one or both N-terminal ends. The diabody has two binding sites, of which zero, one or two may bind to a pharmaceutical target, or to a half-life target (i.e. HSA, IgG, Red Blood Cells, Collagen, etc) or to no target.

FIG. 84b: The diabody may contain zero, one or more drug modules located at the N-terminal or C-terminal end of zero, one or both protein chains.

FIG. 85a shows a single chain Fv fragment, to which a drug module (like IFNα, hGH, etc) can be fused at one or both of the N- and/or C-terminal ends. The scFv has one binding site, which may or may not bind to a pharmaceutical target, or to a half-life target (i.e. HSA (see FIG. 85b), IgG, Red Blood Cells, etc)

FIG. 86 shows the use of rPEG to associate two proteins that belong to the same complex. The affinity between such proteins is often insufficient to keep them associated, but the addition of rPEG stabilizes their interaction and reduces their tendency to form polymers.

FIG. 87 shows a Fab fragment binding to a cell-surface target; the H chain may be fused to Fc (like in whole antibodies) or to a wide variety of other proteins, domains and peptides. Extension of the length of the natural linkers from the usual 2-6 amino acids to 4, 5, 6, 7, 8, 9, 10, 11, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 24, 26, 28, 30, 35, 40, 45, 50, 50, 50, 50, 70, 80, 100 or more amino acids, between the VH and the CH domains, and between the VL and the CL domains, increases the ability of one Fab to crosslink to another Fab by domain swapping, thereby forming a binding complex with higher valency, resulting in higher apparent affinity (avidity). The linker may be rPEG or a different composition. This “Extended Linker” format allows binding with increased affinity specifically at sites with a higher density of target, such as (partially) tumor-specific antigen on tumor cells.

FIG. 88 shows how an association peptide, such as SKVILF(E) (SEQ ID NO: 8) or RARADADA (SEQ ID NO: 9), which bind to another copy of the same sequence in an antiparallel orientation, can be used to create a produg. In this case the drug is protease-cleaved in the last manufacturing step, but the cleavage does not activate the drug since the two chains are still associated by the association peptides. Only after the drug is injected into the blood and the concentration is greatly reduced, the small, non-rPEG-containing protein chain will leave the complex (at a rate that depends on affinity, especially the off-rate) and is likely to be cleaved via the kidney, thereby activating the r-PEG-containing drug module.

FIG. 89 shows the proteolytic cleavage which converts the manufactured single-chain protein into a complex of two protein chains. This cleavage can occur as the last manufacturing step (before injection) or it can occur after injection, by proteases in the patient’s blood.

FIG. 90a shows an antibody Fc fragment, with a hinge region, (optionally) fused to a drug module (e.g. IFNα, hGH, etc.) on one end and (optionally) fused to rPEG on the other end. The sequence between CH2 and CH3 mediates binding to FcRn, the neonatal Fc receptor, unless that function is removed by mutation. FIG. 90b shows a similar construct but without the hinge region.

FIG. 91a shows a protein construct comprising a paired pair of CH3 domains; zero, one or both of these chains may be fused to rPEG on the N-terminal and/or C-terminal end, and to zero, one or more drug modules at the other end. The FcRn binding sequence can either be retained or deleted; retention should yield a longer serum half-life.

FIG. 91b shows a similar protein, but CH2 was fully removed so that the binding of the Fc to the FcRn receptor is no longer functional, reducing half-life.

FIG. 92a shows a protein that is a full Fc, including a hinge, CH2 and CH3 domains, fused at the c-terminus to an rPEG, with the drug/pharmacophore located at the C-terminus.

FIG. 92b shows a full Fc, but without a hinge fused at the c-terminus to an rPEG, with the drug/pharmacophore located at the C-terminus; these molecules can chain swap, potentially resulting in hetero-dimers.

FIG. 93a shows a partial Fc, without hinge and with a CH2 that is truncated but retains FcRn binding and with the drug/pharmacophore located at the C-terminus.

FIG. 93b shows a partial Fc, without hinge and CH2, but retaining CH3 and with the drug/pharmacophore located at the C-terminus. This does not bind FcRn but can dimerize via the CH3 domain.

FIG. 94a shows an rPEG flanked by identical receptor domains (or domains having the same binding function, or domains that can bind simultaneously to the same target). If
both receptors can bind the target simultaneously, then the binding of one receptor stabilizes the binding of the second receptor and the effective/apparent affinity/avidity is increased, typically by 10-100-fold, but at least 3-fold. The rPEG provides serum half-life. One option is to pre-load the product with a ligand. In that case the injected product is inactive for as long as it remains bound to the ligand. This approach reduces peak dose toxicity and also reduces receptor-mediated clearance and may thus be useful in application where this is important.

[0164] FIG. 94 shows a product with rPEG flanked by two different receptors that can bind the ligand simultaneously, which results in mutual stabilization of the complex and increased apparent affinity (avidity), with the rPEG serving as a valency bridge that increases the effective concentration of the receptors.

[0165] FIG. 94: One option is to pre-load the product with a ligand. In that case the injected product is inactive for as long as it remains bound to the ligand. When the ligand unbinds, it is likely to be rapidly cleared via the kidney, resulting in activation of the product, which has a long half-life because of the rPEG tail. This approach reduces peak dose toxicity and also reduces receptor-mediated clearance and may thus be useful in application where this is important.

[0166] As shown in FIG. 94, some pro-drug formats do not need a cleavage or other activation site. A single protein chain can contain two (or more) drug modules separated by rPEG; these modules can be the same (of a single type) or of two or more different types. All drug modules are receptor or all are ligand. This rPEG containing product is complexed with a second, complementary protein to form a receptor-ligand-receptor interaction. In this format the ligand is likely to be dimeric or multimeric, but can also be monomeric, especially if the two drug modules are different. Both modules bind to a third protein. X and Y can be the same or different, and X and Y can be the drug module or bind to the drug module. In each case in FIG. 94, X and Y (and rPEG) comprise one protein chain, and the molecule they bind to is a separate molecule, typically protein or small molecule. It is possible to have more than two binding proteins combined in a single protein chain. The idea is that the complex of a large rPEG-containing protein and a non-rPEG containing protein is inactive when injected, but over 2-24 hours the smaller, non-rPEG-containing protein leaves the complex and is excreted via the kidney, thereby activating the drug module(s). The benefit of this format is that it reduces or removes the initial spike in drug concentration and the associated safety issues, and that the complex minimizes the receptor-mediated clearance while it is complexed, thereby extending the serum secretion half-life.

[0167] FIG. 95 shows an rPEGs flanked on both sides by a VEGF-receptors. Since VEGF is dimeric, this can be the same receptor on both sides of the rPEG, or a different receptor (preferably VEGF-R1 and VEGF-R2, but VEGF-R3 can also be used.

[0168] FIG. 96 shows drug products that are either manufactured (cleaved before injection) or administered as an inactive prodrug (cleaved after injection, in the blood). The inactivation of the drug is mediated by a binding protein that is linked to the drug by rPEG, so that all three modules are manufactured as a single protein chain. If the drug is a receptor, then the binding protein is a ligand (peptide or protein) of that receptor; if the drug is an antibody fragment, then the binding site is a peptide or protein ligand. In these examples, the drug is activated by protease cleavage of a site between the two binding domains, called X and Y. If Protein Y is the active product, then Y must retain the rPEG and the protease cleavage site must be (between X and Y, but) close to X. If Protein X is the active product, then X must retain the rPEG and the cleavage site must be close to Y. There can be one or multiple cleavage sites, as shown by the blue crossbars. The drug module can be a receptor, a ligand, one or more Ig domains, an antibody fragment, a peptide, a microprotein, an epitope for an antibody. The protein that binds to the drug module can be a binding protein, a receptor, a ligand, one or more Ig domains, an antibody fragment, a peptide, a microprotein, an epitope for an antibody. FIG. 96 discloses the “SVILFE” sequence as SEQ ID NO: 524 and the “RARADADA” sequence as SEQ ID NO: 9.

[0169] FIG. 97 shows how an inactive pro-drug can be created by adding a binding peptide to a drug module. The peptide must neutralize the target binding capacity of the drug and the peptide is gradually cleared from the blood at a higher rate than the rPEG-containing drug. Such a peptide can be natural but more typically it would be obtained by phage panning of random peptide libraries against the drug module. The peptide would preferably be made synthetically, but it can be recombinant.

[0170] FIG. 98 shows a single-chain protein drug containing multiple bio-active peptides, which can be at the same end of rPEG or at opposite ends of rPEG. These peptides can have the same activity or different activities. The purpose of having multiple peptides in a single chain is to increase their effective potency through binding avidity, without complicating manufacturing.

[0171] FIG. 99 shows how a Pro-drug-rPEG can increase serum half-life by avoiding receptor-mediated clearance.

[0172] FIG. 100 shows how drug concentration changes over time after IV injection. The goal in typical therapies is maintain the drug at a concentration that is higher than the therapeutic does, but lower than the toxic dose. A typical bolus injection (IV, IM, SC, IP or similar) of a drug with a short half-life results in a peak concentration that is much higher than the toxic dose, followed by an elimination phase that causes the drug concentration to rapidly drop below the therapeutic dose. This PK profile tends to cause toxicity and long periods of ineffective treatment, while the drug is present at therapeutic concentrations for only short time (blue line). The addition of rPEG to a drug decreases the peak concentration and thereby decreases toxicity, and increases the period of time that the drug is present at a therapeutic, non-toxic dose. The creation of a Pro-drug by addition of rPEG plus a drug-binding protein can prevent the "burst release" or toxic peak dose (red line), because the drug is only gradually activated over several hours and the length of time between the toxic dose and the therapeutic dose is increased compared to other formats.

[0173] FIG. 101 shows an N-terminal drug module followed by rPEG and a C-terminal Fc fragment (with hinge). This is a useful format for half-life extension of drug modules that can still be manufactured in the E. coli cytoplasm.

[0174] FIG. 102a shows an alternative format for a Prodrug containing an Fc fragment. The format is similar as described in FIG. 101, with the addition (at the N-terminus) of an inhibitory sequence (in blue) that binds to and inhibits the drug sequence (in red). As before, the drug is separated from the inhibitory sequence by a cleavage site. The N-terminal inhibitory binding sequence is followed by a cleavage site, which is followed by the drug sequence (in red). Before
cleavage, the drug is bound to the inhibitory sequence and thus inactive (pro-drug). After cleavage, the inhibitory binding sequence (blue) is gradually released and cleared, gradually increasing the amount of time that the drug (red) is active. [0175] FIG. 102b shows an alternative Pro-drug format containing an Fc fragment. The formats is similar to the format described in FIG. 101, again with the addition of an inhibitory binding sequence (peptide or domain, shown in red, typically positioned in or near the rPEG) which is separated from the drug (shown in blue) by a cleavage site. Before cleavage, the drug is bound to the inhibitory sequence and thus inactive (pro-drug). After cleavage, the inhibitory binding sequence (blue) is gradually released and cleared, gradually increasing the amount of time that the drug (blue) is active.

[0176] FIG. 103a-d shows the preferred fusion Sites for rPEG to an intact, Whole Antibody (incl. IgG1, 2, 3, 4, IgE, IgA, IgD, IgM). These sites indicated are preferred because they are at the boundary of structured sequences, such as domains, hinges, etc., without disturbing the folding of these functional domains. rPEG can thus be added in 1, 2, 3, 4, 5, 6, 7, 8 or more locations to an antibody (and more than 8 for IgM and IgG) and a single antibody can have 1, 2, 3, 4, 5, 6, 7, 8 or more rPEGs in diverse locations and in any combination of the 8 locations shown.

[0177] FIG. 103c shows the Preferred Fusion Sites for rPEG to Domains and Fragments of an Antibody (IgG1, 2, 3, 4, IgA, IgD, IgM). Fusion sites for N-terminal and/or C-terminal addition of rPEG are shown with red arrows or red lines.

[0178] FIG. 104 shows assays for correct folding of Fc fragments.

[0179] FIG. 105 shows the conversion of an inactive protein to an active protein by a site specific protease, either in serum or before injection. In this example the red sequence is the active therapeutic.

[0180] FIG. 106 shows the conversion of an inactive drug to an active drug by a site specific protease. In this example the blue domain (dAb, seFv, etc.) is the therapeutic entity.

DETAILED DESCRIPTION OF THE INVENTION

[0181] The present invention makes use of the unexpected discovery that biologically active polypeptides modified with accessory polypeptides may have the property of remaining soluble in the cytoplasm and folding into their active form, in conditions in which a biologically active polypeptide without such a modification would aggregate and form inclusion bodies. The methods of the invention may be useful for, among other applications, high throughput screening of proteins in the design phase, the manufacturing of proteins that currently require periplasmic expression, and for manufacturing of proteins that are difficult to refold from aggregates such as inclusion bodies. The invention discloses methods of designing accessory protein sequences, recombinant DNA molecules encoding modified polypeptide, expression vectors for such polypeptides, host cells for expression of such polypeptides and purification processes. For example, the fusion of a long hydrophilic polypeptide sequence to proteins, which may include peptides, proteins, antibodies, and vaccines, and may be eukaryotic or mammalian proteins, results in a soluble fusion protein showing improved folding in the cytoplasm in active form.

[0182] Accessory polypeptides of the invention may be linked to pharmaceutical proteins including GCSF, growth hormone, interferon alpha and to antibody fragments. These four proteins or classes of proteins typically form inclusion bodies when expressed in the cytoplasm of E. coli. However, when linked to a long hydrophilic accessory polypeptide sequence, the folding properties of the biologically active polypeptides may be greatly improved, leading to a greatly increased fraction able to fold correctly into active protein within the cell, as opposed to immediate irreversible aggregation into inclusion bodies which typically occurs for eukaryotic proteins in the absence of an accessory protein. Accessory polypeptides may additionally comprise affinity tags for protein purification by ion exchange, alone or in combination with other known purification tags, such as chitin binding domain, cellulose binding domain, MBP, GST or His-tags.

[0183] This and other aspects of the invention will be described in further detail below.

[0184] General Techniques:


DEFINITIONS

[0186] As used in the specification and claims, the singular form “a”, “an” and “the” include plural references unless the context clearly dictates otherwise. For example, the term “a cell” includes a plurality of cells, including mixtures thereof.

[0187] The terms “polypeptide”, “peptide”, “amino acid sequence” and “protein” are used interchangeably herein to refer to polymers of amino acids of any length. The polymer may be linear or branched, it may comprise modified amino acids, and it may be interrupted by non-amino acids. The terms also encompass an amino acid polymer that has been modified, for example, by disulfide bond formation, glycosylation, lipidation, acetylation, phosphorylation, or any other manipulation, such as conjugation with a labeling component. As used herein the term “amino acid” refers to either natural and/or unnatural or synthetic amino acids, including but not limited to glycine and both the D or L optical isomers, and amino acid analogs and peptidomimetics. Standard single or three letter codes are used to designate amino acids.

[0188] The term “biologically active polypeptide” refers to a polypeptide of any length that exhibits binding specificity to a given target or targets, which can be a therapeutic target and/or an accessory target, such as for cell-, tissue- or organ targeting. Alternatively, or in addition, it refers to a polypeptide that exhibits a desired biological characteristic when used in vitro or in vivo. By way of example, biologically active polypeptides include functional therapeutics or in vivo diagnostic proteins that bind to therapeutic or diagnostic targets. The term “biologically active polypeptide” and “Binding Module” or “BM” are used interchangeably herein. Biologically active polypeptides can be, for example, and
without limitation, linear or cyclic peptides, cysteine-constrained peptides, microproteins, scaffold proteins like fibronectin, ankryns, crystalline, streptavidin, antibody fragments, domain antibodies, peptides, hormones, growth factors, cytokines, or any type of protein domain, human or non-human, natural or non-natural, and they may be based on a natural scaffold or not based on a natural scaffold (i.e., engineered or selected), or based on combinations or fragments of any of the above. Optionally, the biologically active polypeptide can be engineered by adding, removing, or replacing one or multiple amino acids in order to enhance their binding properties, their stability, or other desired properties. Binding modules can be obtained from natural proteins, by design or by genetic package display, including phage display, cellular display, ribosomal display or other display methods, for example. Binding modules may bind to the same copy of the same target, which results in avidity, or they may bind to different copies of the same target (which can result in avidity if these copies are somehow connected or linked, such as by a cell membrane), or they may bind to two unrelated targets (which yields avidity if these targets are somehow linked, such as by a membrane). Binding modules can be identified by screening or otherwise analyzing random libraries of peptides or proteins.

[0189] "Recombinant PEG," "rPEG" or "rPEG polypeptides" or "recombinant PK Enhancing Group" are general terms encompassing a class of polypeptides that can be used to modify biologically active polypeptides, whereby the modification results in a desirable change in biological properties such as serum half-life or in vivo clearance. In general, rPEG polypeptides lack binding specificity to the same given target bound by the biologically active polypeptide. In some aspects, rPEG is a functional analog of PEG that may mimic some, but not necessarily all, well-known properties of PEG. Such properties, described in more detail below, include enhanced ability to increase hydraulic radius, increased resistance to proteases, decreased immunogenicity and decreased specific activity. While rPEG molecules may share broad structural and functional features with PEG, such as linearity or lack of tertiary structure, strict chemical similarity with PEG is not a necessary feature of rPEG.

[0190] "Accessory polypeptide" or "accessory protein" refers to a polypeptide which, when used in conjunction with a biologically active polypeptide, e.g., by way of linking to the biologically active polypeptide, renders a desirable change in biological properties of the entire linked polypeptide. Non-limiting examples of accessory polypeptides include rPEGs and any other polypeptides capable of increasing hydraulic radius, extending serum half-life, and/or modifying in vivo clearance rate. When desired, an accessory polypeptide causes a small increase in predicted molecular weight, but a much larger increase in apparent molecular weight. Although the different names emphasize different features, they refer to the same module and can be used interchangeably.

[0191] The terms "modified polypeptide" and "accessory-modified polypeptide" are used interchangeably to refer to biologically active polypeptides which have been modified with the accessory polypeptides of the invention. These terms may also refer to slow release or other types of formulations comprising biologically active polypeptides modified with accessory polypeptides according to the invention.

[0192] A "repetitive sequence" or "repetitive motif" are used interchangeably herein and refer to an amino acid sequence that can be described as an oligomer of repeating peptide sequences ("repeats"), forming direct repeats, or inverted repeats or alternating repeats of multiple sequence motifs. These repeating oligomer sequences can be identical or homologous to each other, but there can also be multiple repeated motifs. Repetitive sequences are characterized by a very low information content. A repetitive sequence is not a required feature of an accessory polypeptide and in some cases a non-repetitive sequence will in fact be preferred.

[0193] Amino acids can be characterized based on their hydrophobicity. A number of scales have been developed. An example is a scale developed by Levitt, M et al., published in J Mol Biol 104, 59, #3233, which is listed in Hopf, T P, et al. (1981) Proc Natl Acad Sci USA 78, 3824, #3232. Examples of "hydrophilic amino acids" are arginine, lysine, threonine, alanine, asparagine, and glutamine. Of particular interest are the hydrophilic amino acids aspartate, glutamate, and serine, and glycine. Examples of "hydrophobic amino acids" are tryptophan, tyrosine, phenylalanine, methionine, leucine, isoleucine, and valine.

[0194] As used herein, the term "cell surface proteins" refers to the plasma membrane components of a cell. It encompasses integral and peripheral membrane proteins, glycoproteins, polysaccharides and lipids that constitute the plasma membrane. An integral membrane protein is a transmembrane protein that extends across the lipid bilayer of the plasma membrane of a cell. A typical integral membrane protein consists of at least one membrane spanning segment that generally comprises hydrophobic amino acid residues. Peripheral membrane proteins do not extend into the hydrophobic interior of the lipid bilayer and they are bound to the membrane surface via covalent or noncovalent interaction directly or indirectly with other membrane components.

[0195] The terms "membrane", "cytosolic", "nuclear" and "secreted" as applied to cellular proteins specify the extracellular and/or subcellular location in which the cellular protein is mostly, predominantly, or preferentially localized.

[0196] "Cell surface receptors" represent a subset of membrane proteins, capable of binding to their respective ligands. Cell surface receptors are molecules anchored on or inserted into the cell plasma membrane. They constitute a large family of proteins, glycoproteins, polysaccharides and lipids, which serve not only as structural constituents of the plasma membrane, but also as regulatory elements governing a variety of biological functions.

[0197] "Non-naturally occurring" as applied to a protein means that the protein contains at least one amino acid that is different from the corresponding wildtype or native protein. Non-natural sequences can be determined by performing BLAST search using, e.g., the lowest smallest sum probability where the comparison window is the length of the sequence of interest (the queried) and when compared to the non-redundant ("nr") database of Genbank using BLAST 2.0. The BLAST 2.0 algorithm, which is described in Altschul et al. (1990) J. Mol. Biol. 215:403-410, respectively. Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information.

[0198] A "host cell" includes an individual cell or cell culture which can be or has been a recipient for the subject vectors. Host cells include progeny of a single host cell. The progeny may not necessarily be completely identical (in morphology or in genomic of total DNA complement) to the original parent cell due to natural, accidental, or deliberate mutation. A host cell includes cells transfected in vivo with a vector of this invention.
As used herein, the term “isolated” means separated from constituents, cellular and otherwise, with which the polynucleotide, peptide, polypeptide, protein, antibody, or fragments thereof, are normally associated with in nature. As is apparent to those of skill in the art, a non-naturally occurring polynucleotide, peptide, polypeptide, protein, antibody, or fragments thereof, does not require “isolation” to distinguish it from its naturally occurring counterpart. In addition, a “concentrated”, “separated” or “diluted” polynucleotide, peptide, polypeptide, protein, antibody, or fragments thereof, is distinguishable from its naturally occurring counterpart in that the concentration or number of molecules per volume is greater than “concentrated” or less than “separated” than that of its naturally occurring counterpart. In general, a polypeptide made by recombinant means and expressed in a host cell is considered to be “isolated”.

“Conjugated”, “linked” and “fused” or “fusion” are used interchangeably herein. These terms refer to the joining together of two or more chemical elements or components, by whatever means including chemical conjugation or recombinant means. An “in-frame fusion” refers to the joining of two or more open reading frames (ORFs) to form a continuous longer ORF, in a manner that maintains the correct reading frame of the original ORFs. Thus, the resulting recombinant fusion protein is a single protein containing two or more segments that correspond to polypeptides encoded by the original ORFs (which segments are not normally so joined in nature).

In the context of polypeptides, a “linear sequence” or a “sequence” is an order of amino acids in a polypeptide in an amino to carboxyl terminus direction in which residues that neighbor each other in the sequence are contiguous in the primary structure of the polypeptide. A “partial sequence” is a linear sequence of part of a polypeptide which is known to comprise additional residues in one or both directions.

“Heterologous” means derived from a genotypically distinct entity from the rest of the entity to which it is being compared. For example, a glycine rich sequence removed from its native coding sequence and operatively linked to a coding sequence other than the native sequence is a heterologous glycine rich sequence. The term “heterologous” as applied to a polynucleotide, a polypeptide, means that the polynucleotide or polypeptide is derived from a genotypically distinct entity from that of the rest of the entity to which it is being compared.

The terms “polynucleotides”, “nucleic acids”, “nucleotides” and “oligonucleotides” are used interchangeably. They refer to a polymorphic form of nucleotides of any length, either deoxyribonucleotides or ribonucleotides, or analogs thereof. Polynucleotides may have any three-dimensional structure, and may perform any function, known or unknown. The following are non-limiting examples of polynucleotides: coding or non-coding regions of a gene or gene fragment, loci (locus) defined from linkage analysis, exons, introns, messenger RNA (mRNA), transfer RNA, ribosomal RNA, ribozymes, cDNA, recombinant polynucleotides, branched polynucleotides, plasmids, vectors, isolated DNA of any sequence, isolated RNA of any sequence, nucleic acid probes, and primers. A polynucleotide may comprise modified nucleotides, such as methylated nucleotides and nucleotide analogs. If present, modifications to the nucleotide structure may be imparted before or after assembly of the polymer. The sequence of nucleotides may be interrupted by non-nucleotide components. A polynucleotide may be further modified after polymerization, such as by conjugation with a labeling component.

“Recombinant” as applied to a polynucleotide means that the polynucleotide is the product of various combinations of cloning, restriction and/or ligation steps, and other procedures that result in a construct that can potentially be expressed in a host cell.

The terms “gene” or “gene fragment” are used interchangeably herein. They refer to a polynucleotide containing at least one open reading frame that is capable of encoding a particular protein after being transcribed and translated. A gene or a gene fragment may be genomic or cDNA, as long as the polynucleotide contains at least one open reading frame, which may cover the entire coding region or a segment thereof. A “fusion gene” is a gene composed of at least two heterologous polynucleotides that are linked together.

A “vector” is a nucleic acid molecule, preferably self-replicating, which transfers an inserted nucleic acid molecule into and/or between host cells. The term includes vectors that function primarily for insertion of DNA or RNA into a cell, replication of vectors that function primarily for the replication of DNA or RNA, and expression vectors that function for transcription and/or translation of the DNA or RNA. Also included are vectors that provide more than one of the above functions. An “expression vector” is a polynucleotide which, when introduced into an appropriate host cell, can be transcribed and translated into a polypeptide(s). An “expression system” usually composes a suitable host cell comprised of an expression vector that can function to yield a desired expression product.

The “target” as used in the context of accessory polypeptides is a biochemical molecule or structure to which the biologically active polypeptide can bind and where the binding event results in a desired biological activity. The target can be a protein ligand or receptor that is inhibited, activated or otherwise acted upon by the t protein. Examples of targets are hormones, cytokines, antibodies or antibody fragments, cell surface receptors, kinases, growth factors and other biochemical structures with biological activity.

“Serum degradation resistance”—Proteins can be eliminated by degradation in the blood, which typically involves proteases in the serum or plasma. The serum degradation resistance is measured by combining the protein with human (or mouse, rat, monkey, as appropriate) serum or plasma, typically for a range of days (ie 0.25, 0.5, 1, 2, 4, 8, 16 days) at 37°C. The samples for these timepoints are then run on a Western assay and the protein is detected with an antibody. The antibody can be to a tag in the protein. If the protein shows a single band on the western, where the protein’s size is similar to that of the injected protein, then no degradation has occurred. The timepoint where 50% of the protein is degraded, as judged by Western Blots or equivalent techniques, is the serum degradation half-life or “serum half-life” of the protein.

“Apparent Molecular Weight Factor” or “Apparent Molecular Weight” are related terms referring to a measure of the relative increase or decrease in apparent molecular weight exhibited by a particular amino acid sequence. The Apparent Molecular Weight is determined using a size exclusion column that can be calibrated using globular protein standards and is measured in “apparent kDa” units. The Apparent Molecular Weight Factor is measured as the ratio between the apparent molecular weight, as determined on a size exclusion
column calibrated with globular proteins and the actual molecular weight, (i.e., predicted by adding based on amino acid composition the calculated molecular weight of each type of amino acid in the amino acid composition). For example, a 20 kDa poly-Glycine sequence has an apparent molecular weight of 200 kDa by size exclusion chromatography, corresponding to an Apparent Molecular Weight Factor of 10x. The ‘Specific Hydrodynamic Radius’ is the hydrodynamic radius per unit molecular weight (kDa), is a measure for the performance of a half-life extender, which is measured as the serum secretion half-life per unit mass (hours per kDa). Both of these measurements are correlated with the ‘Apparent Molecular Weight Factor’, which is a more intuitive measure.

[0210] The “hydrodynamic radius” of a protein affects its rate of diffusion in aqueous solution as well as its ability to migrate in gels of macromolecules. The hydrodynamic radius of a protein is determined by its molecular weight as well as by its structure, including shape and compactness. Most proteins have globular structures, which is the most compact three-dimensional structure a protein can have with the smallest hydrodynamic radius. Some proteins adopt a random and open, unstructured, or ‘linear’ conformation and as a result have a much larger hydrodynamic radius compared to typical globular proteins of similar molecular weight.

[0211] “Physiological conditions” refer to a set of conditions including temperature, salt concentration, pH that mimic those conditions of a living subject. A host of physiologically relevant conditions for use in in vitro assays have been established. Generally, a physiological buffer contains a physiological concentration of salt and is adjusted to a neutral pH ranging from about 6.5 to 7.8, and preferably from about 6.6 to 7.5. A variety of physiological buffers is listed in Sambrook et al. (1989) supra and hence is not detailed herein. Physiologically relevant temperature ranges from about 25°C to about 38°C, and preferably from about 30°C to about 37°C.

[0212] A “reactive group” is a chemical structure that can be coupled to a second reactive group. Examples for reactive groups are amino groups, carboxyl groups, sulfhydryl groups, hydroxyl groups, aldehyde groups, azide groups. Some reactive groups can be activated to facilitate coupling with a second reactive group. Examples for activation are the reaction of a carboxyl group with carbodiimide, the conversion of a carboxyl group into an activated ester, or the conversion of a carboxyl group into an azide function.

[0213] A “crosslinking component” includes a chemical structure that comprises one or more reactive groups.

[0214] These reactive groups can be identical in their chemical structure allowing the direct construction of crosslinked accessory polypeptides. Cross-linking components can contain reactive groups that have been blocked by protecting groups. This allows one to conjugate several different non-cross-linking components to one cross-linking component in controlled consecutive reactions. Cross-linking components can contain multiple reactive groups that differ in their structure and that can be selectively conjugated with different non-cross-linking components. Proteins that contain multiple high-affinity binding sites can also serve as cross-linking agents. Examples are streptavidin, which can bind up to four molecules of a biotinylated non-cross-linking component. Branched multifunctional polyethylene glycol (PEG) molecules can serve as cross-linking components. A variety of reagents with two to eight functional groups and various lengths of PEG as well as various reactive groups are commercially available. Suppliers include NOF America Corporation and SunBio.

[0215] “Non-crosslinking components” include chemical structures that comprise reactive groups which allow conjugation to a cross-linking component. Non-cross-linking components can contain a variety of modules, including one or more biologically active polypeptides and/or one or more accessory polypeptides. In addition, non-cross-linking components can contain affinity tags that facilitate purification and/or detection, such as Flag-tag, E-tag, Myc-tag, HA-tag, His6-tag (SEQ ID NO: 1), Green Fluorescent protein, etc.

[0216] A “crosslinked rPEG polypeptide”, “crosslinked accessory polypeptide”, “crosslinked rPEG”, “CL-rPEG polypeptide”, “CL-rPEG” are terms referring to conjugates of one or more non-cross-linking components with a crosslinking component.

[0217] “Controlled release agent”, “slow release agent”, “depot formulation” or “sustained release agent” are used interchangeably to refer to an agent capable of extending the duration of release of a modified polypeptide of the invention relative to the duration of release when the modified polypeptide is administered in the absence of agent. Different embodiments of the present invention may have different release rates, resulting in different therapeutic amounts.

[0218] “Vl domain” refers to the variable domain of the light chain of an antibody.

[0219] “VH domain” refers to the variable domain of the heavy chain of an antibody.

[0220] A “variable fragment” (Fv) refers to a portion of an antibody which comprises two non-covalently associated VL and VH domains.

[0221] A “single chain variable fragment” (scFv) refers to a portion of an antibody which comprises one VH linked via a non-natural peptide linker to one VL domain, as a single chain. scFvs can have the structure VH-linker-VL or VL-linker-VH where the linker can be any peptide sequence comprising various numbers of amino acids. A scFv preferentially occurs under physiological conditions as a monomeric structure which requires a peptide linker of preferably more than 12 amino acids.

[0222] Disulfide-stabilized Fv fragments of antibodies (dFv) refers to molecules in which the V_{H}-V_{L} heterodimer is stabilized by an interchain disulfide bond engineered between structurally conserved framework positions distant from complementarity-determining regions (CDRs). This method of stabilization is applicable for the stabilization of many antibody Fvs.

[0223] A “variable domain” refers to the domain that forms the antigen binding site of an antibody. Variable domains can be VH or VL; Differences, between the variable domains, are located on three loops known as hypervariable regions (HV-1, HV-2 and HV-3) or CDR1, CDR2 and CDR3. CDRs are supported within the variable domains by conserved framework regions.

[0224] A “domain antibody” (dAb) refers to a portion of an antibody that is capable of binding the target as a monomer. Domain antibodies correspond to the variable regions of either the heavy (VH) or light (VL) chains of antibodies. dAbs do not generally require a second variable domain (VH or VL) for target binding. dAbs can be generated by phage display or other in vitro methods. Alternatively, dAb domain can be obtained from immunized camels or sharks or other species that generate antibodies that lack a light chain.
A “dIabody” refers to a recombinant antibody that has two Fv heads, each consisting of a V\textsubscript{H} domain from one polypeptide paired with the V\textsubscript{L} domain from another polypeptide. A diabody typically contains two VaL-\textsubscript{L} (or VaL-\textsubscript{L}) chains. Diabody can be constructed by joining the V\textsubscript{H} and V\textsubscript{L} domains of an antibody by a peptide linker. The peptide linker lengths comprise various numbers of amino acids, preferably between 2 and 12 amino acids. A diabody can be monospecific or bispecific.

A “tIrabody” refers to a recombinant antibody that has three Fv heads, each consisting of a V\textsubscript{H} domain from one polypeptide paired with the V\textsubscript{L} domain from a neighboring polypeptide. A triabody contains three V\textsubscript{H}-VaL (or V\textsubscript{L}-H) chains. Triabody can be constructed by joining the V\textsubscript{H} and V\textsubscript{L} domains of an antibody by a peptide linker. The peptide linker lengths comprise various numbers of amino acids, preferably between 2 and 2 amino acids. A triabody can be monospecific or bispecific.

A “tetaIrabody” comprises four V\textsubscript{H}-VaL (or V\textsubscript{L}-H) chains. Tetraabodies can be constructed by joining the V\textsubscript{H} and V\textsubscript{L} domains of an antibody by a peptide linker. The peptide linker lengths comprise various numbers of amino acids, preferably between 2 and 2 amino acids. Tetraabodies can be obtained by truncating various numbers of amino acids, preferably between 1 to 10 amino acids, from the joined ends of the V\textsubscript{H} and V\textsubscript{L} domains.

A “Fab fragment” refers to a region on an antibody which binds to antigens. A Fab fragment is composed of one constant and one variable domain of each of the heavy and the light chain. These domains shape the paratope—the antigen binding site—at the amino terminal end of the monomer. The two variable domains bind the epitope on their specific antigens. A Fab fragment can be linked by a disulfide bond at the C-terminus. Fab fragments can be generated in vitro. The enzyme papain can be used to cleave an immunoglobulin monomer into two Fab fragments and an Fc fragment. The enzyme pepsin cleaves below the hinge region, so a Fab\textsubscript{p} fragment and an Fc fragment is formed. The variable regions of the heavy and light chains can be fused together to form a single chain variable fragment (scFv), which retains the original specificity of the parent immunoglobulin.

The term “antibody fragment” is used herein to include all of the fragments described in the present invention including any antigen binding unit as defined in details below, such as Ab, Fv, Fab, and Fc in any form. Antibody fragments can comprise additional domains of an antibody. An antibody fragment also encompasses a complete or full antibody.

The term “parent antibody” is used herein to refer to the antibody upon which the construction of an antibody fragment is based.

An “antibody fragment based therapeutic” (AFBT) refers to any therapeutic agent or pharmaceutical composition that is based on an antibody fragment as described herein. AFBTs can comprise multiple antibody fragments that can be derived from multiple different parent antibodies. Multiplespecific AFBTs may comprise multiple antibody fragments with specificity against multiple different epitopes. These epitopes can be part of the same target antigen or on multiple different target antigens. Bispecific AFBTs may comprise binding sites (generally two or more, but may be one) with two different binding specificities.

The terms “antigen”, “target antigen” or “immuno- gen” are used interchangeably herein to refer to the structure or binding determinant that an antibody fragment or an antibody fragment-based therapeutic binds to or has specificity against.

The terms “domain reassortment” and “domain swapping” are used interchangeably herein to refer to a process that changes the valency of an antibody fragment or an antibody fragment-based therapeutic. For example, single chain variable fragments (scFv) can reassort to form dimers, trimers etc., as well as diabodies, triabodies, tetaabodies, and the like. Fabcs can exchange whole chains with other Fabcs or even whole antibodies, potentially yielding mismatched chains that result in loss of one or both binding activities. The formation of light chain dimers, called Bence-Jones Protein, is another example. Another example of reassortment is heavy chain reassortment between IgG4 antibodies, which do not have a disulfide-bonded hinge that prevents such exchange, which can lead to bispecific IgG4 antibodies. The rate of domain reassortment is dependent on the reaction conditions such as salt concentration, pH, temperature, and the presence of target antigen.

The term “payload” as used herein refers to a protein or peptide sequence that has biological or therapeutic activity, equivalent to the pharmacophore of small molecules. Examples of payloads include, but are not limited to, cytokines, enzymes and growth factors. Payloads can comprise genetically fused or chemically conjugated moieties. Examples for such chemically conjugated moieties include, but are not limited to, chemotherapeutic agents, antiviral compounds, or contrast agents. These conjugated moieties can be joined to the rest of the AFBT via a linker which may be cleavable or non-cleavable.

“Collagen binding domain” (CBD) refers to a protein domain that binds to or has specificity against collagen. CBDs can be specific for any particular types of collagen such as collagen I. Alternatively, CBDs may bind to a variety of collagen types. An example is fibronectin in which four protein domains are sufficient for collagen binding.

The term “repetitiveness” used in the context of a polypeptide, for example, an accessory polypeptide PEG, refers to the degree of internal homology in a peptide sequence. A repetitive sequence may contain multiple identical or homologous copies of an amino acid sequence. Repetitiveness can be measured by analyzing the frequency of identical subsequences. For instance, a polypeptide sequence of interest may be divided into n-mer sub-sequences and the number of identical subsequences can be counted. Highly repetitive sequences contain a large fraction of identical subsequences.

“Total charge density” as used herein is calculated by adding the number of negatively charged amino acids with the number of positively charged amino acids, and dividing the sum by the total number of amino acids in a polypeptide. For example: hlgG1 Fc sequence: (MDKTHTPCP...)

Number of negatively charged residues: 22; Number of positively charged residues: 22; Total number of residues: 224; Total charge density of Fc alone: (22\times24)/224 = 0.20.
“Net charge density” as used herein is calculated by subtracting the number of positively charged amino acids from the number of negatively charged amino acids, and dividing the difference by the total number of amino acids in a polypeptide. For example: hlgG1 Fc sequence: (MDKTH-TCPPCPAPELLGPGSVTLFFPKK-KDTLMISRTEVIPCVVYDUSHEDPEVK-FNWYVDGVEV-HIASCKKPREQYNSTRVSVLTLYHQDNWINGKYCKVSNKALPAEKITISKAKQGPLRPQVYTLPSRDELTKQQVSTLTCVKGFYPSDIAVEWESENSQFPENNYKTTIPVLDSDSFFLYSKLTVDKS-RWQQGNFSCSVMHEALH-NHYTQKSLSL (SEQ ID NO: 10) Number of negatively charged residues: 24; Number of positively charged residues: 22; Total number of residues: 224; Net charge density of Fc alone: (24-22)/224=2-224=0.99%.

“Predicted solubility” as used herein is calculated by adding the net charge of folded protein to the total charge of an unstructured protein (e.g. rPEG), and dividing the sum by the total number of amino acids in the protein. For example, the predicted solubility of Fc-rPEG50 is (-2+192)/224+576)=190/800=23.75%.

Design of Accessory Polypeptides for Improving Solubility During Expression of Biologically Active Polypeptides.

Expression of soluble modified biologically active polypeptides may be optimized by modifying the net charge density of the modified polypeptide. In some cases, the net charge density is above +0.1 or below 0.1 charges/residue. In other cases, the charge density is above +0.2 or below -0.2 charges per residue. Charge density may be controlled by modifying the content of charged amino acids such as arginine, lysine, glutamic acid and aspartic acid within accessory polypeptides linked to the biologically active polypeptide. If desired, the accessory polypeptide may be composed exclusively of a short stretch of charged residues. Alternatively, the accessory polypeptide may comprise charged residues separated by other residues such as serine or glycine, which may lead to better expression or purification behavior. Higher expression may be obtained. Use of serine may lead to higher expression levels.

The net charge that is required for the accessory protein to make a fusion protein soluble and fold in the cytoplasm depends on the biologically active polypeptide, specifically its size and net charge. The net charge of the modified polypeptide may be positive or negative. In some applications, accessory polypeptide sequences rich in negative amino acids such as glutamic acid or aspartic acid may be desirable. In other applications, accessory polypeptide sequences rich in positive amino acids such as lysine or arginine may be preferred. The use of both positively and negatively charged amino acids may lead to charge neutralization, which could potentially neutralize the advantage of the invention. For example, accessory polypeptide sequences rich in positive amino acids with 16%, 25% or 33% negatively charged residues may provide up to 96 total charges, which is sufficient to achieve a charge density of 0.1 for a neutral fusion protein of up to 960 amino acids, or a non-fusion protein of 672 amino acids. In one specific example, an accessory polypeptide comprising 33% glutamic acid residues might be used to make even larger and difficult to express proteins soluble.

To impart solubility on the binding protein, the net positive or negative charge of the accessory polypeptide may be greater than 5, 10, 15 or 20 or even greater than 30, 40, 50, 60, 70, 80, 90 or 100. Charges can be concentrated in a short sequence of 5, 10, 15, 20, 25, 30, 40, 50 amino acids, or can be spaced out over a longer sequence of 60, 80, 100, 150, 200, 250, 300, 400, or 500 or more amino acids. The sequence of a negative accessory polypeptide may contain over 5, 10, 15, 25, 30, 40, 50, 60, 70, 80, 90 or 100 percent of glutamic or aspartic acid, while a positive accessory polypeptide may contain over 5, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90 or 100 percent of arginine or lysine. Non-charged residues may be used as such as the relatively hydrophilic residues Serine and Glocine.

Additional Considerations in the Design of Accessory Polypeptides

One aspect of the present invention is the design of accessory polypeptides, e.g., rPEG accessory polypeptides and the like for the modification of biologically active polypeptides (FIG. 1). The accessory polypeptides are particularly useful in generating recombinant proteins of therapeutic and/or diagnostic value.

A variety of accessory polypeptide sequences can be designed and these may be rich in glycine and/or serine, as well as other amino acids such as glutamate, aspartate, alanine or proline. Accessory polypeptide sequences may be rich in hydrophilic amino acids and contain a low percentage of hydrophobic or aromatic amino acids. Accessory polypeptide sequences can be designed to have at least 30, 40, 50, 60, 70, 80, 90 or 100% glycine and/or serine residues. In some cases, accessory polypeptide sequences contain at least 50, 55, 60, 65% glycine and/or serine. In other cases, accessory polypeptide sequences may contain at least 70, 75, 80, 85, 90% glycine and/or serine residues.

The compositions of the present invention will typically contain accessory polypeptide sequences consisting of a total of at least 40 amino acids. However, the products can contain multiple accessory polypeptide sequences and some or all of these individual accessory polypeptide sequences may be shorter than 40 amino acids as long as the combined length of all accessory polypeptide sequences of a product is at least 40 amino acids. In some embodiments, the combined length of accessory polypeptide sequences that are attached to a protein can be 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 120, 140, 160, 180, 200, 250, 300, 350, 400, 500, 600, 700, 800, 900 or more than 1000 or 2000 amino acids. In some modified biologically active polypeptides the combined length of accessory polypeptide sequences exceeds 60, 70, 80, 90 or more amino acids. In other modified polypeptides the combined length of accessory polypeptide sequences exceeds 100, 120, 140, 160 or 180 amino acids, and even 200, 250, 300, 350, 400, 500, 600, 700, 800 or even more than 1000 amino acids.

One or several accessory polypeptide sequences can be fused to a biologically active polypeptide, for example to the N- or C-terminus of the biologically active polypeptide or inserted into loops of a polypeptide of interest to give the resulting modified polypeptide improved properties relative to the unmodified polypeptide. Fusion of accessory sequences (to a (therapeutic) protein leads to a significant increase in the hydrodynamic radius of the resulting fusion protein relative to the unmodified protein, which can be detected by ultrafiltration, size exclusion chromatography, or light scattering, for example.

Accessory polypeptide sequences can be designed to avoid one or more types of amino acids to yield a desired
property. For instance, one can design accessory polypeptide sequences to contain few or none of the following amino acids: cysteine (to avoid disulfide formation and oxidation), methionine (to avoid oxidation), asparagine and glutamine (to avoid deamination) and aspartate. Accessory polypeptide sequences can be designed to contain proline residues that tend to reduce sensitivity to proteolytic degradation.

[0248] Accessory polypeptide sequences can be designed such as to optimize protein production. This can be achieved by avoiding or minimizing repetitiveness of the encoding DNA. Accessory polypeptide sequences such as poly-glycine or poly-serine may have very desirable pharmaceutical properties but their manufacturing can be difficult due to the high GC-content of DNA sequences encoding for poly-glycine and due to the presence of repeating DNA sequences that can lead to recombination.

[0249] Accessory polypeptides, including simple sequences composed of short, repeated motifs rich in sequences rich in G, S and E, may cause relatively high antibody titers of >1000 in multiple species despite the absence of T-cell epitopes in these sequences. This may be caused be the repetitive nature of the accessory polypeptides, as it is has been shown that immunogens with repetitive epitopes, including protein aggregates, cross-linked immunogens, and repetitive carbohydrates are highly immunogenic. (Johansson, J., et al. (2007) Vaccine, 25: 1676-82; Yankai, Z., et al. (2006) Biochem Biophys Res Commun, 345: 1365-71, Hsu, C. T., et al. (2000) Cancer Res, 60: 3701-5). B-cells displaying pentavalent IgM molecules are stimulated by repetitive immunogens even if the monovalent binding affinity of an immunogen for the IgM is very low, such as at micromolar concentrations (Fig. 74). Simultaneous binding of linked repeats to multiple coupled IgM domains located on the same molecule or on the same cell may cause a large (thousand, million or perhaps even billion-fold) increase in the apparent (effective) affinity of the interaction, which may stimulate B-cells. To avoid this type of effect, accessory polypeptides may be screened for immunogenicity (as well as for effects on half-life and other properties) in multiple species of animals (such as rats, rabbits, mice, or guinea pigs. Multiple injections may be performed, with pharmacokinetic properties being measured in the same animals before and after immunization). In addition, accessory polypeptide sequences may be designed to be non-repetitive (comprising only 1 identical copy of each sequence motif) or to have a minimal number of copies of each sequence motif. Accessory polypeptide sequences that are less-repetitive may comprise binding sites for different IgMs, but they may be less able to bind multivalently to the same IgM molecule or to the same B-cell, since each B-cell generally secretes only one type of IgM and each IgM typically only has one type of binding site. This mechanism is illustrated in FIGS. 74a and b. In some embodiments, accessory polypeptides may contain exclusively sequences that occur at 1, 2, 3, 4, 5 or 6 copies per accessory polypeptide. Polypeptides with a lower number of repeats, may have a lower expected avidity may be less likely to induce a substantial immune response. Such sequences may comprise multiple types of amino acids, such as two types (for example, G and E or S and E), three types of amino acids (like G, E and S) or even four or more. Such accessory polypeptide may also comprise, for example, 30-80% glycine, 10-40% serine and 15-50% glutamate of the total amino acid composition. Such sequences may provide an optimal balance of desired properties such as expression level, serum and E. coli protease resistance, solubility, aggregation, and immunogenicity.

[0250] FIG. 74 compares the interactions of a repetitive (74a) and a non repetitive accessory polypeptide sequence (74b) with B cells that recognize epitopes in said sequences. A repetitive sequence will be recognized by few B cells in an organism as it contains a relatively small number of different epitopes. However, a repetitive sequence can form multivalent contacts with these few B cells and as a consequence it can stimulate their proliferation as illustrated in FIG. 74a. A non repetitive sequence can make contacts with many different B cells as it contains many different epitopes. However, each individual B cell can only make one or a small number of contacts with an individual non-repetitive accessory polypeptide (“nrURP”) due to the lack of repetitiveness as illustrated in FIG. 74b. As a result, non-repetitive accessory polypeptides may have a much lower tendency to stimulate proliferation of B cells and thus an immune response.

[0251] An additional advantage of non-repetitive accessory polypeptides relative to repetitive accessory polypeptides is that non-repetitive accessory polypeptides form weaker contacts with antibodies relative to repetitive accessory polypeptides. Antibodies are multivalent molecules. For instance, IgGs have two identical binding sites and IgMs contain 10 identical binding sites. Thus antibodies against repetitive sequences can form multivalent contacts with such repetitive sequences with high avidity, which can affect the potency and/or elimination of such repetitive sequences. In contrast, antibodies against non-repetitive accessory polypeptides tend to form mostly monovalent interactions with antibodies as said non-repetitive accessory polypeptides contain few repeats of each epitope.

[0252] Repetitiveness describes the degree of internal homology in a peptide sequence. In the extreme case a repetitive sequence can contain multiple identical copies of an amino acid sequence. Repetitiveness can be measured by analyzing the frequency of identical subsequences. For instance one can divide a sequence of interest into n-mer subsequences and count the number of identical or homologous subsequences. Highly repetitive sequences will contain a large fraction of identical or homologous subsequences.

[0253] The repetitiveness of a gene can be measured by computer algorithms. An example is illustrated in FIG. 75. Based on the query sequence on can perform a pair wise comparison of all subsequences of a particular length. These subsequences can be compared for identity or homology. The example in FIG. 75 compares subsequences of 4 amino acids for identity. In the example, most 4-mer subsequences occur just once in the query sequence and 3 4mer subsequences occur twice. One can average the repetitiveness in a gene. The length of the subsequences can be adjusted. Where desired, the length of the subsequences can reflect the length of sequence epitopes that can be recognized by the immune system. Thus analysis of subsequences of 4-15 amino acids can be performed. Genes encoding non-repetitive accessory polypeptides can be assembled from oligonucleotides using standard techniques of gene synthesis. The gene design can be performed using algorithms that optimize codon usage and amino acid composition. In addition, one can avoid amino acid sequences that are protease sensitive or that are known to be epitopes that can easily recognized by the human immune system. Computer algorithms can be applied during sequence design to minimize the repetitiveness of the resulting
amino acid sequences. One can evaluate the repetitiveness of large numbers of gene designs that match preset criteria such as amino acid composition, codon usage, avoidance of protease sensitive subsequence, avoidance of epitopes, and chose the least repetitive sequences for synthesis and subsequent evaluation.

[0254] An alternative approach to the design of non-repetitive accessory polypeptide genes is to analyze the sequences of existing collections of non-repetitive accessory polypeptides that show high level expression, low aggregation tendency, high solubility, and good resistance to proteases. A computer algorithm can design non-repetitive accessory polypeptide sequences based on such pre-existing non-repetitive accessory polypeptide sequences by re-assembly of sequence fragments. The algorithm generates a collection of subsequences from these non-repetitive accessory polypeptide sequences and evaluates multiple ways to assemble non-repetitive accessory polypeptide sequences from such subsequences. These assembled sequences can be evaluated for repetitiveness to identify a non-repetitive accessory polypeptide sequence that is only composed of subsequences of previously identified non-repetitive accessory polypeptides.

[0255] Non-repetitive accessory polypeptide-encoding genes can be assembled from libraries of short accessory polypeptide segments as illustrated in FIG. 77. One can first generate large libraries of accessory polypeptide segments. Such libraries can be assembled from partially randomized oligonucleotides. The randomization scheme can be optimized to control amino acid choices for each position as well as codon usage. One may clone the library of accessory polypeptide segments into an expression vector. Alternatively, one may clone the library of accessory polypeptide segments into an expression vector fused to an indicator gene like GFP. Subsequently, one can screen library members for a number of properties such as level of expression, protease stability, binding to antiserum. One can determine the amino acid sequence of the library members to identify segments that have a particularly desirable amino acid composition, segment length, or to identify segments that have a low frequency of internal repeats. Subsequently, one can assemble non-repetitive accessory polypeptide sequences from collections of accessory polypeptide segments by random dimerization or multimerization. Dimerization or multimerization can be achieved by ligation or PCR assembly. This process results in a library non-repetitive accessory polypeptide sequences that can be evaluated for a number of properties to identify the non-repetitive accessory polypeptide sequences with the best properties. One can repeat the process of dimerization or multimerization to further increase the length of non-repetitive accessory polypeptide sequences.

[0256] In a specific embodiment, an accessory polypeptide comprises a mixture of the following 8 amino acid motifs: GESEGSGESE (SEQ ID NO: 11), GEGGSSESRGE (SEQ ID NO: 12), GEGSGGSE (SEQ ID NO: 13), GEGSGGESG (SEQ ID NO: 14), GEGSSEGGE (SEQ ID NO: 15), GEGSSSEGGE (SEQ ID NO: 16), GEGSSEGGE (SEQ ID NO: 17), GEGSEGSEGGE (SEQ ID NO: 18), GEGSGGGE (SEQ ID NO: 19), or GEGGSGG (SEQ ID NO: 20). This design has an average of 33% E and 11-22% Serine content, depending on the ratio of the numbers of motifs relative to each other. In another specific embodiment, an accessory polypeptide comprises a mixture of the following 12 amino acid motifs: GXEGGSEGGE (SEQ ID NO: 21), GEXEGGSEGGE (SEQ ID NO: 22), GEXGGSEGGE (SEQ ID NO: 23), GEXGGSEGGE (SEQ ID NO: 24), GSSGGSEGGE (SEQ ID NO: 25), GSSGGSEGGE (SEQ ID NO: 26), GSSGGSEGGE (SEQ ID NO: 27) or GSSGGSEGGE (SEQ ID NO: 28), where X represents either S or E with equal likelihood. This design has an average of 25% E and around 1% S, depending on the specific ratios chosen. Suitable specific ratios may be 1:1:1:1:1:1:1:1 or any other ratio, and may be to fine-tune the composition.

[0257] Accessory polypeptide sequences can be designed to be highly repetitive, less repetitive or non-repetitive at the amino acid level. For example, highly repetitive accessory polypeptide sequences may contain only a small number of overlapping 9-mer peptide sequences and in this way the risk of eliciting an immune reaction can be reduced.

[0258] Examples of single-aminoc acid-type accessory polypeptide sequences are: poly-glycine, poly-glutamic acid, poly-aspartic acid, poly-serine, poly-threonine, wherein the length is at least 20 residues. Examples of accessory polypeptides with two types of amino acids are (GX)n (SEQ ID NO: 29), (SX)n (SEQ ID NO: 30), where G is glycine and S is serine, and X is aspartic acid, glutamic acid, threonine, or proline and n is at least 10. Another example is (GEX)n (SEQ ID NO: 31) or (SSX)n (SEQ ID NO: 32), where X is aspartic acid, glutamic acid, threonine, or proline and n is at least 7. Another example is (GGX)n (SEQ ID NO: 33) or (SSX)n (SEQ ID NO: 34), where X is aspartic acid, glutamic acid, threonine, or proline and n is at least 5. Another example is (GGG)X (SEQ ID NO: 35) or (SSS)X (SEQ ID NO: 36), where X is aspartic acid, glutamic acid, threonine, or proline and n is at least 4. Other examples are (GZ)n (SEQ ID NO: 37) and (Sz)n (SEQ ID NO: 38) and where X is aspartic acid, glutamic acid, threonine, or proline, n is at least 10, and z is between 1 and 20.

[0259] The number of these repeats can be any number between 5 and 300 or more. Products of the invention may contain accessory polypeptide sequences that are semi-random sequences. Examples are semi-random sequences containing at least 30, 40, 50, 60 or 70% glycine in which the glycines are well dispersed and in which the total concentration of tryptophan, phenylalanine, tyrosine, valine, leucine, and isoleucine is less than 70, 60, 50, 40, 30, 20, or 10% when combined. A preferred semi-random accessory polypeptide sequence contains at least 40% glycine and the total concentration of tryptophan, phenylalanine, tyrosine, valine, leucine, and isoleucine is less than 10% and a more preferred semi-random accessory polypeptide sequence contains at least 50% glycine and the total concentration of tryptophan, phenylalanine, tyrosine, valine, leucine, and isoleucine less than 5%. Accessory polypeptide sequences can be designed by combining the sequences of two or more shorter accessory polypeptide sequences or fragments of accessory polypeptide sequences. Such a combination allows one to better modulate the pharmaceutical properties of the product containing the accessory polypeptide sequences and it allows one to reduce the repetitiveness of the DNA sequences encoding the accessory polypeptide sequences, which can improve expression and reduce recombination of the accessory polypeptide sequences encoding sequences.

[0260] Where high level of solubility is desired, a high fraction of charged residues, preferably >25% glutamate (E) with the rest being mostly glycine or serine may be employed. High-level expression favors 10-50% serine (E), since serine has 6 codons which generally yields a much higher expression level than glycine (4 codons). There is generally a trade-
off in solubility and rapid clearance when utilizing high glutamate content in a sequence. Where desired, a glutamate content of less than 50%, preferably less than 30%, is used to provide desired solubility and to avoid rapid clearance in animals.

Non-Glycine Residues can be Selected to Optimize Properties

[0261] Of particular interest are accessory polypeptide sequences that are rich in glycine and/or serine. The sequences of non-gly, non-ser residues in these gly-rich or ser-rich sequences can be selected to optimize the properties of the protein. For instance, one can optimize the sequences of accessory polypeptides to enhance the selectivity of the biologically active polypeptide for a particular tissue. Such tissue-selective accessory polypeptide sequences can be obtained by generating libraries of random or semi-random accessory polypeptide sequences, injecting them into animals or patients, and determining sequences with the desired tissue selectivity in tissue samples. Sequence determination can be performed by mass spectrometry. Using similar methods one can select accessory polypeptide sequences that facilitate oral, buccal, intestinal, nasal, thecal, peritoneal, pulmonary, rectal, or dermal uptake. Of particular interest are accessory polypeptide sequences that contain regions that are relatively rich in the positively charged amino acids arginine or lysine which favor cellular uptake or transport through membranes; such accessory polypeptides may be useful for intracellular delivery of proteins.

[0262] As described in more detail below, accessory polypeptide sequences can be designed to contain one or several protease-sensitive sequences. Such accessory polypeptide sequences can be cleaved once the product of the invention has reached its target location. This cleavage may trigger an increase in potency of the pharmacologically active domain (pro-drug activation) or it may enhance binding of the cleavage product to a receptor. This is currently not possible for antibodies. However, in the case of PEGylated or accessory protein modified biologically active polypeptides, it is possible to provide a cleavage site for a foreign protease such as Tomato Etch Virus Protease or a similar site-specific, non-human protease. If the protease site is between the accessory protein and the therapeutic protein, or close to the therapeutic protein, then the injection of the protease will remove the accessory protein tail from the drug resulting in a shorter half-life and removal from the patient’s system. The concentration of the drug in the serum will drop 10-100-fold, effectively terminating treatment. This would be desirable, for example, if treatment needs to be stopped suddenly, such as due to an infection during treatment with a TNF-inhibitory microprotein (such as TNFα-Receptor-PcEG). An example would be to add a protease to the treatment regime that cleaves off the accessory protein, thereby sharply reducing the half-life of the active, TNF-inhibitory part of the protein which is then rapidly cleared. This approach would allow the infection to be controlled.

[0263] Accessory polypeptide sequences can also be designed to carry excess negative charges by introducing aspartic acid or glutamic acid residues. Of particular interest are accessory polypeptide that contain 8, 10, 15, 20, 25, 30, 40 or even 50% glutamic acid and less than 2% lysine or arginine. Such accessory polypeptides carry a high net negative charge and as a result they have a tendency to adopt open conformations due to electrostatic repulsion between individual negative charges of the peptide. Such a net negative charge leads to an effective increase in their hydrodynamic radius and as a result it can lead to reduced kidney clearance of such molecules. Thus, one can modulate the effective net charge and hydrodynamic radius of an accessory polypeptide sequence by controlling the frequency and distribution of negatively charged amino acids in the accessory polypeptide sequences. Most tissues and surfaces in a human or animal have a net negative charge. By designing accessory polypeptide sequences to have a net negative charge one can minimize non-specific interactions between the accessory polypeptide-therapeutic protein and various surfaces such as blood vessels, healthy tissues, or various receptors.

[0264] Other accessory polypeptides useful in the present invention exhibit one or more following features.

[0265] The accessory polypeptide can be characterized by enhanced hydrodynamic radius, wherein the accessory polypeptide increases the Apparent Molecular Weight Factor of the biologically active polypeptide to which it is linked. Because the Apparent Molecular Weight Factor is a predictor of serum secretion half-life (assuming the predicted molecular weight is constant), accessory polypeptides with higher Apparent Molecular Weight Factor are expected to show longer serum half-lives. In some embodiments, Apparent Molecular Weight Factors for accessory polypeptides can be greater than 3, 5, 7 or even 9. The Apparent Molecular Weight Factor can be measured by a variety of methods including but not limited to ultrafiltration through membranes with controlled pore sizes, or by size exclusion gel filtration (SEC). The Apparent Molecular Weight Factor can be affected by the concentration of salts and other solutes. It should generally be measured under conditions that are similar to physiological conditions, such as in blood or saline.

[0266] The accessory polypeptide can also be characterized by the effect wherein upon its incorporation into a biologically active polypeptide, the biologically active polypeptide exhibits a longer serum half-life as compared to the corresponding protein that lacks the accessory polypeptide. Methods of ascertaining serum half-life are known in the art (see e.g., Alvarez, P., et al. (2004) J Biol Chem, 279: 3375-81). One can readily determine whether the resulting protein has a longer serum half-life as compared to the unmodified protein by practicing any methods available in the art or exemplified herein.

[0267] The accessory polypeptide can also increase the solubility of the protein to which it is attached. For example, whereas human Interferon-alpha, human Growth Hormone and human G-CSF typically form inclusion bodies when expressed in the cytoplasm of E. coli, attachment of an accessory polypeptide (such as (SSGSSE)₃ (SEQ ID NO: 39) or (SSESSSSSSESSSE)₂ (SEQ ID NO: 40), (GEGGGE)₅ (SEQ ID NO: 41), or others) increases the solubility of the expressed polypeptide such that it no longer forms inclusion bodies but remains soluble in the cytoplasm from where it can be easily purified in active form and at high expression levels and efficiency, avoiding the need for refolding from inclusion bodies.

[0268] Accessory polypeptides can have a high degree of conformational flexibility under physiological conditions and they tend to have large hydrodynamic radii (Stokes' radius) compared to globular proteins of similar molecular weight, leading to a large 'specific volume' (volume per unit mass). Thus, the accessory polypeptide can behave like denatured peptide sequences lacking well defined secondary and
tertiary structures under physiological conditions. Denatured conformation describes the state of a peptide in solution that is characterized by a large conformational freedom of the peptide backbone. Most peptides and proteins adopt a denatured conformation in the presence of high concentrations of denaturants or at elevated temperature. Peptides in denatured conformation have characteristic CD spectra and they are characterized by a lack of long range interactions as determined by NMR. "Denatured conformation" and "unfolded conformation" are used synonymously herein. A variety of methods have been established in the art to discern the presence or absence of secondary and tertiary structures of a given polypeptide. For example, the secondary structure of a polypeptide can be determined by CD spectroscopy in the "far-UV" spectral region (190-250 nm). Secondary structure elements, such as alpha-helix, beta-sheet, and random coil structures each give rise to a characteristic shape and magnitude of CD spectra. Secondary structure can also be ascertained via certain computer programs or algorithms such as the Chou-Fasman algorithm (Chou, P. Y., et al. (1974) Biochemistry, 13: 222-45). For a given accessory sequence, the algorithm can predict whether there exists some or no secondary structure at all. In many cases, accessory sequences will have spectra that resemble denatured sequences due to their low degree of secondary and tertiary structure. In other cases, accessory sequences can adopt secondary structure, especially helices such as alpha-helices, or sheets such as beta-sheets. While unstructured amino acid polymers are generally preferred for the present invention, it is possible to use amino acid sequences that adopt some secondary structure, especially alpha-helices and to a lesser extent beta-sheets. Tertiary structure is generally undesirable due to its low specific hydrodynamic radius. Sequences with secondary structure are likely to have a lower hydrodynamic radius than sequences with less secondary structure, but they may still be useful. If the accessory sequence adopts tertiary structure (such as in protein domains), the hydrodynamic radius is expected to be even smaller. Whereas polyglycine has the highest ratio of hydrodynamic radius to mass (glycine is only 70D), globular proteins have the smallest ratio of hydrodynamic radius to mass. An exception is the inclusion in the accessory polypeptide of peptides with 0, 1, 2, 3 or 4 disulfides and varying degrees of secondary and tertiary structure) that bind to serum-exposed targets and increase the serum secretion halflife by a different mechanism.

The accessory polypeptides can be sequences with low immunogenicity. Low immunogenicity can be a direct result of the conformational flexibility of accessory sequences. Many antibodies recognize so-called conformational epitopes in protein antigens. Conformational epitopes are formed by regions of the protein surface that are composed of multiple discontinuous amino acid sequences of the protein antigen. The precise folding of the protein brings these sequences into a well-defined special configuration that can be recognized by antibodies. Preferred accessory polypeptides are designed to avoid formation of conformational epitopes. For example, of particular interest are accessory sequences having a low tendency to adopt compactly folded conformations in aqueous solution. In particular, low immunogenicity can be achieved by choosing sequences that resist antigen processing in antigen presenting cells, choosing sequences that do not bind MHC well and/or by choosing sequences that are derived from human sequences. Accessory polypeptide sequences can also reduce the immunogenicity of the biologically active polypeptide.

The accessory polypeptides can be sequences with a high degree of protease resistance. Protease resistance can also be a result of the conformational flexibility of accessory sequences, e.g., due to their high entropy. Protease resistance can be designed by avoiding known protease recognition sites for both endo- and exo-proteases, and by including a high glycine content. Alternatively, protease resistant sequences can be selected by phage display or related techniques from random or semi-random sequence libraries. Where desired for special applications, such as slow release from a depot protein, serum protease cleavage sites can be built into an accessory polypeptide. In such cases, the compositions of the present invention may dissolve or degrade (or may be intended to dissolve or degrade) during use. In general, degradation attributable to biodegradability involves the degradation of a polymer into its constituents (including, without limitation, the modified polypeptides and resulting degradation products). The degradation rate of a polymer often depends in part on a variety of factors, including the identity of any constituents that form the polymer (such as a protease sensitive site), the ratio of any substituents, and how the composition is formed or treated (e.g. whether substituents are protected). Of interest, however, are also accessory sequences with high stability (e.g., long serum half-life, less prone to cleavage by proteases present in bodily fluid) in blood or in the bodily tissue that is relevant for the application. Accessory polypeptides can also improve the protease resistance of a protein as they shield it from protease attack. An example of a natural unstructured, repetitive sequence composed of 3 amino acids is the linker in the pIII protein of M13 phage, which has the repeat (GGGS)n (SEQ ID NO: 42) and is known to be exceptionally stable to a vast array of proteases. An accessory protein with the motif (GGGS)n (SEQ ID NO: 42) is predicted to be very useful. For long sequences, one may prefer (GGG)n (SEQ ID NO: 43) to achieve higher solubility which may be needed at the increased length.

Accessory polypeptides with good solubility in water, blood and other bodily fluids under physiological conditions are also desirable to facilitate bioavailability. Such sequences can be obtained by designing sequences that are rich in hydrophilic amino acids such as glycine, serine, aspartate, glutamate, lysine, arginine, threonine and that contain few hydrophobic amino acids such as tryptophan, phenylalanine, tyrosine, leucine, isoleucine, valine, methionine. As a result of their amino acid composition, accessory polypeptides have a low tendency to form aggregates in aqueous formulations and the fusion of an accessory polypeptide to other proteins or peptides tends to enhance their solubility and reduce their tendency to form aggregates, which is a separate mechanism to reduce immunogenicity.

The accessory polypeptide can, in some cases, display enhanced non-specific binding to tissues or serum proteins (FIG. 28), which can function to prolong their serum half-life. Serum protein binding can be measured using a variety of methods. Examples for binding assays are ELISA, Bicore, Kinexa, or Forte Bio. Since most animal tissue surfaces have a (net) weak negative charge, proteins with a net negative charge show less non-specific tissue binding than proteins with a net positive charge. Creating a net weak negative charge by the addition of negative charges or by the
deletion of positive charges can make a protein bind more specifically or at least reduce non-specific binding.

[0273] However, if the net negative charge (or the net charge density) is too high, it can result in non-specific binding to surfaces with local patches of positive charge, such as parts or proteins that bind to extracellular matrix, or to DNA or RNA (e.g., VEGF, histones). In contrast, creating a protein with net positive charge by the addition of positive charges (such as K, R) or by the deletion of negative charges can make a protein bind non-specifically to tissues, which results in an extension of half-life.

[0274] The charge type and density of the accessory polypeptide itself can be modified. The negatively charged amino acids are E, D, (C) and the positively charged amino acids are R, K, (H). Changes generally involve exchanging one negatively charged residue for another, such as E for D or vice versa. In some instances, E is preferred, because D can isomerize leading to chemical instability that is undesirable for manufacturing. Changes in charge type, from positive charge to negative charge or vice versa, involve replacing K or R with E or D (positive replaced by a negative). Changes in charge also include replacing a non- or weakly charged amino acid (A, C, F, G, H, I, L, M, N, P, Q, S, T, V, W, Y) with a charged amino acid (E, D, K, R) or vice versa. “Charge density” is the number of charged amino acids as a percentage of total residues. Changing the charge density involves increasing or reducing the number of negatively charged amino acids (specifically E, D) or positively charged amino acids (K, R) as a percentage of total amino acids. In contrast, the ‘net charge density’ is the sum of all positively charged amino acids minus the sum of all negatively charged amino acids (“net charge”) as a percentage of the total number of residues.

[0275] The “net charge” and the “net charge density (net charge per AA)” can influence the solubility of the accessory polypeptide and of the accessory-modified polypeptide, as well as its ability to bind to other molecules. The accessory polypeptide can modify the charge type and density of fusion proteins, which can enhance serum half-life and can be exploited to enhance desirable interactions or to reduce non-desirable interactions of the fusion protein with other proteins or materials.

[0276] The accessory polypeptide can, in some cases, display enhanced non-specific binding to tissues or serum proteins, which can function to prolong their serum half-life. This can be measured as an extension of serum half-life compared to an accessory sequence that does not show non-specific binding, or it can be shown by ELISA as a weak binding affinity for proteins a high density of the opposite charge.

[0277] Accessory polypeptides can consist partially or entirely of a single amino acid, such as (E)n, (G)n or (S)n (also referred to as poly-E, poly-G, GGGGGG (SEQ ID NO: 44), poly-S, SSSSSS (SEQ ID NO: 46)), or even a homo-polymer of one of A, C, D, F, H, I, K, L, M, N, P, Q, R, S, T, V, W, Y; ie AAAAAA (SEQ ID NO: 47). The best single amino acid motifs (E, G, S) are immunologically the least complex (only one type of 9 amino acid peptide can be created), but each has some drawbacks. Glycine is weakly hydrophobic and poly-G allows limited solubility. An advantage of glycine is its high entropy. In some instances, serine may be preferred over glycine because the corresponding DNA sequence is likely to have a more balanced GC-ratio and generally provides a higher expression level, likely due to its 6 codons. The four charged amino acids, including Glutamic acid (E), have the highest solubility of the 20 natural amino acids, followed by Glycine and Serine. However, at a high negative net charge density the proteins start binding non-specifically to positively charged proteins and surfaces, such as VEGF (basic exons that bind ECM), histones, DNA/RNA-binding proteins and also to bone. Others have reported that a long string of poly-E causes a reduced half-life, instead of the desired extended half-life.

[0278] Serine and poly-Serine offer high solubility without a risk of aggregation and with the best codon use and expression level. The six codons for serine offer a balanced GC content, but more importantly, they allow poly-S or S-rich sequences to be encoded by exceptionally diverse DNA sequences that offer a greater degree of codon usage optimization and expression level optimization than other amino acids such as poly-E or poly-G (FIGS. 14 and 15).

[0279] The accessory polypeptides can be of any length necessary to effect the functional changes described above. The length of an accessory sequence that only contains 1, 2, 3 or more types of amino acids can have a lower limit of 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90 or 100 amino acids and an upper limit of 30, 40, 50, 60, 70, 80, 90, 100, 125, 150, 175, 200, 250, 300, 350, 400, 500, 600 or even 1000 amino acids.

[0280] The amino acid composition of the accessory polypeptide can be chosen such that the desirable properties of the resulting polypeptide are maximized. For example, for the extension of serum half-life a high ratio of apparent molecular weight to predicted molecular weight is preferred. The unstructured accessory polypeptides that offer more hydrodynamic radius for the same mass are constructed with amino acids that do not support structures such as alpha helices or beta-sheets. According to the rating of amino acid residues by the Chou-Fasman algorithm, residues A, D, E, Q, I, L, K, M, F, W, V support alpha-helical structure and residues C, Q, L, M, F, T, W, Y, V support beta-sheet structures. The amino acids that the Chou-Fasman algorithm considers most unstructured, because they are turn-forming, are, in order from most to least unstructured: G, N, P, D, S, C, Y, K. On balance, the residues that least support structure are G, N, P, S.

[0281] To achieve better fine tuning of the properties of the polymer, especially solubility and charge density, accessory polypeptides composed of two or three amino acids are generally preferred over those composed of a single amino acid. Accessory polypeptides that are composed of two or three types of amino acids are preferred because they offer the best balance of immunological simplicity (yielding only a small number of different 9-mer peptides can bind MHC complexes or 8-mer peptides that form epitopes for antibody binding), with the optimization of solubility, protease resistance, charge type and density, absence of structure, entropy, and non-specific binding to tissues (which can be undesirable but can also be used as a half-life mechanism). In general, the larger the number of non-human 8-mer or 9-mer peptides that can be created from the accessory protein sequence, the higher the risk of immunogenicity. Accordingly, in some aspects the accessory polypeptide comprises a small number of different 8-mer or 9-mer, and wherein all or most of these peptide sequences occur in the human proteome, preferably with many copies.

[0282] Where desired, a blend of two or three amino acids types can be optimized for obtaining the desired balance of
properties. The 20 natural amino acids (AA) can be separated into groups with related properties. Residues E, D (and to a lesser extent C) are negatively charged at physiological (neutral) pH, and residues K, R and to a lesser extent H are positively charged at neutral pH. The presence of charged residues E, D, K and R may be desirable for maximizing the water solubility of long polypeptides. For some biological applications it is desirable to have a high but equal or similar frequency of negative and positive residues that result in an uncharged, or nearly uncharged polypeptide that has high charge density but low net charge, such polypeptide tend to have low tendency for non-specific interactions with receptors that bind charged polymers such as heparin. For some biological applications, a single charge type (negative) which (unlike D) is chemically stable; thus favoring E (glutamate). The question is what the percentage of amino acids should be E, and whether the majority of non-charges amino acids should be G or S, and whether the sequence should be highly repetitive or less repetitive.

[0283] A high frequency of negatively charged residues E, D is likely to make the polymer bind to molecules with a large number of strong charges, like DNA binding proteins, histones and other K, R-rich sequences. A high frequency of positively charged residues K, R is likely to make the polymer bind to surfaces with a large number of negative charges, which includes most cell surfaces. Binding to cell surfaces is generally not desirable but a low degree of such non-specific binding may be useful to increase the half-life. The polar, hydrophilic amino acids N, Q, S, T, K, R, H, D, E, and additionally the amino acids C or Y can be useful in making accessible polypeptides that are relatively water soluble. Q and N can be glycosylation sites, offering a separate mechanism for increasing the hydrodynamic radius and thereby half-life. Non-polar, hydrophobic residues such as A, V, I, L, P, Y, F, W, M, C are less useful when creating a sequence with high water solubility, but it may be desirable to incorporate one or more of these residues at a low frequency, such that they constitute less than 10-20% of total. For example, a limited number of substitutions of hydrophobic residues can increase half-life by increasing non-specific binding to serum-exposed sites. Similarly, free thiols from cysteine residues may function as a mechanism for half-life extension by binding to one or more free thiols, such as the free thiol in human serum albumin. Also, these less-preferred amino acids can be used to create peptides that bind to serum-exposed proteins, thereby adding a second half-life extension mechanism, rather than hydrodynamic radius, to the accessory protein.

[0284] Glycine is a preferred residue that can be used in accessory polypeptides due to its high ratio of hydrodynamic radius to mass, or apparent molecular weight to predicted molecular weight. Glycine does not have a side chain and thus is the smallest residue, at 70Da. Because of its small size it provides maximal rotational freedom and maximum entropy. This makes it difficult for proteins to bind to sequences with higher frequencies of glycine, and glycine-rich sequences are highly protease resistant.

[0285] Residues C, W, N, Q, S, T, Y, K, R, H, D, E can form hydrogen bonds with other residues and thereby support structure (intermolecular hydrogen bonds) and binding to other proteins (intermolecular hydrogen bonds). These can be excluded in places where structure is not desired, or included if some degree of binding (specific or non-specific) is required for extension of half-life. The sulfur-containing residues C and M are typically avoided in accessory polypeptides, but cysteine can be included to provide half-life via its free thiol and can also be used in cyclic peptides as low-immunogenicity binding elements to extend half-life by binding to serum-exposed proteins or to obtain tissue targeting or modulated biodistribution by binding to tissue specific sites.

[0286] In one embodiment, the accessory polypeptide contains no or minimal repetitive sequence. IgM is pentavalent and exhibits propensity for recognizing repetitive sequences. Even low affinity contacts with IgM may lead to significant apparent affinity (avidity) due to the pentavalency of IgM. One way to build sequences with a reduced degree of repetition and reduced likelihood of IgM binding is to use repeat sequences that are long (ie 7, 8, 9, 10, 12, 14, 16, 20, 30, 40, 50, 70, 100, 150, 200 amino acids in repeat length). Examples of sequences with reduced repetition are (SESSSESSSE)n (SEQ ID NO: 48), (SESSSESSSSSSSSE)n (SEQ ID NO: 49), (SESSSESSSSSSSSSSSSE)n (SEQ ID NO: 50), or (SESSSESSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS
In another embodiment, the amino acids are chosen from the group consisting of E, S, G, and R. In yet another embodiment, the amino acids are E, G, and S (in any order).

In a related embodiment, the accessory polypeptide of the invention contains three different types of amino acids organized in repetitive sequence motifs, wherein each repeated sequence motif is longer than three consecutive amino acids. Exemplary sequences for this embodiment are shown in Table 1. Repetitive sequence motifs can be direct or inverted and 1, 2, 3, 4, or more different types of motifs can occur separately or intermixed in the same protein. The repeats may be perfect or imperfect, having 1, 2, 3, 4, 5, or more mismatched residues, and the repeats can be contiguous or dispersed, meaning they are separated by other, unrelated sequences that are not comprised of the same motif. In some embodiments, repetitive sequences constitute a majority of the accessory polypeptide, while non-repetitive sequences predominate in other embodiments. In one particular embodiment, a repetitive sequence contains interspersed single amino acids which break the strictly repetitive nature of the sequence. Exemplary sequences for this embodiment are shown in Table 1. In another related embodiment, the accessory polypeptide contains primarily three types of amino acids, organized in repetitive or non-repetitive sequences, together with a smaller number of amino acids of a different type, wherein the said three types of amino acid make up for more than 50%, 60%, 70%, 80%, 90%, 95%, 98%, or 99% of the entire sequence.

Another example of a sequence comprising multiple types of repeated motifs is GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG
TABLE 1—continued

Accessory polypeptide sequences containing three different types of amino acids.

<table>
<thead>
<tr>
<th>Accessory polypeptide sequences containing three different types of amino acids.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(EBSGGSGGG)n (SEQ ID NO: 71), (EBSGG)n (SEQ ID NO: 72), (EBSGG)n (SEQ ID NO: 73),</td>
</tr>
<tr>
<td>(EGSG)n (SEQ ID NO: 74), (EGSSEG)n (SEQ ID NO: 75)</td>
</tr>
<tr>
<td>(EGG)n (R) (EGG)n (SEQ ID NO: 76)</td>
</tr>
<tr>
<td>(EGG)n (P) (EGG)n (SEQ ID NO: 77)</td>
</tr>
</tbody>
</table>

[0298] Accessory Polypeptides Containing Two Different Types of Amino Acids:
[0299] In one embodiment, the accessory polypeptide comprises a sequence containing two different types of amino acids, wherein one of the amino acids is glycine and the other is D, E, K, R, R, F, S, T, A, H, N, R, L, V, W, M, F, I or C. A more specific embodiment provides an accessory polypeptide comprising a sequence containing two different types of amino acids, wherein one of the amino acids is glycine, and wherein glycine makes up 0%, half or less than half of the entire sequence. In related embodiments, the accessory polypeptide comprises 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% or even 100% glycine residues.

[0301] In a related embodiment, the accessory polypeptide comprises a sequence containing two different types of amino acids, wherein one of the amino acids is serine and the other is D, E, K, P, R, G, T, A, H, N, Y, L, V, W, M, F, I or C. A more specific embodiment provides an accessory polypeptide comprising a sequence containing two different types of amino acids, wherein one of the amino acids is serine, and wherein serine makes up 0%, half or less than half of the entire sequence. In related embodiments, the accessory polypeptide comprises 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% or even 100% serine residues.

[0302] In a related embodiment, the accessory polypeptide comprises two different types of amino acids, wherein the amino acids are represented in equal or about equal amounts (1:1 ratio). In related embodiments, the two types of amino acids are represented in 1:2, 1:3, 2:3, 3:4 ratios. Example sequences are shown in Table.

[0303] An alternative embodiment of the present invention provides an accessory polypeptide comprising a sequence containing two different types of amino acids, wherein half or less than half of the total amino acids are A, T, G, D, E or H.

[0304] An alternative embodiment of the present invention provides an accessory polypeptide comprising a sequence containing two different types of amino acids, wherein half or more of the amino acids are G and half or less than half of the total amino acids are A, S, T, D, E or H.

[0305] Another embodiment of the present invention provides an accessory polypeptide comprising a sequence containing two different types of amino acids, wherein half or more of the amino acids are S and half or less than half of the total amino acids are A, T, G, D, E or H.

[0306] Another embodiment of the present invention provides an accessory polypeptide comprising a sequence containing two different types of amino acids, wherein half or less than half of the total amino acids are P, R, L, V, Y, W, M, F, I, K or C.

[0307] Accessory polypeptides are also envisioned comprising repeating sequence motifs, wherein the sequence motifs can consist of 2, 3, 4, 5, 6, 7, 8, 9 or more amino acids.

[0308] The composition of amino acids in the motif or in the polymeric sequence can be balanced (for example, 50% A and 50% B), or unbalanced (i.e., 75% A and 25% B).

[0309] The accessory polypeptide repeats can be located at the N-terminus of the protein, at the C-terminus of the protein or 1, 2, 3, 4, 5, 6, 10, 20, 30 or more amino acid residues away from the N-terminus or C-terminus. The polynaminob acid can also lie between two protein domains.

[0310] The number of repeats of a motif in a polynaminob acid can have a lower limit of 4, 5, 6, 7, 8, 9, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 45, 50, 50, 70, 80, 90, 100 and an upper limit of 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 45, 50, 50, 70, 80, 90, 100, 150, 200, 250, 300, 400, 500, or even 600.

[0311] Possible motifs comprising two amino acids are AD, AE, AF, AG, AH, AI, AK, AL, AM, AN, AP, AQ, AR, AS, AT, AV, AW, AX, YA, DA, DE, DF, DG, DH, DI, DK, DL, DM, DN, DP, DQ, DR, DS, DT, DV, DW, DX, DY, E, ED, EE, EF, EG, EH, EI, EK, EL, EM, EN, EP, EQ, ER, ES, ET, EV, EW, EX, FA, FD, FE, FG, FH, FI, FK, FL, FM, FN, FP, FQ, FR, FS, FT, FY, FY, GA, GD, GE, GF, GH, GL, GK, GM, GN, GP, GQ, GR, GS, GT, GV, GW, GY, HA, HD, HE, HF, HG, HL, HK, HL, HM, HN, HP, HQ, HR, HS, HT, HV, HW, HX, HY, IA, ID, IE, IF, IG, IH, IL, IM, IN, IP, IQ, IR, IS, IT, IV, IW, IY, KA, KD, KE, KE, KG, KH, KL, KM, KN, KP, KQ, KR, KS, KT, KV, KW, KY, LA, LD, LE, LF, LG, LH, LI, LK, LM, LN, LP, LQ, LS, LT, LV, LW, LY, MA, MD, ME, MF, MG, MH, MI, MK, ML, MN, MP, MQ, MR, MS, MT, MV, MW, MY, NA, ND, NE, NF, NG, NH, NI, NK, NL, NM, NN, NP, NQ, NR, NS, NT, NV, NR, NY, PA, PD, PE, PF, PG, PH, PL, PK, PL, PM, PN, PQ, PR, PS, PT, PW, PY, Q, QA, QE, QF, QG, QH, QI, QJ, QK, QM, QN, QP, QQ, QS, QT, QV, QW, QY, RA, RD, RE, RF, RG, RH, RL, RM, RN, RP, RQ, RR, RS, RT, RV, RW, RY, SA, SD, SE, SF, SG, SH, SI, SK, SL, SM, SN, SP, SQ, SR, SS, ST, SV, SW, SY, TA, TD, TE, TF, TG, TH, TI, TK, TL, TM, TN, TP, TQ, TR, TS, TV, TW, TY, VA, VA, VE, VF, VG, VH, VL, VK, VL, VM, VN, VP, VR, VS, VT, VW, VV, VY, WA, WD, WE, WF, WG, WH, WI, WK, WL, WM, WN, WP, WQ, WR, WS, WT, WV, WY, YA, YD, YE, YF, YG, YH, YI, YK, YL, YM, YN, YP, YQ, YR, YS, YT, YV, YW. Of these, the preferred 2 amino acid motifs are EG and GE (forming the polymer EGEGEGEGEGE (SEQ ID NO: 78) and other variants), GS and SG (forming the polymer SGSGSGSGSGSGS (SEQ ID NO: 79) and other variants), and SE and SE (forming the polymer SSESESESESESESESESESESE (SEQ ID NO: 80) and other variants). These can also repeat 3, 4, 5, 6 or 7 amino acid residues. It is possible for the repeats to comprise 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 or even 20 residues. Each such repeat may contain 2, 3, 4, 5 or more types of amino acids, up to the number of residues present in the repeat.
Possible motifs comprising two amino acids are

Accessory polypeptide sequences that are closely related to sequences of human proteins are desirable in some applications as they carry a diminished risk of inducing an immune reaction in patients. Such sequences may be used as accessory polypeptides in embryos of the present invention. The relationship of accessory sequences to human sequences can be assessed by determining the abundance of partial sequences of said accessory polypeptide sequences in the human genome. Table 3 shows an example for the occurrence of 8mer partial sequences. Accessory polypeptides can be cleaved into a small number of 8mer sequences as illustrated in Table 3, where the 8mer sequences are underlined. For each 8mer sequence one can perform a data base search to identify the number of matches in a data base of human protein sequences. A similar analysis can be performed for 7mers, 9mers, 10mers, 11mers, or longer oligomers. One can perform database analysis searching for complete matches of these partial sequences or one can search for close homologues. Thus, the stringency of the search can be tuned to allow a ranking of accessory polypeptides for their relationship to human proteins. The data in Table 3 shows several examples of accessory polypeptides which are chosen based on their close relatedness to human proteins. Of particular interest are accessory proteins of sequence (SSSSE), (SEQ ID NO: 53), (SSSSEEESSSS), (SEQ ID NO: 55), and (SSSSESSSSSSSS), (SEQ ID NO: 51) where all 8mer subsequences can be found in several human proteins.

**TABLE 2**

<table>
<thead>
<tr>
<th>Accessory polypeptides containing two different types amino acids</th>
<th>(DDSE)</th>
<th>(DDDG)</th>
<th>(DDGQ)</th>
<th>(DDGP)</th>
<th>(DDGR)</th>
<th>(DDHE)</th>
<th>(DDHEE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(SEQ ID NO: 81)</td>
<td>(SEQ ID NO: 92)</td>
<td>(SEQ ID NO: 93)</td>
<td>(DDPP)</td>
<td>(SEQ ID NO: 84)</td>
<td>(DDSD)</td>
<td>(SEQ ID NO: 85)</td>
<td>(DDSE)</td>
</tr>
<tr>
<td>(SEQ ID NO: 87)</td>
<td>(SEQ ID NO: 88)</td>
<td>(SEQ ID NO: 89)</td>
<td>(SEQ ID NO: 90)</td>
<td>(SEQ ID NO: 91)</td>
<td>(SEQ ID NO: 92)</td>
<td>(SEQ ID NO: 93)</td>
<td>(SEQ ID NO: 94)</td>
</tr>
<tr>
<td>(DDKQ)</td>
<td>(DDKQ)</td>
<td>(DDKQ)</td>
<td>(DDKQ)</td>
<td>(DDKQ)</td>
<td>(DDKQ)</td>
<td>(DDKQ)</td>
<td>(DDKQ)</td>
</tr>
<tr>
<td>(SEQ ID NO: 95)</td>
<td>(SEQ ID NO: 96)</td>
<td>(SEQ ID NO: 97)</td>
<td>(SEQ ID NO: 98)</td>
<td>(SEQ ID NO: 99)</td>
<td>(SEQ ID NO: 100)</td>
<td>(SEQ ID NO: 101)</td>
<td>(SEQ ID NO: 102)</td>
</tr>
<tr>
<td>(SIGN)</td>
<td>(SIGN)</td>
<td>(SIGN)</td>
<td>(SIGN)</td>
<td>(SIGN)</td>
<td>(SIGN)</td>
<td>(SIGN)</td>
<td>(SIGN)</td>
</tr>
<tr>
<td>(SEQ ID NO: 103)</td>
<td>(SEQ ID NO: 104)</td>
<td>(SEQ ID NO: 105)</td>
<td>(SEQ ID NO: 106)</td>
<td>(SEQ ID NO: 107)</td>
<td>(SEQ ID NO: 108)</td>
<td>(SEQ ID NO: 109)</td>
<td>(SEQ ID NO: 110)</td>
</tr>
<tr>
<td>(SSSSE)</td>
<td>(SSSSE)</td>
<td>(SSSSE)</td>
<td>(SSSSE)</td>
<td>(SSSSE)</td>
<td>(SSSSE)</td>
<td>(SSSSE)</td>
<td>(SSSSE)</td>
</tr>
<tr>
<td>(SEQ ID NO: 111)</td>
<td>(SEQ ID NO: 112)</td>
<td>(SEQ ID NO: 113)</td>
<td>(SEQ ID NO: 114)</td>
<td>(SEQ ID NO: 115)</td>
<td>(SEQ ID NO: 116)</td>
<td>(SEQ ID NO: 117)</td>
<td>(SEQ ID NO: 118)</td>
</tr>
<tr>
<td>(SSSSEQ)</td>
<td>(SSSSEQ)</td>
<td>(SSSSEQ)</td>
<td>(SSSSEQ)</td>
<td>(SSSSEQ)</td>
<td>(SSSSEQ)</td>
<td>(SSSSEQ)</td>
<td>(SSSSEQ)</td>
</tr>
<tr>
<td>(SEQ ID NO: 119)</td>
<td>(SEQ ID NO: 120)</td>
<td>(SEQ ID NO: 121)</td>
<td>(SEQ ID NO: 122)</td>
<td>(SEQ ID NO: 123)</td>
<td>(SEQ ID NO: 124)</td>
<td>(SEQ ID NO: 125)</td>
<td>(SEQ ID NO: 126)</td>
</tr>
<tr>
<td>(SSSST)</td>
<td>(SSSST)</td>
<td>(SSSST)</td>
<td>(SSSST)</td>
<td>(SSSST)</td>
<td>(SSSST)</td>
<td>(SSSST)</td>
<td>(SSSST)</td>
</tr>
<tr>
<td>(SEQ ID NO: 127)</td>
<td>(SEQ ID NO: 128)</td>
<td>(SEQ ID NO: 129)</td>
<td>(SEQ ID NO: 130)</td>
<td>(SEQ ID NO: 131)</td>
<td>(SEQ ID NO: 132)</td>
<td>(SEQ ID NO: 133)</td>
<td>(SEQ ID NO: 134)</td>
</tr>
</tbody>
</table>

<p>| (SSSSSEEESSSS) | (SEQ ID NO: 135) | (SSSSSEEESSSS) | (SEQ ID NO: 136) | (SSSSSEEESSSS) | (SEQ ID NO: 137) | (SSSSSEEESSSS) | (SEQ ID NO: 138) |
| (SEQ ID NO: 139) | (SEQ ID NO: 140) | (SEQ ID NO: 141) | (SEQ ID NO: 142) | (SEQ ID NO: 143) | (SEQ ID NO: 144) | (SEQ ID NO: 145) | (SEQ ID NO: 146) |</p>
<table>
<thead>
<tr>
<th>(DE)n (DG)n (DK)n (DP)n (DS)n (DT)n (ED)n (EG)n (EK)n (EP)n (ER)n (ET)n (GD)n (GE)n (GI)n (GS)n (GT)n (RD)n (RG)n (RS)n (TP)n (TR)n (TS)n</th>
</tr>
</thead>
<tbody>
<tr>
<td>(DE)n (DG)n (DK)n (DP)n (DS)n (DT)n (ED)n (EG)n (EK)n (EP)n (ER)n (ET)n (GD)n (GE)n (GI)n (GS)n (GT)n (RD)n (RG)n (RS)n (TP)n (TR)n (TS)n</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TABLE 2-continued</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accessory polypeptides containing two different types amino acids</td>
</tr>
</tbody>
</table>

146, (KEEKSS) (SEQ ID NO: 147), (KEETTTT) (SEQ ID NO: 148), (KKKKDDD) (SEQ ID NO: 149), (KKKKEEE) (SEQ ID NO: 150), (KKKKEGGG) (SEQ ID NO: 151), (KKKKEPPP) (SEQ ID NO: 152), (KKKKESSS) (SEQ ID NO: 153), (KKKKEEEE) (SEQ ID NO: 154), (KKKKEEEE) (SEQ ID NO: 155), (PPPDDD) (SEQ ID NO: 156), (PPPPEEE) (SEQ ID NO: 157), (PPPPGGG) (SEQ ID NO: 158), (PPPKEKK) (SEQ ID NO: 159), (PPPPPEE) (SEQ ID NO: 160), (PPPPSSS) (SEQ ID NO: 161), (PPPFTTT) (SEQ ID NO: 162), (PPRRRRD) (SEQ ID NO: 163), (PPRRREE) (SEQ ID NO: 164), (PPRRRGGG) (SEQ ID NO: 165), (PPRRRRKK) (SEQ ID NO: 166), (PPRRRRPP) (SEQ ID NO: 167), (PPRRRREE) (SEQ ID NO: 168), (PPRRRRTT) (SEQ ID NO: 169), (PPSSSSDD) (SEQ ID NO: 170), (PPSSSSSE) (SEQ ID NO: 171), (PPSSSSGG) (SEQ ID NO: 172), (PPSSSSKE) (SEQ ID NO: 173), (PPSSSSPP) (SEQ ID NO: 174), (PPSSSSRR) (SEQ ID NO: 175), (PPSSSSTT) (SEQ ID NO: 176), (TTTTSSDD) (SEQ ID NO: 177), (TTTTSSEE) (SEQ ID NO: 178), (TTTTSSGG) (SEQ ID NO: 179), (TTTTTEKK) (SEQ ID NO: 180), (TTTTTPPP) (SEQ ID NO: 181), (TTTTTRRR) (SEQ ID NO: 182), (TTTTSSSS) (SEQ ID NO: 183).
**TABLE 2-continued**

Accessory polypeptides containing two different type amino acids


[0315] Accessory Polypeptide Sequences that are Related to Human Sequences

[0316] Accessory polypeptide sequences that are closely related to sequences of human proteins are desirable in some applications as they carry a diminished risk of inducing an immune reaction in patients. Such sequences may be used as accessory polypeptides in some embodiments of the present invention. The relationship of accessory sequences to human sequences can be assessed by determining the abundance of partial sequences of said accessory polypeptide sequences in the human genome. Table 3 shows an example for the occurrence of 8mer partial sequences. Accessory polypeptides can be cleaved into a small number of 8mer sequences as illustrated in Table 3, where the 8mer sequences are underlined.

For each 8mer sequence one can perform a data base search to identify the number of matches in a data base of human protein sequences. A similar analysis can be performed for 7mers, 9mers, 10mers, 11mers, or longer oligomers. One can perform database analysis searching for complete matches of these partial sequences or one can search for close homologues. Thus, the stringency of the search can be tuned to allow a ranking of accessory polypeptides for their relationship to human proteins. The data in Table 3 shows several examples of accessory polypeptides which are chosen based on their close relatedness to human proteins. Of particular interest are accessory proteins of sequence (SSSSE), (SSSSESS), and (SSSSESSSSS), where all 8mer subsequences can be found in several human proteins.
<table>
<thead>
<tr>
<th>Repeating unit</th>
<th>Seq ID NO: 49</th>
<th>Seq ID NO: 53</th>
<th>Seq ID NO: 54</th>
<th>Seq ID NO: 55</th>
<th>Seq ID NO: 369</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSSSSSSESSSE</td>
<td>364 4</td>
<td>365 10</td>
<td>366 14</td>
<td>367 50</td>
<td>369 58</td>
</tr>
<tr>
<td>SSSSSSSESSSE</td>
<td>364 4</td>
<td>365 9</td>
<td>366 10</td>
<td>367 14</td>
<td>369 14</td>
</tr>
<tr>
<td>SSSSSSSESSSE</td>
<td>364 25</td>
<td>366 9</td>
<td>367 10</td>
<td>367 25</td>
<td>369 10</td>
</tr>
<tr>
<td>SSSSSSSESSSE</td>
<td>364 12</td>
<td>366 0</td>
<td>367 25</td>
<td>367 25</td>
<td>369 9</td>
</tr>
<tr>
<td>SSSSSSSESSSE</td>
<td>364 0</td>
<td>366 0</td>
<td>367 25</td>
<td>367 25</td>
<td>369 25</td>
</tr>
</tbody>
</table>

**[0317]** Unstructured Recombinant Polymers (URPs): One aspect of the present invention is the use of unstructured recombinant polymers (URPs) as accessory polypeptides. The subject URPs are particularly useful for generating recombinant proteins of therapeutic and/or diagnostic value. The subject URPs exhibit one or more following features.

**[0318]** The subject URPs comprise amino acid sequences that typically share commonality with denatured peptide sequences under physiological conditions. URP sequences typically behave like denatured peptide sequences under physiological conditions. URP sequences lack well defined secondary and tertiary structures under physiological conditions. A variety of methods have been established in the art to ascertain the second and tertiary structures of a given polypeptide. For example, the secondary structure of a polypeptide can be determined by CD spectroscopy in the "far-UV" spectral region (190-250 nm). Alpha-helix, beta-sheet, and random coil structures each give rise to a characteristic shape and magnitude of CD spectra. Secondary structure can also be ascertained via certain computer programs or algorithms such as the Chou-Fasman algorithm (Chou, P.Y., et al. (1974) *Biochemistry*, 13: 222-45). For a given URP sequence, the algorithm can predict whether there exists some or no secondary structure at all. In general, URP sequences will have spectra that resemble denatured sequences due to their low degree of secondary and tertiary structure. Where desired, URP sequences can be designed to have predominantly denatured conformations under physiological conditions. URP sequences typically have a high degree of conformational flexibility under physiological conditions and they tend to have large hydrodynamic radii (Stokes' radius) compared to globular proteins of similar molecular weight. As used herein, physiological conditions refer to a set of condi-
tions including temperature, salt concentration, pH that mimic those conditions of a living subject. A host of physiologically relevant conditions for use in in vitro assays have been established. Generally, a physiological buffer contains a physiological concentration of salt and is adjusted to a neutral pH ranging from about 6.5 to about 7.8, and preferably from about 7.0 to about 7.5. A variety of physiological buffers is listed in Sambrook et al. (1989) supra and hence is not detailed herein. Physiologically relevant temperature ranges from about 25°C to about 38°C, and preferably from about 30°C to about 37°C.

[0320] The subject URP can be sequences with low immunogenicity. Low immunogenicity can be a direct result of the conformational flexibility of UR sequences. Many antibodies recognize so-called conformational epitopes in protein antigens. Conformational epitopes are formed by regions of the protein surface that are composed of multiple discontinuous amino acid sequences of the protein antigen. The precise folding of the protein brings these sequences into a well-defined special configuration that can be recognized by antibodies. Preferred URP are designed to avoid formation of conformational epitopes. For example, of particular interest are URP sequences having a low tendency to adopt compactly folded conformations in aqueous solution. In particular, low immunogenicity can be achieved by choosing sequences that resist antigen processing in antigen presenting cells, choosing sequences that do not bind MHC well and/or by choosing sequences that are derived from human sequences.

[0321] The subject URP can be sequences with a high degree of protease resistance. Protease resistance can also be a result of the conformational flexibility of URP sequences. Protease resistance can be designed by avoiding known protease recognition sites. Alternatively, protease resistant sequences can be selected by phage display or related techniques from random or semi-random sequence libraries. Where desired for special applications, such as slow release from a depot protein, serum protease cleavage sites can be built into an URP. Of particular interest are URP sequences with high stability (e.g., long serum half-life, less prone to cleavage by proteases present in bodily fluid) in blood.

[0322] The subject URP can also be characterized by the effect of wherein upon incorporation of it into a biologically active polypeptide, the modified polypeptide exhibits a longer serum half-life and/or higher solubility as compared to an unmodified biologically active polypeptide. The subject URP can be of any length necessary to effect (a) an increase in stability of protein comprising the URP; (b) an increase in solubility of the resulting protein; (c) an increased resistance to protease; and/or (d) a reduced immunogenicity of the resulting protein that comprises the URP. Typically, the subject URP has about 30, 40, 50, 60, 70, 80, 90, 100, 150, 200, 300, 400 or more contiguous amino acids. When incorporated into a protein, the URP can be fragmented such that the resulting protein contains multiple URP, or multiple fragments of URP. Some or all of these individual URP sequences may be shorter that 40 amino acids as long as the combined length of all URP sequences in the resulting protein is at least 40 amino acids. Preferably, the resulting protein has a combined length of URP sequences exceeding 40, 50, 60, 70, 80, 90, 100, 150, 200 or more amino acids.

[0323] URP may have an isoelectric point (pI) of 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5, 5.0, 5.5, 6.0, 6.5, 7.0, 7.5, 8.0, 8.5, 9.0, 9.5, 10.0, 10.5, 11.0, 11.5, 12.0, 12.5 or even 13.0.

[0324] In general, URP sequences are rich in hydrophilic amino acids and contain a low percentage of hydrophobic or aromatic amino acids. Suitable hydrophilic residues include but are not limited to glycine, serine, aspartate, glutamate, lysine, arginine, and threonine. Hydrophobic residues that are less favored in construction of URP include tryptophan, phenylalanine, tyrosine, leucine, isoleucine, valine, and methionine. URP sequences can be rich in glycine but URP sequences can also be rich in the amino acids glutamate, aspartate, serine, threonine, alanine or proline. Thus the predominant amino acid may be G, E, D, S, T, A or P. The inclusion of proline residues tends to reduce sensitivity to proteolytic degradation.

[0325] The inclusion of hydrophilic residues typically increases URP’s solubility in water and aqueous media under physiological conditions. As a result of their amino acid composition, URP sequences have a low tendency to form aggregates in aqueous formulations and the fusion of URP sequences to other biologically active polypeptides or peptides tends to enhance their solubility and reduce their tendency to form aggregates, which is a separate mechanism to reduce immunogenicity.

[0326] URP sequences can be designed to avoid certain amino acids that confer undesirable properties to the biologically active polypeptide. For instance, one can design URP to contain few or none of the following amino acids: cysteine (to avoid disulfide formation and oxidation), methionine (to avoid oxidation), asparagine and glutamine (to avoid deamidation).

[0327] Glycine-Rich URP:

[0328] In one embodiment, the subject URP comprises a glycine rich sequence (GRS). For example, glycine can be present predominantly such that it is the most prevalent residues present in the sequence of interest. In another example, URP sequences can be designed such that glycine residues constitute at least about 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 100% of the total amino acids. URP can also contain 100% glycines. In yet another example, the URP contains at least 30% glycine and the total concentration of tryptophan, phenylalanine, tyrosine, valine, leucine, and isoleucine is less than 20%. In still another example, the URP contains at least 40% glycine and the total concentration of tryptophan, phenylalanine, tyrosine, valine, leucine, and isoleucine is less than 10%. In still yet another example, the URP contains at least 50% glycine and the total concentration of tryptophan, phenylalanine, tyrosine, valine, leucine, and isoleucine is less than 5%.

[0329] The length of GRS can vary between about 5 amino acids and 200 amino acids or more. For example, the length of a single, contiguous GRS can contain 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 70, 80, 90, 100, 120, 140, 160, 180, 200, 240, 280, 320 or 400 or more amino acids. GRS may comprise glycine residues at both ends.

[0330] GRS can also have a significant content of other amino acids, for example Ser, Thr, Ala, or Pro. GRS can contain a significant fraction of negatively charged amino acids including but not limited to Asp and Glu. GRS can contain a significant fraction of positively charged amino acids including but not limited to Arg or Lys. Where desired, URP can be designed to contain only a single type of amino acid (i.e., Gly or Glu), sometimes only a few types of amino acid, e.g., two to five types of amino acids (e.g., selected from G, E, D, S, T, A and P), in contrast to typical proteins and typical linkers which generally are composed of most of the
twenty types of amino acids. URP s may contain negatively
treated residues (Asp, Glu) in 30, 25, 20, 15, 12, 10, 9, 8, 7,
6, 5, 4, 3, 2, or 1 percent of the amino acids positions.

[0331] Typically, the subject GRS-containing URP has
about 30, 40, 50, 60, 70, 80, 90, 100, or more contiguous
peptide chains. When incorporated into a biologically active
polymerase, the URP can be fragmentated such that the resulting
modified polymer contains multiple URP s, or multiple
fragments of URP s. Some or all of these individual URP
sequences may be shorter that 40 amino acids as long as the
combined length of all URP sequences in the resulting polymer
is at least 30 amino acids. Preferably, the resulting
polymer has a combined length of URP sequences exceeding 40, 50, 60, 70, 80, 90, 100, or more amino acids.

[0332] The GRS-containing URP s are of particular interest
duo to, in part, the increased conformational freedom of gly-
cine-containing peptides. Denatured peptides in solution
have a high degree of conformational freedom. Most of that
conformational freedom is lost upon binding of said peptides
to a target like a receptor, an antibody, or a protease. This loss of
entropy needs to be offset by the energy of interaction
between the peptide and its target. The degree of conforma-
tional freedom of a denatured peptide is dependent on its
amino acid sequence. Peptides containing many amino acids
and with small side chains tend to have more conformational
freedom than peptides that are composed of amino acids with
larger side chains. Peptides containing the amino acid glycine
have particularly large degrees of freedom. It has been esti-
imated that glycine-containing peptide bonds have about 3.4
times more entropy in solution as compared to corresponding
Proteins, 25: 143-56). This factor increases with the number
of glycine residues in a sequence. As a result, such peptides
tend to lose more entropy upon binding to targets, which
reduces their overall ability to interact with other proteins as
well as their ability to adopt defined three-dimensional struc-
tures. The large conformational flexibility of glycine-peptide
bonds is also evident when analyzing Ramachandran plots of
protein structures where glycine peptide bonds occupy areas
that are rarely occupied by other peptide bonds (Venkatachala-
et al. studied a database of 12,320 residues from 61 nonho-
logous, high resolution crystal structures to determine the
phi, psi conformational preferences of each of the 20 amino
acids. The observed distributions in the native state of pro-
teins are assumed to also reflect the distributions found in
the denatured state. The distributions were used to approximate
the energy surface for each residue, allowing the calculation
of relative conformational entropies for each residue relative
to glycine. In the most extreme case, replacement of glycine
by proline, conformational entropy changes will stabilize the
native state relative to the denatured state by $-0.82+/-0.08$
132). These observations confirm the special role of glycine
among the 20 natural amino acids.

[0333] In designing the subject URP s, natural or non-natu-
ral sequences can be used. For example, a host of natural
sequences containing high glycine content is provided in
Table 4. Table 5, Table 6, and Table 7. One skilled in the art
may adopt any one of the sequences as an URP, or modify the
sequences to achieve the intended properties. Where immu-
nogenicity to the host subject is of concern, it is preferable
to design GRS-containing URR s based on glycine rich
sequences derived from the host. Preferred GRS-containing
URR s are sequences from human proteins or sequences that
share substantial homology to the corresponding glycine rich
sequences in the reference human proteins.

<table>
<thead>
<tr>
<th>PDB file</th>
<th>Protein function</th>
<th>Glycine rich sequences</th>
</tr>
</thead>
<tbody>
<tr>
<td>1N3V</td>
<td>Porcine Parvovirus capsid</td>
<td>ggggggggsggggggggg (SEQ ID NO: 370)</td>
</tr>
<tr>
<td>1FPV</td>
<td>Feline Panleukopenia Virus</td>
<td>tgggpggggsggggggggg (SEQ ID NO: 371)</td>
</tr>
<tr>
<td>1IJS</td>
<td>CpV strain D, mutant</td>
<td>tgggpggggsggggggggg (SEQ ID NO: 371)</td>
</tr>
<tr>
<td>1MVV</td>
<td>Mm (strain I) virus</td>
<td>ggggggggsggggggggg (SEQ ID NO: 372)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Accession</th>
<th>Organism</th>
<th>Gly (%)</th>
<th>GRS length</th>
<th>Gene length</th>
<th>Predicted Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>NP_97499</td>
<td>Arabidopsis thaliana</td>
<td>64</td>
<td>509</td>
<td>579</td>
<td>unknown</td>
</tr>
<tr>
<td>ZP_0045077</td>
<td>Bacteriobacteria cecopinacea</td>
<td>66</td>
<td>373</td>
<td>518</td>
<td>putative lipoprotein</td>
</tr>
<tr>
<td>NP_427841</td>
<td>Oryza sativa</td>
<td>74</td>
<td>371</td>
<td>422</td>
<td>unknown</td>
</tr>
<tr>
<td>NP_910409</td>
<td>Oryza sativa</td>
<td>75</td>
<td>368</td>
<td>400</td>
<td>putative cell-wall precursor</td>
</tr>
<tr>
<td>NP_510800</td>
<td>Drosophila melanogaster</td>
<td>66</td>
<td>322</td>
<td>610</td>
<td>transposable element</td>
</tr>
</tbody>
</table>
TABLE 6

<table>
<thead>
<tr>
<th>Accession</th>
<th>Gly (%)</th>
<th>GRS length</th>
<th>Gene length</th>
<th>Hydrophobics</th>
<th>Predicted Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>NP_000217</td>
<td>62</td>
<td>135</td>
<td>622</td>
<td>yes</td>
<td>keratin 9</td>
</tr>
<tr>
<td>NP_651961</td>
<td>61</td>
<td>73</td>
<td>592</td>
<td>yes</td>
<td>TBP-associated factor 15 isoform 1</td>
</tr>
<tr>
<td>NP_476429</td>
<td>65</td>
<td>70</td>
<td>629</td>
<td>yes</td>
<td>keratin 3</td>
</tr>
<tr>
<td>NP_006418</td>
<td>70</td>
<td>66</td>
<td>316</td>
<td>yes</td>
<td>laminin, cell envelope</td>
</tr>
<tr>
<td>NP_056932</td>
<td>60</td>
<td>66</td>
<td>638</td>
<td>yes</td>
<td>cytokeratin 2</td>
</tr>
</tbody>
</table>

TABLE 7

<table>
<thead>
<tr>
<th>Accession</th>
<th>Sequences</th>
<th>Number of amino acids</th>
</tr>
</thead>
<tbody>
<tr>
<td>NP_006228</td>
<td>GPGGPGGGPGGPGGPGGPGGPGGPGGPGGPGGPGG</td>
<td>37</td>
</tr>
<tr>
<td>NP_787059</td>
<td>GAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG</td>
<td>33</td>
</tr>
<tr>
<td>NP_009060</td>
<td>GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG</td>
<td>32</td>
</tr>
<tr>
<td>NP_031393</td>
<td>GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG</td>
<td>27</td>
</tr>
<tr>
<td>NP_005850</td>
<td>GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG</td>
<td>25</td>
</tr>
<tr>
<td>NP_061856</td>
<td>GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG</td>
<td>23</td>
</tr>
<tr>
<td>NP_787059</td>
<td>GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG</td>
<td>33</td>
</tr>
<tr>
<td>NP_009060</td>
<td>GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG</td>
<td>32</td>
</tr>
<tr>
<td>NP_031393</td>
<td>GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG</td>
<td>27</td>
</tr>
<tr>
<td>NP_115818</td>
<td>GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG</td>
<td>21</td>
</tr>
<tr>
<td>XP_376532</td>
<td>GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG</td>
<td>18</td>
</tr>
<tr>
<td>NP_065104</td>
<td>GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG</td>
<td>12</td>
</tr>
</tbody>
</table>

[0334] (Table 7 discloses SEQ ID NOS 373-378, 374-376 and 379-381, respectively, in order of appearance.)

POU domain, class 4, transcription factor 1 [Homo sapiens]

YEATS domain containing 2 [Homo sapiens]

AT rich interactive domain 1B (SWI1-like) isoform 2; BRG1-binding protein ELD/OSA1; Eld (eyelid)/Osa protein [Homo sapiens]

purine-rich element binding protein \Lambda; purine-rich single-stranded DNA-binding protein alpha; transcriptional activator protein PUR-alpha [Homo sapiens]

regulatory factor X1; trans-acting regulatory factor 1; enhancer factor C; MHC class II regulatory factor RFX [Homo sapiens]
bromo domain-containing protein disrupted in leukemia [Homo sapiens]

unknown protein [Homo sapiens]

PREDICTED: hypothetical protein XP_050256 [Homo sapiens]

zinc finger protein 281; ZNP-99 transcription factor [Homo sapiens]

RNA binding protein (autoantigenic, hnRNP-associated with lethal yellow) short isoform; RNA-binding protein (autoantigenic); RNA-binding protein (autoantigenic, hnRNP-associated with lethal yellow) [Homo sapiens]

signal recognition particle 68 kDa [Homo sapiens]

KIAA0265 protein [Homo sapiens]

engrailed homolog 2; Engrailed-2 [Homo sapiens]

RNA binding protein (autoantigenic, hnRNP-associated with lethal yellow) long isoform; RNA-binding protein (autoantigenic); RNA-binding protein (autoantigenic, hnRNP-associated with lethal yellow) [Homo sapiens]

androgen receptor; dihydrotestosterone receptor [Homo sapiens]

homeo box D11; homeo box 4F; Hox-4.6, mouse, homolog of; homeobox protein Hox-D11 [Homo sapiens]
PREDICTED: similar to THO complex subunit 4 (Tho4) (RNA and export factor binding protein 1) (REF1-I) (Aly of AML-1 and LEF-1) (Aly/REF) [Homo sapiens]

GOTREGTRGTRGGRGGRGGRG (SEQ ID NO: 407)

PREDICTED: similar to THO complex subunit 4 (Tho4) (RNA and export factor binding protein 1) (REF1-I) (Aly of AML-1 and LEF-1) (Aly/REF) [Homo sapiens]

GOTREGTRGTRGGRGGRGGRG (SEQ ID NO: 407)

POU domain, class 3, transcription factor 3 [Homo sapiens]

GAGGGGGGGGGGGGGGGGGGGGG (SEQ ID NO: 408)
nucleolar protein family A, member 1; GAR1 protein [Homo sapiens]

GRRGRRGRRGRRGRRGRRG (SEQ ID NO: 378)
fibrillin; 34-kD nucleolar scleroderma antigen; RNA, U3 small nuclear interacting protein 1 [Homo sapiens]

GRRGRRGRRGRRGRRGRRG (SEQ ID NO: 409)
zinc finger protein 579 [Homo sapiens]

GRRGRRGRRGRRGRRGRRG (SEQ ID NO: 410)
calpain, small subunit 1; calcium-activated neutral proteinase; calpain, small polypeptide; calpain 4, small subunit (30K); calcium-dependent protease, small subunit [Homo sapiens]

GAAAAAAGAAAAAGAAAAAGAAAA (SEQ ID NO: 411)
keratin 9 [Homo sapiens]

GSGSGGGGSGGSGGSGGSGGSGG (SEQ ID NO: 412)
forkhead box D1; forkhead-related activator 4; Forkhead, drosophila, homolog-like 8; forkhead (Drosophila)-like 8 [Homo sapiens]

GAGGGGGGGGGGGGGGGGGGGGG (SEQ ID NO: 413)
PREDICTED: similar to RIKEN cDNA C230094B 15 [Homo sapiens]

GGAGGAGGAGGAGGAGGAGGAGG (SEQ ID NO: 414)
cadherin 22 precursor; ortholog of rat PB-cadherin [Homo sapiens]

GGGGGGGGGGGGGGGGGGGGGG (SEQ ID NO: 415)

AT-binding transcription factor 1; AT motif-binding factor 1 [Homo sapiens]

GSSSSSSSSSSSSSSSSSSSSSSSSSS (SEQ ID NO: 416)
comesdermin; t box, brain, 2; comesdermin (Xenopus laevis) homolog [Homo sapiens]

GPAAGAGAGAGAGAGAGAGAGAGA (SEQ ID NO: 417)
phosphatidylinositol transfer protein, membrane-associated 2; PYK2 N-terminal domain-interacting receptor 3; retinal degeneration B alpha 2 (Drosophila) [Homo sapiens]

GSSSSSSSSSSSSSSSSSSSSSSSSSS (SEQ ID NO: 418)
sperm associated antigen 8 isoform 2; sperm membrane protein 1 [Homo sapiens]

GSSSGPGSGPSGPGPSGPGSSG (SEQ ID NO: 419)
PREDICTED: RNA binding motif protein 27 [Homo sapiens]

GPAGPGPGPGPGPGPGPGPG (SEQ ID NO: 420)
AP1 gamma subunit binding protein 1 isoform 1; gamma-synergin; adaptor-related protein complex 1 gamma subunit-binding protein 1 [Homo sapiens]

GSSSSSSSSSSSSSSSSSSSSSSSSSS (SEQ ID NO: 421)
AP1 gamma subunit binding protein 1 isoform 2; gamma-synergin; adaptor-related protein complex 1 gamma subunit-binding protein 1 [Homo sapiens]

GSSGAGAGAGAGAGAGAGAGAGAG (SEQ ID NO: 422)
ankyrin repeat and sterile alpha motif domain containing 1; ankyrin repeat and SAM domain containing 1 [Homo sapiens]

GSSSSSSSSSSSSSSSSSSSSSSSSSS (SEQ ID NO: 423)
methyl-CpG binding domain protein 2 isoform 1 [Homo sapiens]

GGRGGRGRRGGRGGRGGRG (SEQ ID NO: 424)
triple functional domain (PTPRF interacting) [Homo sapiens]

GSSSSSSSSSSSSSSSSSSSSSSSSSS (SEQ ID NO: 425)
forkhead box D3 [Homo sapiens]
sperm associated antigen 8 isoform 1; sperm membrane protein 1 [Homo sapiens]

methyl-CpG binding domain protein 2 testis-specific isoform [Homo sapiens]

cell death regulator aven; programmed cell death 12 [Homo sapiens]

regulator of nonsense transcripts 1; delta helicase; up-frame-shift mutation 1 homolog (S. cerevisiae); nonsense mRNA reducing factor 1; yeast Upf1p homolog [Homo sapiens]

small conductance calcium-activated potassium channel protein 2 isoform a; aminophosphoryl-sensitive small-conductance Ca2+-activated potassium channel [Homo sapiens]

SRY (sex determining region Y)-box 1; SRY-related HMGB gene 1 [Homo sapiens]

transcription factor 20 isoform 2; stromelysin-1 platelet-derived growth factor-responsive element binding protein; stromelysin 1 PDGF-responsive element binding protein; SPRE-binding protein; nuclear factor SPBP [Homo sapiens]

transcription factor 20 isoform 1; stromelysin-1 platelet-derived growth factor-responsive element binding protein; stromelysin 1 PDGF-responsive element binding protein; SPRE-binding protein; nuclear factor SPBP [Homo sapiens]

Ras-interacting protein 1 [Homo sapiens]

BMP-2 inducible kinase isoform b [Homo sapiens]

BMP-2 inducible kinase isoform a [Homo sapiens]

forkhead box C1; forkhead-related activator 3; Forkhead, drosophila, homolog-like 7; forkhead (Drosophila)-like 7; iridogoniodysgenesis type 1 [Homo sapiens]

splicing factor p54; arginine-rich 54 kDa nuclear protein [Homo sapiens]

v-maf musculoaponeurotic fibrosarcoma oncogene homolog; Avian musculoaponeurotic fibrosarcoma (MAF) protooncogene; v-maf musculoaponeurotic fibrosarcoma (avian) oncogene homolog [Homo sapiens]

small nuclear ribonucleoprotein D1 polypeptide 16 kDa; snRNP core protein D1; Sm-D autoantigen; small nuclear ribonucleoprotein D1 polypeptide (16 kD) [Homo sapiens]

hypothetical protein H41 [Homo sapiens]

URPs Containing Non-Glycine Residues (NGR): The sequences of non-glycine residues in these GRS can be selected to optimize the properties of URPs and hence the biologically active polypeptides that contain the desired URPs. For instance, one can optimize the sequences of URPs to enhance the selectivity of the resulting modified polypeptide for a particular tissue, specific cell type or cell lineage. For example, one can incorporate protein sequences that are not ubiquitously expressed, but rather are differentially expressed in one or more of the body tissues including heart, liver, prostate, lung, kidney, bone marrow, blood, skin, bladder, brain, muscles, nerves, and selected tissues that are affected by diseases such as infectious diseases, autoimmune disease, renal, nervous, cardiac disorders and cancers. One can employ sequences representative of a specific developmental origin, such as those expressed in an embryo or an adult, during ectoderm, endoderm or mesoderm formation in a multi-cellular organism. One can also utilize sequence involved in a specific biological process, including but not limited to cell cycle regulation, cell differentiation, apoptosis, chemotaxis, cell motility and cytoskeletal rearrangement. One can also utilize other non-ubiquitously expressed protein sequences to direct the resulting protein to a specific subcellular locations: extracellular matrix, nucleus, cytoplasm, cytoskeleton, plasma and/or intracellular membranous structures which include but are not limited to coated pits, Golgi apparatus, endoplasmic reticulum, endosome, lysosome, and mitochondria.

URPs can be derived from human sequences. The human genome contains many subsequences that are rich in one particular amino acid. Of particular interest are such
amino acid sequences that are rich in a hydrophilic amino acid like serine, threonine, glutamate, aspartate, or glycine. Of particular interest are such subsequences that contain few hydrophobic amino acids. Such subsequences are predicted to be unstructured and highly soluble in aqueous solution. Such human subsequences can be modified to further improve their utility. For example, dentin sialophosphoprotein contains a 670-amino acid subsequence in which 64% of the residues are serine and most other positions are hydrophilic amino acids such as aspartate, asparagines, and glutamate. The sequence is extremely repetitive and as a result it has a low information content. One can directly use subsequences of such a human protein. Where desired, one can modify the sequence in a way that preserves its overall character but which makes it more suitable for pharmaceutical applications. Examples of sequences that are related to dentin sialophosphoprotein are (SSD)$_n$, (SEQ ID NO: 438), (SSDSSSN)$_n$, (SEQ ID NO: 439), (SSE)$_n$, (SEQ ID NO: 440), where $n$ is between about 4 and 200.

[0339] Of particular interest are URP sequences that contain regions that are relatively rich in the positively charged amino acids arginine or lysine which favor cellular uptake or transport through membranes. URP sequences can be designed to contain one or several protease-sensitive subsequences. Such URP sequences can be cleaved once the product of the invention has reached its target location. This cleavage may trigger an increase in potency of the pharmaceutically active domain (pro-drug activation) or it may enhance binding of the cleavage product to a receptor. URP sequences can be designed to carry excess negative charges by introducing aspartic acid or glutamic acid residues. Of particular interest are URP that contain greater than 5%, greater than 6%, 7%, 8%, 9%, 10%, 15%, 30% or more glutamic acid and less than 2% lysine or arginine. Such URP carry an excess negative charge and as a result they have a tendency to adopt open conformations due to electrostatic repulsion between individual negative charges of the peptide. Such an excess negative charge leads to an effective increase in their hydrodynamic radius and as a result it can lead to reduced kidney clearance of such molecules. Thus, one can modulate the effective net charge and hydrodynamic radius of a URP sequence by controlling the frequency and distribution of negatively charged amino acids in the URP sequences. Most tissues and surfaces in a human or animal carry excess negative charges. By designing URP sequences to carry excess negative charges one can minimize non-specific interactions between the resulting modified polypeptide comprising the URP sequences and various surfaces such as blood vessels, healthy tissues, or various receptors.

[0340] URP may have a repetitive amino acid sequence of the format (Motif)$_n$, in which a sequence motif forms a direct repeat (ie ABCABCABCA or an inverted repeat (ABC-ABCAABCCBA) and the number of these repeats can be 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36 or more. URP or the repeats inside URP contain only 1, 2, 3, 4, 5 or 6 different types of amino acids. URP typically consist of repeats of human amino acid sequences that are 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 22, 24, 26, 28, 30, 32, 34, 36 or more amino acids long, but URP may also consist of non-human amino acid sequences that are 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50 amino acids long.

[0341] URP Derived from Human Sequences:

[0342] URP can be derived from human sequences. The human genome contains many subsequences that are rich in one particular amino acid. Of particular interest are such amino acid sequences that are rich in a hydrophilic amino acid like serine, threonine, glutamate, aspartate, or glycine. Of particular interest are subsequences that contain few hydrophobic amino acids. Such subsequences are predicted to be unstructured and highly soluble in aqueous solution. Such human subsequences can be modified to further improve their utility. For example, dentin sialophosphoprotein contains a 670-amino acid subsequence in which 64% of the residues are serine and most other positions are hydrophilic amino acids such as aspartate, asparagines, and glutamate. The sequence is extremely repetitive and as a result it has a low information content. One can directly use subsequences of such a human protein. Where desired, one can modify the sequence in a way that preserves its overall character but which makes it more suitable for pharmaceutical applications. Examples of sequences that are related to dentin sialophosphoprotein are (SSD)$_n$, (SSDSSN)$_n$, (SSE)$_n$, where $n$ is between about 4 and 200.

[0343] The use of sequences from human proteins is particularly desirable in design of URP with reduced immunogenicity in a human subject. A key step for eliciting an immune response to a foreign protein is the presentation of peptide fragments of said protein by MHC class II receptors. These MHCII-bound fragments can then be detected by T cell receptors, which triggers the proliferation of T helper cells and initiates an immune response. The elimination of T cell epitopes from pharmaceutical proteins has been recognized as a means to reduce the risk of eliciting an immune reaction (Stickler, M., et al. (2005). Immuno Methods, 281:95-108). MHCII receptors typically interact with an epitope having e.g., a 9-amino acid long region of the displayed peptides. Thus, one can reduce the risk of eliciting an immune response to a protein in patients if all or most of the possible 9mer subsequences of the protein can be found in human proteins and if so, these sequences and repeats of these sequences will not be recognized by the patient as foreign sequences. One can incorporate human sequences into the design of URP sequences by oligomerizing or concatenating human sequences that have suitable amino acid compositions. These can be direct repeats or inverted repeats or mixtures of different repeats. For instance one can oligomerize the sequences shown in table 5. Such oligomers have reduced risk of being immunogenic. However, the junction sequences between the monomer units can still contain T cell epitopes that can trigger an immune reaction. One can further reduce the risk of eliciting an immune response by designing URP sequences based on multiple overlapping human sequences. An URP sequence may be designed as an oligomer based on multiple human sequences such that each 9mer subsequences of the oligomer can be found in a human protein. In these designs, every 9-mer subsequence is a human sequence. For example an URP sequence may be based on three human sequences. It is also possible to design URP sequences based on a single human sequence such that all possible 9mer subsequences in the oligomeric URP sequences occur in the same human protein. Non-oligomeric URP sequences can be designed based on human proteins as well. The primary conditions are that all 9mer sub-sequences can be found in human sequences. The amino acid composition of the sequences preferably contains few hydrophobic residues. Of particular
interest are URP sequences that are designed based on human sequences and that contain a large fraction of glycine residues.

[0344] Utilizing this or a similar scheme, one can design a class of URPs that comprise repeat sequences with low immunogenicity to the host of interest. Host of interest can be any animals, including vertebrates and invertebrates. Preferred hosts are mammals such as primates (e.g., chimpanzees and humans), cetaceans (e.g., whales and dolphins), chiropterans (e.g., bats), perrissiodactyls (e.g., horses and rhinoceroses), rodents (e.g., rats), and certain kinds of insectivores such as shrews, moles and hedgehogs. Where human is selected as the host, the URPs typically contain multiple copies of the repeat sequences or units, wherein the majority of segments comprising about 6 to about 15 contiguous amino acids are present in one or more native human proteins. One can also design URPs in which the majority of segments comprising between about 9 to about 15 contiguous amino acids are found in one or more native human proteins. As used herein, majority of the segments refers to more than about 50%, preferably 60%, preferably 70%, preferably 80%, preferably 90%, preferably 100%. Where desired, each of the possible segments between about 6 to about 15 amino acids, preferably between about 9 to 15 amino acids within the repeating units are present in one or more native human proteins. The URPs can comprise multiple repeating units or sequences, for example having 2, 3, 4, 5, 6, 7, 8, 9, 10, or more repeating units.

[0345] Design of URPs that are Substantially Free of Human T-Cell Epitopes:

[0346] Non-limiting examples of URPs containing repeating amino acids are: poly-glycine, poly-glycine acid, poly-aspartic acid, poly-seryl, poly-threonine, (GX)n (SEQ ID NO: 441) where G is glycine and X is serine, aspartic acid, glutamic acid, threonine, or proline and n is at least 8, (GGG)n (SEQ ID NO: 442) where X is serine, aspartic acid, glutamic acid, threonine, or proline and n is at least 13, (GGG)n (SEQ ID NO: 443) where X is serine, aspartic acid, glutamic acid, threonine, or proline and n is at least 10, (GGG)n (SEQ ID NO: 444) where X is serine, aspartic acid, glutamic acid, threonine, or proline and n is at least 8, (GGG)n (SEQ ID NO: 445) where X is serine, aspartic acid, glutamic acid, threonine, or proline, n is at least 15, and z is between 1 and 20.

[0347] URPs can be designed to optimize protein production. This can be achieved by avoiding or minimizing repetitiveness of the encoding DNA. URPs sequences such as poly-glycine may have very desirable pharmaceutical properties but their manufacturing can be difficult due to the high GC-content of DNA sequences encoding for GRs and due to the presence of repeating DNA sequences that can lead to recombination.

[0348] As noted above, URPs can be designed to be highly repetitive at the amino acid level. As a result the URPs have very low information content and the risk of eliciting an immune reaction can be reduced.

[0349] Non-limiting examples of URPs containing repeating amino acids are: poly-glycine, poly-glycine acid, poly-aspartic acid, poly-serine, poly-threonine, (GX)n where G is glycine and X is serine, aspartic acid, glutamic acid, threonine, or proline and n is at least 20, (GGG)n (SEQ ID NO: 446) where X is serine, aspartic acid, glutamic acid, threonine, or proline and n is at least 13, (GGGG)n (SEQ ID NO: 447) where X is serine, aspartic acid, glutamic acid, threonine, or proline and n is at least 10, (GGGGGG)n (SEQ ID NO: 448) where X is serine, aspartic acid, glutamic acid, threonine, or proline and n is at least 15, and z is between 1 and 20.

[0350] The number of these repeats can be any number between 10 and 100. Products of the invention may contain URPs sequences that are semi-random sequences. Examples are semi-random sequences containing at least 30, 40, 50, 60 or 70% glycine in which the glycines are well dispersed and in which the total concentration of tryptophan, phenylalanine, tyrosine, valine, leucine, and isoleucine is less than 70, 60, 50, 40, 30, 20, or 10% when combined. A preferred semi-random URPs sequence contains at least 40% glycine and the total concentration of tryptophan, phenylalanine, tyrosine, valine, leucine, and isoleucine is less than 40%. A more preferred semi-random URPs sequence contains at least 50% glycine and the total concentration of tryptophan, phenylalanine, tyrosine, valine, leucine, and isoleucine is less than 50%. URPs sequences can be designed by combining the sequences of two or more shorter URPs sequences or fragments of URPs sequences. Such a combination allows one to better modulate the pharmaceutical properties of the product containing the URPs sequences and it allows one to reduce the repetitiveness of the DNA sequences encoding the URPs sequences, which can improve expression and reduce recombination of the URPs encoding sequences.

[0351] URPs sequences can be designed and selected to possess several of the following desired properties: a) high genetic stability of the coding sequences in the production host, b) high level of expression, c) low (predicted/calculated) immunogenicity, d) high stability in presence of serum proteases and/or other tissue proteases, e) large hydrodynamic radius under physiological conditions. One exemplary approach to obtain URPs sequences that meet multiple criteria is to construct a library of candidate sequences and to identify from the library the suitable subsequences. Libraries can comprise random and/or semi-random sequences. Of particular utility are codon libraries, which is a library of DNA molecules that contains multiple codons for the identical amino acid residue. Codon randomization can be applied to selected amino acid positions of a certain type or to most or all positions. True codon libraries encode only a single amino acid sequence, but they can easily be combined with amino acid libraries, which is a population of DNA molecules encoding a mixture of (related or unrelated) amino acids at the same residue position. Codon libraries allow the identification of genes that have relatively low repetitiveness at the DNA level but that encode highly repetitive amino acid sequences. This is useful because repetitive DNA sequences tend to recombine, leading to instability. One can also construct codon libraries that encode limited amino acid diversity. Such libraries allow introduction of a limited number of amino acids in some positions of the sequence while other positions allow for codon variation but all codons encode the same amino acid. One can then synthesize partially random oligonucleotides by incorporating mixtures of nucleotides at the same position during oligonucleotide synthesis. Such partially random oligonucleotides can be fused by overlap PCR or ligation-based approaches. In particular, one can utilize one or more semi-random oligonucleotides that encode glycine-rich sequences. These oligonucleotides can differ in length and sequences and codon usage. As a result, one obtains a library of candidate URPs sequences. Another method to generate libraries is to synthesize a starting sequence and subsequently
subject said sequence to partial randomization. This can be
done by cultivation of the gene encoding the URP sequences
in a mutator strain or by amplification of the encoding gene
under mutagenic conditions (Leung, D., et al. (1989) Tech-
nique, 1: 11-15). URP sequences with desirable properties
can be identified from libraries using a variety of methods.
Sequences that have a high degree of genetic stability can be
enriched by cultivating the library in a production host.
Sequences that are unstable will accumulate mutations,
which can be identified by DNA sequencing. Variants of URP
sequences that can be expressed at high level can be identified
by screening or selection using multiple protocols known to
someone skilled in the art. For instance one can cultivate
multiple isolates from a library and compare expression lev-
els. Expression levels can be measured by gel analysis, ana-
lytical chromatography, or various ELISA-based methods.
The determination of expression levels of individual sequence variants can be facilitated by fusing the library of
candidate URP sequences to sequence tags like myc-tag,
His-tag, HA-tag. Another approach is to fuse the library to an
enzyme or other reporter protein like green fluorescent pro-
gain. Of particular interest is the fusion of the library to a
selectable marker like beta-lactamase or kanamycin-acyl
transferase. One can use antibiotic selection to enrich for
variants with high level of expression and good genetic sta-
bility. Variants with good protease resistance can be identified
by screening for intact sequences after incubation with pro-
teases. An effective way to identify protease-resistant URP
sequences is bacterial plasmid display or related display meth-
ods. Multiple systems have been described where sequences
that undergo rapid proteolysis can be enriched by plasmid
display. These methods can be easily adopted to enrich for pro-
tease resistant sequences. For example, one can clone a
library of candidate URP sequences between an affinity tag
and the PI(1) protein of M13 plasmid. The library can then be
exposed to proteases or protease-containing biological
samples like blood or lysosomal preparations. Phage that
contain protease-resistant sequences can be captured after
protease treatment by binding to the affinity tag. Sequences
that resist degradation by lysosomal preparations are of par-
ticular interest because lysosomal degradation is a key step
during antigen presentation in dendritic and other antigen
presenting cells. Phage display can be utilized to identify
candidate URP sequences that do not bind to a particular
immunogenicity. One can immunize animals with a candi-
date URP sequence or with a library of URP sequences to
raise antibodies against the URP sequences in the library. The
resulting serum can then be used for phage panning to remove
or identify sequences that are recognized by antibodies in the
resulting immune serum. Other methods like bacterial dis-
play, yeast display, ribosomal display can be utilized to iden-
tify variants of URP sequences with desirable properties.
Another approach is the identification of URP sequences of
interest by mass spectrometry. For instance, one can incu-
bate a library of candidate URP sequences with a protease or
biological sample of interest and identify sequences that
resist degradation by mass spectrometry. In a similar
approach one can identify URP sequences that facilitate oral
uptake. One can feed a mixture of candidate URP sequences
to animals or humans and identify variants with the highest
transfer or uptake efficiency across some tissue barrier (ie
dermal, etc) by mass spectrometry. In a similar way, one can
identify URP sequences that favor other uptake mechanisms
like pulmonary, intranasal, rectal, transdermal delivery. One
can also identify URP sequences that favor cellular uptake or
URP sequences that resist cellular uptake.

[0352] URP sequences can be designed by combining URP
sequences or fragments of URP sequences that were designed
by any of the methods described above. In addition, one can
apply semi-random approaches to optimize sequences that
were designed based on the rules described above. Of par-
ticular interest is codon optimization with the goal of improv-
ing expression of the enhanced polypeptides and to improve
the genetic stability of the encoding gene in the production
hosts. Codon optimization is of particular importance for
URP sequences that are rich in glycine or that have very
repetitive amino acid sequences. Codon optimization can be
performed using computer programs (Gustafsson, C., et al.
mize ribosomal pausing (Coda Genomics Inc.). When
designing URP sequences one can consider a number of
properties. One can minimize the repetitiveness in the encod-
ing DNA sequences. In addition, one can avoid or minimize
the use of codons that are rarely used by the production host
(ie the AGG and AGA arginine codons and one Leucine
codon in E. coli) DNA sequences that have a high level of
glycine tend to have a high GC content that can lead to
instability or low expression levels. Thus, when possible it
is preferred to choose codons such that the GC-content of URP-
encoding sequence is suitable for the production organism
that will be used to manufacture the URP.

[0353] URP encoding genes can be made in one or more
steps, either fully synthetically or by synthesis combined with
enzymatic processes, such as restriction enzyme-mediated
cloning, PCR and overlap extension. URP accessory
polypeptides can be constructed such that the URP accessory
polypeptide-encoding gene has low repetitiveness while the
encoded amino acid sequence has a high degree of repetitiv-
ness. As a first step, one constructs a library of relatively short
URP sequences. This can be a pure codon library such that
each library member has the same amino acid sequence but
many different coding sequences are possible. To facilitate
the identification of well-expressing library members one can
construct the library as fusion to a reporter protein. Examples of
suitable reporter genes are green fluorescent protein,
luciferase, alkaline phosphatase, beta-galactosidase. By
screening one can identify short URP sequences that can be
expressed in high concentration in the host organism of
choice. Subsequently, one can generate a library of random
URP dimers and repeat the screen for high level of expres-
sion. Dimerization can be performed by ligation, overlap
extension or similar cloning techniques. This process of
dimerization and subsequent screening can be repeated mul-
tiple times until the resulting URP sequence has reached the
desired length. Optionally, one can sequence clones in the
library to eliminate isolates that contain undesirable
sequences. The initial library of short URP sequences can
allow some variation in amino acid sequence. For instance
one can randomize some codons such that a number of hydro-
phobic amino acids can occur in said position. During the
process of iterative multimerization one can screen library
members for other characteristics like solubility or protease
resistance in addition to a screen for high-level expression.
Instead of dimerizing URP sequences one can also generate
longer multimers. This allows one to faster increase the length of
URP accessory polypeptides.
Many URP sequences contain particular amino acids at high fraction. Such sequences can be difficult to produce by recombinant techniques as their coding genes can contain repetitive sequences that are subject to recombination. Furthermore, genes that contain particular codons at very high frequencies can limit expression as the respective loaded tRNAs in the production host become limiting. An example is the recombinant production of GRS. Glycine residues are encoded by 4 triplets, GGG, GCC, GGA, and GGT. As a result, genes encoding GRS tend to have high GC-content and tend to be particularly repetitive. An additional challenge can result from codon bias of the production host. In the case of E. coli, two glycine codons, GGA and GGG, are rarely used in highly expressed proteins. Thus codon optimization of the gene encoding URP sequences can be very desirable. One can optimize codon usage by employing computer programs that consider codon bias of the production hosts (Gustafsson, C., et al. (2004) Trends Biotechnol, 22: 346-53). As an alternative, one can construct codon libraries where all members of the library encode the same amino acid sequence but where codon usage is varied. Such libraries can be screened for highly expressing and genetically stable members which are particularly suitable for the large-scale production of URP-containing products.

Multivalent Unstructured Recombinant Proteins (MURPs):

As noted above, the subject URPs are particularly useful as accessory polypeptides for the modification of biologically active polypeptides. Accordingly, the present invention provides proteins comprising one or more subject URPs. Such proteins are termed herein Multivalent Unstructured Recombinant Proteins (MURPs).

To construct MURPs, one or more URP sequences can be fused to the N-terminus or C-terminus of a protein or inserted in the middle of the protein, e.g., into loops of a protein or in between modules of the biologically active polypeptide of interest, to give the resulting modified polypeptide improved properties relative to the unmodified protein. The combined length of URP sequences that are attached to a protein can be 40, 50, 60, 70, 80, 90, 100, 150, 200 or more amino acids.

The subject MURPs exhibit one or more improved properties as detailed below.

Improved Half-Life:

Adding a URP sequences to a biologically active polypeptide can improve many properties of that protein. In particular, adding a long URP sequence can significantly increase the serum half-life of the protein. Such URP typically contain amino acid sequences of at least about 40, 50, 60, 70, 80, 90, 100, 150, 200 or more amino acids.

The URP can be fragmented such that the resulting protein contains multiple URPs, or multiple fragments of URPs. Some or all of these individual URP sequences may be shorter than 40 amino acids as long as the combined length of all URP sequences in the resulting protein is at least 30 amino acids. Preferably, the resulting protein has a combined length of URP sequences exceeding 40, 50, 60, 70, 80, 90, 100, 150, 200 or more amino acids. In one aspect, the fused URPs can increase the hydrodynamic radius of a protein and thus reduces its clearance from the blood by the kidney. The increase in the hydrodynamic radius of the resulting fusion protein relative to the unmodified protein can be detected by ultracentrifugation, size exclusion chromatography, or light scattering.

Improved Tissue Selectivity:

Increasing the hydrodynamic radius can also lead to reduced penetration into tissues, which can be exploited to minimize side effects of a biologically active polypeptide. It is well documented that hydrophilic polymers have a tendency to accumulate selectively in tumor tissue which is caused by the enhanced permeability and retention (EPR) effect. The underlying cause of the EPR effect is the leaky nature of tumor vasculature (McDonald, D. M., et al. (2002) Cancer Res. 62: 5381-5) and the lack of lymphatic drainage in tumor tissues. Therefore, the selectivity of biologically active polypeptides for tumor tissues can be enhanced by adding hydrophilic polymers. As such, the therapeutic index of a given biologically active polypeptide can be increased via incorporating the subject URPs.

Protection from Degradation and Reduced Immunogenicity:

Adding URP sequences can significantly improve the protease resistance of a protein. URP sequences themselves can be designed to be protease resistant and by attaching them to a protein one can shield that protein from the access of degrading enzymes. URP sequences can be added to biologically active polypeptides with the goal of reducing undesirable interactions of the protein with other receptors or surfaces. To achieve this, it can be beneficial to add the URP sequences to the biologically active polypeptide in proximity to the site of the protein that makes such undesirable contacts. In particular, one can add URP sequences to biologically active polypeptides with the goal of reducing their interactions with any component of the immune system to prevent an immune response against the product of the invention. Adding a URP sequence to a biologically active polypeptide can reduce interaction with pre-existing antibodies or B-cell receptors. Furthermore, the addition of URP sequences can reduce the uptake and processing of the product of the invention by antigen presenting cells. Adding one or more URP sequence to a protein is a preferred way of reducing its immunogenicity as it will suppress an immune response in many species allowing one to predict the expected immunogenicity of a product in patients based on animal data. Such species independent testing of immunogenicity is not possible for approaches that are based on the identification and removal of human T cell epitopes or sequences comparison with human sequences.

Interruption of T Cell Epitopes:

URP sequences can be introduced into proteins in order to interrupt T cell epitopes. This is particularly useful for proteins that combine multiple separate functional modules. The formation of T cell epitopes requires that peptide fragments of a protein antigen bind to MHC. MHC molecules interact with a short segment of amino acids typically 9 contiguous residues of the presented peptides. The direct fusion of different binding modules in a protein molecule can lead to T cell epitopes that span two neighboring domains. By separating the functional modules by URP accessory polypeptides prevents the generation of such module-spanning T cell epitopes. The insertion of URP sequences between functional modules can also interfere with proteolytic processing in antigen presenting cells, which will lead to an additional reduction of immunogenicity.
Improved Solubility:

Functional modules of a protein can have limited solubility. In particular, binding modules tend to carry hydrophobic residues on their surface, which can limit their solubility and can lead to aggregation. By spacing or flanking such functional modules with URP accessory polypeptides one can improve the overall solubility of the resulting product. This is in particular true for URP accessory polypeptides that carry a significant percentage of hydrophilic or charged residues. By separating functional modules with soluble URP modules one can reduce intramolecular interactions between these functional modules.

Improved pH Profile and Homogeneity of Product Charge:

URP sequences can be designed to carry an excess of negative or positive charges. As a result they confer an electrostatic field to any fusion partner which can be utilized to shift the pH profile of an enzyme or a binding interaction. Furthermore, the electrostatic field of a charged URP sequence can increase the homogeneity of pKa values of surface charges of a protein product, which leads to sharpened pH profiles of ligand interactions and to sharpened separations by isoelectric focusing or chromatofocusing.

Improved Purification Properties Due to Sharper Product pKa:

Each amino acid in solution by itself has a single, fixed pKa, which is the pH at which its functional groups are half protonated. In a typical protein you have many types of residues and due to proximity and protein breathing effects, they also change each other’s effective pKas in variable ways. Because of this, at a wide range of pH conditions, typical proteins can adopt hundreds of differently ionized species, each with a different molecular weight and net charge, due to large numbers of combinations of charged and neutral amino acid residues. This is referred to as a broad ionization spectrum and makes the analysis (eg by mass spectrometry) and purification of such proteins more difficult.

PEG is unchanged and does not affect the ionization spectrum of the protein it is attached to, leaving it with a broad ionization spectrum. However, an URP with a high content of Gly and Glu in principle exist in only two states: neutral (—COOH) when the pH is below the pKa of Glutamate and negatively charged (—COO−) when the pH is above the pKa of Glutamate. URP accessory polypeptides can form a single, homogeneously ionized type of molecule and can yield a single mass in mass spectrometry.

Where desired, MURPs can be expressed as a fusion with an URP having a single type of charge (Glu) distributed at constant spacing through the URP accessory polypeptide. One may choose to incorporate 25-50 Glu residues per 20 kD of URP and all of these 25-50 residues would have very similar pKas.

In addition, adding 25-50 negative charges to a small protein like IFN, IGFH or GCSF (with only 20 charged residues) will increase the charge homogeneity of the product and sharpen its isoelectric point, which will be very close to the pKa of free glutamate.

The increase in the homogeneity of the charge of the protein population has favorable processing properties, such as in ion exchange, isoelectric focusing, mass spec, etc. compared to traditional PEylation.

Biologically Active Polypeptides

Suitable polypeptides that can be linked to the accessory polypeptide include all biologically active polypeptides exhibiting a binding specificity to a given target or another desired biological characteristic when used in vitro or in vivo. In particular, any protein of therapeutic or diagnostic interest can be modified by accessory polypeptides. Of particular interest are polypeptides for which modification of certain properties such as serum half-life or in vivo clearance is desirable. Such modifications can be envisioned in the context of therapeutic applications, for example if one desires to prolong the half-life of an administered protein therapeutic drug. Modification with accessory polypeptides could also show utility in diagnostic applications, for example to reduce non-specific binding of a diagnostic protein or imaging agent to other molecules.

Biologically active polypeptides can include, but are not limited to cytokines, chemokines, lymphokines, ligands, receptors, hormones, enzymes, antibodies and antibody fragments, and growth factors. Examples of receptors include TNF type I receptor, IL-1 receptor type II, IL-1 receptor antagonist, IL-4 receptor and any chemically or genetically modified soluble receptors. Examples of enzymes include activated protein C, factor VII, collagenase (e.g., marketed by Advance Biofactures Corporation under the name Santyl); agalasidase-beta (e.g., marketed by Genzyme under the name Fabrazyme); dornase-alpha (e.g., marketed by Genentech under the name Pulmozyme); alteplase (e.g., marketed by Genentech under the name Activase); pegylated-asparaginase (e.g., marketed by Enzon under the name Oncaspars); asparaginase (e.g., marketed by Merck under the name Elspar); and imiglucerase (e.g., marketed by Genzyme under the name Ceravease). Examples of specific polypeptides or proteins include, but are not limited to granulocyte macrophage colony stimulating factor (GM-CSF), granulocyte colony stimulating factor (G-CSF), macrophage colony stimulating factor (M-CSF), colony stimulating factor (CSF), interferon beta (IFN-β), interferon gamma (IFNγ), interferon gamma inducing factor 1 (IFNγ1), transforming growth factor beta (TGF-β), RANTES (regulated upon activation, normal T-cell expressed and presumably secreted), macrophage inflammatory proteins (e.g., MIP-1α and MIP-1β), Leishmania elongation initiating factor (LEIF), platelet derived growth factor (PDGF), tumor necrosis factor (TNF), growth factors, e.g., epidermal growth factor (EGF), vascular endothelial growth factor (VEGF), fibroblast growth factor (FGF), nerve growth factor (NGF), brain derived neurotrophic factor (BDNF), neurotrophin-2 (NT-2), neurotrophin-3 (NT-3), neurotrophin-4 (NT-4), neurotrophin-5 (NT-5), glial cell line derived neurotrophic factor (GDNF), ciliary neurotrophic factor (CNTF), TNF a type II receptor, erythropoietin (EPO), insulin and soluble glycoproteins e.g., gp120 and gp160 glycoproteins. The gp120 glycoprotein is a human immunodeficiency virus (HIV) envelope protein, and the gp160 glycoprotein is a known precursor to the gp120 glycoprotein.

By way of example, the following are several examples of biologically active polypeptides which may be suitable for modification according to the present invention.

In one embodiment, the biologically active polypeptide is GLP-1. GLP-1 is an approximately 30 amino acid polypeptide that is currently being investigated as a possible therapy for diabetes. GLP-1 suppresses glucagon release and increases insulin release. Both responses to GLP-1 result in a decrease in the serum concentration of glucose. GLP-1 is rapidly cleaved by dipeptidyl peptidase-4 in the body and as a result has an extremely short serum half-life, ~2 min. The successful development of GLP-1 as a therapeutic protein
requires formulations to increase the serum half life and delivery of the protein. This example describes the preparation of an rPEG-GLP-1 fusion protein based on rPEG (L.288) and its encapsulation in a polymer matrix to improve the half-life of GLP-1 for therapeutic use.

[0382] In another embodiment, the biologically active polypeptide is nesiritide, human B-type natriuretic peptide (hBNP). Nesiritide can be manufactured in E. coli using recombinant DNA technology. In a specific embodiment, nesiritide consists of a 32 amino acid sequence with a molecular weight of 3464g/mol.

[0383] In yet another embodiment, the biologically active polypeptide is secretin, which is a peptide hormone composed of an amino acid sequence identical to the naturally occurring porcine secretin consisting of 27 amino acids. After intravenous bolus administration of 0.4 mcg/kg of unmodified polypeptide, synthetic human secretin concentration rapidly declines to baseline secretin levels within 90 to 120 minutes. The elimination half-life of synthetic human secretin (not modified with ascorbyl polypeptide) is approximately 45 minutes.

[0384] In an alternative embodiment, the biologically active polypeptide is enfuvirtide, a linear 36-amino acid synthetic polypeptide which is an inhibitor of the fusion of HIV-1 with CD4+ cells.

[0385] In an additional embodiment, the biologically active polypeptide is bivalirudin, a specific and reversible direct thrombin inhibitor. A more specific embodiment provides for an biologically active polypeptide which is a synthetic, 20 amino acid peptide with a molecular weight of 1280 daltons.

[0386] Alternatively, Anthemiphilic Factor (AHF) may be selected as the biologically active polypeptide. AHF is a glycoprotein amenable to synthesis in a genetically engineered Chinese Hamster Ovary (CHO) cell line. It is also known as HEMOFIL M™ AHF (Baxter) or Anthemiphilic Factor (Human) [AHF (Human)]. The mean in vivo half-life of HEMOFIL M™ AHF is known to be 14.7±5.1 hours (n=61).

[0387] In another embodiment, erythropoietin is the biologically active polypeptide. Erythropoietin is a 165 amino acid glycoprotein manufactured by recombinant DNA technology and has the same biological effects as endogenous erythropoietin. In a specific embodiment, erythropoietin has a molecular weight of 30,400 daltons and is produced by mammalian cells into which the human erythropoietin gene has been introduced. The product may contain the identical amino acid sequence of isolated natural erythropoietin. In adult and pediatric patients with chronic renal failure, the elimination half-life of unmodified plasma erythropoietin after intravenous administration is known to range from 4 to 13 hours.

[0388] In still another embodiment, the biologically active polypeptide is Reteplase. Reteplase is a non-glycosylated delirion mutein of tissue plasminogen activator (tPA), comprising the kringle 2 and the protease domains of human tPA. Reteplase contains 355 of the 527 amino acids of native tPA (amino acids 1-3 and 176-527). The polypeptide may be produced by recombinant DNA technology in E. coli, and may be isolated as inactive inclusion bodies from E. coli, converted into its active form by an in vitro folding process and purified by chromatographic separation. The molecular weight of unmodified Reteplase is 39,571 daltons. Based on the measurement of thrombolytic activity, the effective half-life of unmodified Reteplase is known to be approximately 15 minutes.

[0389] A further embodiment provides for a biologically active polypeptide which is Anakinra, a recombinant, nonglycosylated form of the human interleukin-1 receptor antagonist (IL-1Ra). In one case, Anakinra consists of 155 amino acids and has a molecular weight of 17.3 kilodaltons. It may be produced by recombinant DNA technology using an E. coli bacterial expression system. In vivo half-life of unmodified Anakinra is known to range from 4 to 6 hours.

[0390] Becaplermin may also be selected as the biologically active polypeptide. Becaplermin is a recombinant human platelet-derived growth factor (rPDGF-BB) for topical administration. Becaplermin may be produced by recombinant DNA technology by insertion of the gene for the B chain of platelet derived growth factor (PDGF) into the yeast strain Saccharomyces cerevisiae. One form of Becaplermin has a molecular weight of approximately 25 kDa and is a homodimer composed of two identical polypeptide chains that are bound together by disulfide bonds.

[0391] The biologically active polypeptide may be Oprelvekin, which is a recombinant form of interleukin eleven (IL-11) that is produced in Escherichia coli (E. coli) by recombinant DNA technology. In one embodiment, the selected biologically active polypeptide has a molecular mass of approximately 19,000 daltons, and is non-glycosylated. The polypeptide is 177 amino acids in length and differs from the 178 amino acid length of native IL-11 only in lacking the amino-terminal proline residue, which is known not to result in measurable differences in bioactivity either in vitro or in vivo. The terminal half-life of unmodified Oprelvekin is known to be approximately 7 hrs.

[0392] Yet another embodiment provides for a biologically active polypeptide which is Glucagon, a polypeptide hormone identical to human glucagon that increases blood glucose and relaxes smooth muscles of the gastrointestinal tract. Glucagon may be synthesized in a special non-pathogenic laboratory strain of E. coli bacteria that have been genetically altered by the addition of the gene for glucagon. In a specific embodiment, glucagon is a single-chain polypeptide that contains 29 amino acid residues and has a molecular weight of 3,483. The in vivo half-life is known to be short, ranging from 8 to 18 minutes.

[0393] G-CSF may also be chosen as a biologically active polypeptide. Recombinant granulocye-colony stimulating factor or G-CSF is used following various chemotherapy treatments to stimulate the recovery of white blood cells. The reported half life of recombinant G-CSF is only 3.5 hours.

[0394] Alternatively, the biologically active polypeptide can be interferon alpha (IFN alpha). Chemically PEG-modified interferon-alpha 2a is clinically validated for the treatment of hepatitis C. This PEGylated protein requires weekly injection and slow release formulations with longer half-life are desirable.

[0395] Additional cellular proteins which may be modified with accessory polypeptides, or to which biologically active polypeptides may be targeted are VEGF, VEGF-R1, VEGF-R2, VEGF-R3, Her-1, Her-2, Her-3, EGF-1, EGF-2, EGF-3, Alphau3, cMet, ICOS, CD40L, LFA-1, c-Met, ICOS, LFA-1, IL-6. B7.1, B7.2, OX-40, IL-1b, TAC1, IgE, BAFF, or BLyS, TPO-R, CD19, CD20, CD22, CD33, CD28, IL-1-R1, TNFalpha, TRAIL-R1, Complement Receptor 1, FGFR, Octinicipant, Vitronecin, Ephrin A1-A5, Ephrin B1-B3, alpha-2-macro-


[097] Numerous human ion channels are targets of particular interest. Non-limiting examples include 5-hydroxytryptamine 3 receptor B subunit, 5-hydroxytryptamine 3 receptor precursor, 5-hydroxytryptamine receptor 3 subunit C, AAD14 protein, Acetylcholine receptor protein, alpha subunit precursor, Acetylcholine receptor protein, beta subunit precursor, Acetylcholine receptor protein, delta subunit precursor, Acetylcholine receptor protein, epsilon subunit precursor, Acetylcholine receptor protein, gamma subunit precursor, Acid sensing ion channel 3 splice variant b, Acid sensing ion channel 3 splice variant c, Acid sensing ion channel 4, ADP-ribose pyrophosphatase, mitochondrial precursor, Alpha-1-voltage-dependent, calcium channel, Amiloride-sensitive cation channel 1, neuronal, Amiloride-sensitive cation channel 2, neuronal Amiloride-sensitive cation channel 4, isoform 2, Amiloride-sensitive channel subunit, Amiloride-sensitive sodium channel alpha-subunit, Amiloride-sensitive sodium channel beta-subunit, Amiloride-sensitive sodium channel delta-subunit, Amiloride-sensitive sodium channel gamma-subunit, Annexin A7, Apical-like protein, ATR-sensitive inward rectifier potassium channel 1, ATP-sensitive inward rectifier potassium channel 10, ATP-sensitive inward rectifier potassium channel 15, ATP-sensitive inward rectifier potassium channel 18, Calcium channel alpha12.2 subunit, Calcium channel channel alpha12.2 subunit, Calcium channel channel alpha12.2 subunit, delta19 delta4 delta46 splice variant, Calcium-activated potassium channel alpha subunit 1, Calcium-activated...
potassium channel beta subunit 1, Calcium-activated potassium channel beta subunit 2, Calcium-activated potassium channel subunit 3, Calcium-dependent chloride channel-1, Cation channel TRPM4, cDNA FLJ90453 fs, clone NT12R3P001542, highly similar to Potassium channel tetramerisation domain containing 6, cDNA FLJ90663 fs, clone PLACE1005301, highly similar to Chloride intracellular channel protein 5, CGMP-gated cation channel beta subunit, Chloride channel protein, Chloride channel protein 2, Chloride channel protein 3, Chloride channel protein 4, Chloride channel protein 5, Chloride channel protein 6, Chloride channel protein C1C-Ka, Chloride channel protein C1C-Kb, Chloride channel protein, skeletal muscle, Chloride intracellular channel 6, Chloride intracellular channel protein 3, Chloride intracellular channel protein 4, Chloride intracellular channel protein 5, CHRNA3 protein, Clcn3e protein, CLCNKB protein, CNGA4 protein, Cullin-5, Cyclic GMP gated potassium channel, Cyclic-nucleotide-gated cation channel 4, Cyclic-nucleotide-gated cation channel alpha 3, Cyclic-nucleotide-gated cation channel beta 3, Cyclic-nucleotide-gated olfactory channel, Cystic fibrosis transmembrane conductance regulator, Cytochrome B-245 heavy chain, Dihydropyridine-sensitive 1-type, calcium channel alpha-2/delta subunits precursor, FXDY domain-containing ion transport regulator 3 precursor, FXDY domain-containing ion transport regulator 5 precursor, FXDY domain-containing ion transport regulator 6 precursor, FXDY domain-containing ion transport regulator 7, FXDY domain-containing ion transport regulator 8 precursor, G protein-activated inward rectifier potassium channel 1, G protein-activated inward rectifier potassium channel 2, G protein-activated inward rectifier potassium channel 3, G protein-activated inward rectifier potassium channel 4, Gamma-aminobutyric-acid receptor alpha-1 subunit precursor, Gamma-aminobutyric-acid receptor alpha-2 subunit precursor, Gamma-aminobutyric-acid receptor alpha-3 subunit precursor, Gamma-aminobutyric-acid receptor alpha-4 subunit precursor, Gamma-aminobutyric-acid receptor alpha-5 subunit precursor, Gamma-aminobutyric-acid receptor alpha-6 subunit precursor, Gamma-aminobutyric-acid receptor beta-1 subunit precursor, Gamma-aminobutyric-acid receptor beta-2 subunit precursor, Gamma-aminobutyric-acid receptor beta-3 subunit precursor, Gamma-aminobutyric-acid receptor delta subunit precursor, Gamma-aminobutyric-acid receptor epsilon subunit precursor, Gamma-aminobutyric-acid receptor gamma-1 subunit precursor, Gamma-aminobutyric-acid receptor gamma-3 subunit precursor, Gamma-aminobutyric-acid receptor gamma-pi subunit precursor, Gamma-aminobutyric-acid receptor rho-1 subunit precursor, Gamma-aminobutyric-acid receptor rho-2 subunit precursor, Gamma-aminobutyric-acid receptor theta subunit precursor, Gl1ur6 kainate receptor, Glutamate receptor 1 precursor, Glutamate receptor 2 precursor, Glutamate receptor 3 precursor, Glutamate receptor 4 precursor, Glutamate receptor 7, Glutamate receptor B, Glutamate receptor delta-1 subunit precursor, Glutamate receptor, ionotropic kainate 1 precursor, Glutamate receptor, ionotropic kainate 2 precursor, Glutamate receptor, ionotropic kainate 3 precursor, Glutamate receptor, ionotropic kainate 4 precursor, Glutamate receptor, ionotropic kainate 5 precursor, Glutamate [NMDA] receptor subunit 3A precursor, Glutamate [NMDA] receptor subunit 3B precursor, Glutamate [NMDA] receptor subunit epsilon 1 precursor, Glutamate [NMDA] receptor subunit epsilon 2 precursor, Glutamate [NMDA] receptor subunit epsilon 4 precursor, Glutamate [NMDA] receptor subunit zeta 1 precursor, Glycine receptor alpha-1 chain precursor, Glycine receptor alpha-2 chain precursor, Glycine receptor alpha-3 chain precursor, Glycine receptor beta chain precursor, HACA ribonucleoprotein complex subunit 1, High affinity immunoglobulin epsilon receptor beta-subunit, Hypothetical protein DKFZp1310334, Hypothetical protein DKFZp761M1724, Hypothetical protein FLJ12242, Hypothetical protein FLJ14389, Hypothetical protein FLJ14798, Hypothetical protein FLJ14995, Hypothetical protein FLJ16180, Hypothetical protein FLJ16802, Hypothetical protein FLJ32069, Hypothetical protein FLJ37401, Hypothetical protein FLJ38750, Hypothetical protein FLJ40162, Hypothetical protein FLJ41415, Hypothetical protein FLJ90576, Hypothetical protein FLJ90590, Hypothetical protein FLJ90622, Hypothetical protein KCTD15, Hypothetical protein MGC15619, Inositol 1,4,5-trisphosphate receptor type 1, Inositol 1,4,5-trisphosphate receptor type 2, Inositol 1,4,5-trisphosphate receptor type 3, Intermediate conductance calcium-activated potassium channel protein 4, Inward rectifier potassium channel 13, Inward rectifier potassium channel 16, Inward rectifier potassium channel 4, Inward rectifying K(+) channel negative regulator Kir2.2v, Kainate receptor subunit KA2a, KCNHS protein, KCTD17 protein, KCTD22 protein, Keratinocyes associated transmembrane protein 1, Kv channel-interacting protein 4, Melastatin 1, Membrane protein MCL1, MGC15619 protein, Mucolipin-1, Mucolipin-2, Mucolipin-3, Multidrug resistance-associated protein 4, N-methyl-D-aspartate receptor 2C subunit precursor, NADPH oxidase homolog 1, Nav1.5, Neuronal acetycholine receptor protein, alpha-10 subunit precursor, Neuronal acetycholine receptor protein, alpha-2 subunit precursor, Neuronal acetycholine receptor protein, alpha-3 subunit precursor, Neuronal acetycholine receptor protein, alpha-4 subunit precursor, Neuronal acetycholine receptor protein, alpha-5 subunit precursor, Neuronal acetycholine receptor protein, alpha-6 subunit precursor, Neuronal acetycholine receptor protein, alpha-7 subunit precursor, Neuronal acetycholine receptor protein, alpha-8 subunit precursor, Neuronal acetycholine receptor protein, alpha-9 subunit precursor, Neuronal acetycholine receptor protein, beta-2 subunit precursor, Neuronal acetycholine receptor protein, beta-3 subunit precursor, Neuronal acetycholine receptor protein, beta-4 subunit precursor, Neuronal voltage-dependent calcium channel alpha 2D subunit, P2X purinoreceptor 1, P2X purinoreceptor 2, P2X purinoreceptor 3, P2X purinoreceptor 4, P2X purinoreceptor 5, P2X purinoreceptor 6, P2X purinoreceptor 7, Pancreatic potassium channel TALK-1b, Pancreatic potassium channel TALK-1d, Phospholemman precursor, Plasminogen, Polycystic kidney disease 2 related protein, Polycystic kidney disease 2-like 1 protein, Polycystic kidney disease 2-like 2 protein, Polycystic kidney disease and receptor for egg jelly related protein precursor, Polycystin-2, Potassium channel regulator, Potassium channel subfamily K member 1, Potassium channel subfamily K member 10, Potassium channel subfamily K member 12, Potassium channel subfamily K member 13, Potassium channel subfamily K member 15, Potassium channel subfamily K member 16, Potassium channel subfamily K member 17, Potassium channel subfamily K member 2, Potassium channel subfamily K member 3, Potassium channel subfamily K member 4, Potassium channel subfamily K member 5, Potassium channel subfamily K member 6, Potassium channel subfamily K member 7, Potassium channel subfamily K member 9, Potassium channel tetramerisation domain con-
taining 3, Potassium channel tetramerisation domain containing protein 12, Potassium channel tetramerisation domain containing protein 14, Potassium channel tetramerisation domain containing protein 2, Potassium channel tetramerisation domain containing protein 4, Potassium channel tetramerisation domain containing protein 5, Potassium channel tetramerisation domain containing 10, Potassium channel tetramerisation domain containing protein 13, Potassium channel tetramerisation domain-containing 1, Potassium voltage-gated channel subfamily A member 1, Potassium voltage-gated channel subfamily A member 2, Potassium voltage-gated channel subfamily A member 4, Potassium voltage-gated channel subfamily A member 5, Potassium voltage-gated channel subfamily A member 6, Potassium voltage-gated channel subfamily B member 1, Potassium voltage-gated channel subfamily B member 2, Potassium voltage-gated channel subfamily C member 1, Potassium voltage-gated channel subfamily C member 3, Potassium voltage-gated channel subfamily C member 4, Potassium voltage-gated channel subfamily D member 1, Potassium voltage-gated channel subfamily D member 2, Potassium voltage-gated channel subfamily D member 3, Potassium voltage-gated channel subfamily E member 1, Potassium voltage-gated channel subfamily E member 2, Potassium voltage-gated channel subfamily E member 3, Potassium voltage-gated channel subfamily E member 4, Potassium voltage-gated channel subfamily F member 1, Potassium voltage-gated channel subfamily G member 1, Potassium voltage-gated channel subfamily G member 2, Potassium voltage-gated channel subfamily G member 3, Potassium voltage-gated channel subfamily G member 4, Potassium voltage-gated channel subfamily H member 1, Potassium voltage-gated channel subfamily H member 2, Potassium voltage-gated channel subfamily H member 3, Potassium voltage-gated channel subfamily H member 4, Potassium voltage-gated channel subfamily H member 5, Potassium voltage-gated channel subfamily H member 6, Potassium voltage-gated channel subfamily H member 7, Potassium voltage-gated channel subfamily H member 8, Potassium voltage-gated channel subfamily KQT member 1, Potassium voltage-gated channel subfamily KQT member 2, Potassium voltage-gated channel subfamily KQT member 3, Potassium voltage-gated channel subfamily KQT member 4, Potassium voltage-gated channel subfamily KQT member 5, Potassium voltage-gated channel subfamily S member 1, Potassium voltage-gated channel subfamily S member 2, Potassium voltage-gated channel subfamily S member 3, Potassium voltage-gated channel subfamily S member 4, Potassium voltage-gated channel subfamily S member 5, Potassium voltage-gated channel subfamily S member 6, Potassium voltage-gated channel subfamily S member 7, Potassium voltage-gated channel subfamily S member 8, Potassium voltage-gated channel subfamily S member 9, Putative transmembrane protein 3, Putative potassium channel subfamily H member 7, isoform 2, Putative potassium/sodium hyperpolarization-activated cyclic nucleotide-gated channel 1, Potassium/sodium hyperpolarization-activated cyclic nucleotide-gated channel 2, Putative potassium/sodium hyperpolarization-activated cyclic nucleotide-gated channel 3, Potassium/sodium hyperpolarization-activated cyclic nucleotide-gated channel 4, Probable mitochondrial import receptor subunit TOM40 homolog, Purinergic receptor P2X5, isoform A, Putative 4 repeat voltage-gated ion channel, Putative chloride channel protein 7, Putative G1R6 kainate receptor, Putative ion channel protein CATSPER2 variant 1, Putative ion channel protein CATSPER2 variant 2, Putative ion channel protein CATSPER2 variant 3, Putative regulator of potassium channels protein variant 1, Putative tyrosine-protein phosphatase TPTE, Rybomine receptor 1, Rybomine receptor 2, Rybomine receptor 3, SDF KBP1 binding protein 1, Short transient receptor potential channel 1, Short transient receptor potential channel 4, Short transient receptor potential channel 5, Short transient receptor potential channel 6, Short transient receptor potential channel 7, Small conductance calcium-activated potassium channel protein 1, Small conductance calcium-activated potassium channel protein 2, isoform b, Small conductance calcium-activated potassium channel protein 3, isoform b, Small-ducte-activated calcium-activated potassium channel SK2, Small-conductance calcium-activated potassium channel SK3, Sodium channel channel beta-1 subunit precursor, Sodium channel protein type II alpha subunit, Sodium channel protein type III alpha subunit, Sodium channel protein type IV alpha subunit, Sodium channel protein type IX alpha subunit, Sodium channel protein type V alpha subunit, Sodium channel protein type VII alpha subunit, Sodium channel protein type VIII alpha subunit, Sodium channel protein type X alpha subunit, Sodium channel protein type XI alpha subunit, Sodium- and chloride-activated ATP-sensitive potassium channel, Sodium/potassium-transporting ATPase gamma chain, Sperm-associated cation channel 1, Sperm-associated cation channel 2, isoform 4, Syntaxin-1B1, Transient receptor potential cation channel subfamily A member 1, Transient receptor potential cation channel subfamily A member 2, Transient receptor potential cation channel subfamily M member 2, Transient receptor potential cation channel subfamily M member 3, Transient receptor potential cation channel subfamily M member 6, Transient receptor potential cation channel subfamily M member 7, Transient receptor potential cation channel subfamily V member 1, Transient receptor potential cation channel subfamily V member 2, Transient receptor potential cation channel subfamily V member 3, Transient receptor potential cation channel subfamily V member 4, Transient receptor potential cation channel subfamily V member 5, Transient receptor potential cation channel subfamily V member 6, Transient receptor potential channel 4 epsilon splice variant, Transient receptor potential channel 4 zeta splice variant, Transient receptor potential channel 7 gamma splice variant, Tumor necrosis factor, alpha-induced protein 1, endothelial, Two-pore calcium channel protein 2, NDAC4 protein, Voltage gated potassium channel Kv3.2b, Voltage gated sodium channel beta1B subunit, Voltage-dependent anion channel, Voltage-dependent anion channel 2, Voltage-dependent anion-selective channel protein 1, Voltage-dependent anion-selective channel protein 2, Voltage-dependent anion-selective channel protein 3, Voltage-dependent calcium channel gamma-1 subunit, Voltage-dependent calcium channel gamma-2 subunit, Voltage-dependent calcium channel gamma-3 subunit, Voltage-dependent calcium channel gamma-4 subunit, Voltage-dependent calcium channel gamma-5 subunit, Voltage-dependent calcium channel gamma-6 subunit, Voltage-dependent calcium channel gamma-7 subunit, Voltage-dependent calcium channel gamma-8 subunit, Voltage-dependent L-type calcium channel alpha-1C subunit, Voltage-dependent L-type calcium channel alpha-1D subunit, Voltage-dependent L-type calcium channel alpha-1S subunit, Voltage-dependent L-type calcium channel beta-1 subunit, Voltage-dependent L-type calcium channel beta-2 subunit, Voltage-dependent L-type calcium channel beta-3 subunit, Voltage-dependent L-type calcium channel beta-4 subunit, Voltage-dependent N-type calcium channel alpha-1B subunit, Voltage-dependent P/Q-type calcium channel alpha-1A subunit, Voltage-dependent R-type calcium channel alpha-1E subunit, Voltage-dependent T-type calcium
channel alpha-1G subunit, Voltage-dependent T-type calcium channel alpha-1H subunit, Voltage-dependent T-type calcium channel alpha-1I subunit, Voltage-gated L-type calcium channel alpha-1 subunit, Voltage-gated potassium channel beta-1 subunit, Voltage-gated potassium channel beta-2 subunit, Voltage-gated potassium channel beta-3 subunit, Voltage-gated potassium channel KCNA7. The Nav1.6 family of human voltage-gated sodium channels also a particularly promising target. This family includes, for example, channels Nav1.6 and Nav1.8.

[0398] Many of the microproteins used as scaffolds in this application have native activity against G-Protein Coupled Receptors (GPCRs) and offer ideal starting points to create novel GPCR modulators (including agonists, antagonists and modulators of any property of the GPCR). Exemplary GPCRs include but are not limited to Class A Rhodopsin like receptors such as Muscarinic (Mus.) acetylcholine Vertebrate type I, Mus. acetylcholine Vertebrate type 2, Mus. acetylcholine Vertebrate type 3, Mus. Acetylcholine Vertebrate type 4; Adrenoceptors (Alpha Adrenoceptors type I, Alpha Adrenoceptors type II, Beta Adrenoceptors type I, Beta Adrenoceptors type II, Dopamine Vertebrate type I, Dopamine Vertebrate type 2, Dopamine Vertebrate type 3, Dopamine Vertebrate type 4, Histamine type I, Histamine type 2, Histamine type 3, Histamine type 4, Serotonin type 1, Serotonin type 2, Serotonin type 3, Serotonin type 4, Serotonin type 5, Serotonin type 6, Serotonin type 7, Serotonin type 8, other Serotonin types, Tracer amine, Angiotensin type I, Angiotensin type II, Bombesin, Bradykinin, C'N ana-
phyatocephal, Fmet-leu-phe, IAP like, Interleukin-8 type A, Interleukin-8 type B, Interleukin-8 type others, C-3-C Chemokine type 1 through type 11 and other types, C-3-X-C Chemokine (types 2 through 6 and others), C-X-C Chemokine, Cholecystokinin CCK, CCK type A, CCK type B, CCK others, Endothelin, Melanocortin (Melanocyte stimulating hormone, Adrenocorticotropic hormone, Melanocortin hormone), Dlfy antigen, Prolactin-releasing peptide (GPR10), Neuropeptide Y (type I through 7), Neuropeptide Y, Neuropeptide Y other, Neurotensin, Opioid type (D, K, M, X), Somatostatin (type I through 5), Thyryokin (Substance P(NK1), Substance K (NK2), Neurokinin K (NK3), Thyryokin like 1, Thyryokin like 2, Vasopressin/vasotonin (type I through 2), Vasotonin, Oxytocin/nesotocin, Conopressin, Galanin like, Proteinase-activated like, Orexin & neuropeptides P, QRFP, Chemokine receptor-like, Neuro-
medin U like (Neuromedin U, PRXamide), hormone protein (Follicle stimulating hormone, Luteinizing hormone, Thyroid secretion, Gonadotropic hormone type I, Gonadotropic hormone type II), (Rhodopsin, Rhodopsin Vertebrate type 1-5, Rhodopsin Vertebrate type 5, Rhodopsin Arthropod, Rhodopsin Arthropod type I, Rhodopsin Arthropod type 2, Rhodopsin Arthropod type 3, Rhodopsin Mollusc, Rhodopsin, Offactory (olfactory II fam 1 through 13), Prostaglandin (prostaglandin E2 subtype EP1, Prostaglandin E2D2 subtype EP2, Prostaglandin E2 subtype EP3, Prostaglandin E2 subtype EP4, Prostaglandin E2-F alpha, Prostacyclin, Throm-
boxane, Adenosine type 1 through 3, Purinceptors, Purinceptors P2Y2-R2Y5, 6, 11 GPR91, Purinceptors P2Y5, 8, 9, 10 GPR35,92,174, Purinceptors P2Y2-R12-GP487 (UDP-Glucose), Cannabinoid, Platelet activating factor, Gonadotropin-releasing hormone, Gonadotropin-releasing hormone type I, Gonadotropin-releasing hormone type II, Adipokine hormone like, Corazonin, Thyrotropin-releasing hormone & Secretagogue, Thyrotropin-releasing hor-
mine, Growth hormone secretagogue, Growth hormone secretagogue like, Ecdysis-triggering hormone (EThrH), Melatonin, Lysophospholipid & LPA (EDG), Sphingosine 1-phosphate Edg-1, Sphingosine 1-phosphate Edg-2, Sphingosine 1-phosphate Edg-3, Sphingosine 1-phosphate Edg-4, Sphingosine 1-phosphate Edg-5, Sphingosine 1-phosphate Edg-6, Sphingosine 1-phosphate Edg-7, Sphingosine 1-phosphate Edg-8, Edg Other Leukotriene B4 receptor, Leukotriene B4 receptor BLT1, Leukotriene B4 receptor BLT2, Class A Orphan/other, Putative neurotransmitters, SREB, Mas proto-phenycyrene & Mas-related (MRGs), GPR45 like, Cysteinyl leukotriene, G-protein coupled bile acid receptor, Free fatty acid receptor (GPR40, GPR41, GPR43), Class B Secretin like, Calcitonin, Corticotropin releasing factor, Gastric inhibitory peptide, Glucagon, Growth hormone-releasing hormone, Parathyroid hormone, PACAP, Secretin, VIP, Opioid intestinal poly peptide, Latrophilin, Latrophilin type 1, Latrophilin type 2, Latrophilin type 3, ET-1 receptors, Brain-specific angiogenesis inhibitor (BAI), Methuselah-like proteins (MTH), Cadherin EGF LAG (CELSR), Very large G-protein coupled receptor, Class C Metabotropic glutamate/phero-
mone, Metabotropic glutamate group I through III, Calcium-sensing like, Extracellular calcium-sensing, Pheromone, calcium-sensing like other, Putative pheromone receptors, GABA-B, GAB A-B subtype 1, GAB A-B subtype 2, GAB A-B like, Orphan GPRC3, Orphan GPCR6, Bride of sevenless proteins (BOSS), Taste receptors (TIR), Class D Fungal pheromone, Fungal pheromone A-Factor like (STE2, STE3), Fungal pheromone B like (BAR, BBR, RCB, PRA), Class E-CAMP receptors, Ocular albinism proteins, Frizzled, Smoothened family, frizzled Group A (Fz1 &2 & & &5 &7 &9), frizzled Group B (Fz 3 & 6), frizzled Group C (other), Vomeronasal receptors, Nematode chemoreceptors, Insect odorant receptors, and Class Z Archaeal/bacterial/fungal opsins.

[0399] Of particular utility is the fusion of accessory sequences to any of the following active poly peptides: BOTOX, Myobloc, Neurobloc, Dysport (or other serotypes of botulinum neurotoxins), algalucosidase alfa, daptomycin, YH-16, choriogonadotropin alfa, filgrastim, cetuximab, interleukin-2, aldesleukin, teceleukin, denileukin difitox, interferon alfa-n3 (injection), interferon alfa-n1, DL-8234, interferon, Sontury (gamma-1 a), interferon gamma, thyrosin alpha 1, tasonerin, DigiFab, Viperin Tab, EchiTab, CroFab, nesiritide, abactepe, alefacept, Rifab, eptofetalin bal, teri-
paradine (osteoporosis), calcitonin injectable (bone disease), calcitonin (nasal, osteoporosis), etanercept, hemoglobin glucamer 250 (bovine), drotrecogin alfa, collagenase, carper-
tide, recombinant human epidural growth factor (topical gel, wound healing), DWP-401, durepocita alfa, nesiritide, omega, epeotin beta, epeotin alfa, desirudin, lepirudin, bival-
irinidin, nonocog alfa, Mononine, eptocog alfa (activated), recombinant Factor VIII+VF, Recombine, recombinant Factor VIII, Factor VIII (recombinant), Alphanate, octocog alfa, Factor VIII, palifermin, Indinavir, tenofovir, atelapte, pamitaple, retaple, nateple, montaple, fel-
itorpin alfa, rfSH, hFSH, micaflungin, pegfilgrastim, lenogrostam, nartogristam, sermorelin, glucagon, exendate, pramlintide, imiglucerase, galsulfase, Leucotropin, molgam-
mostin, triporelin acetate, histrelin (subcutaneous implant, Hydron), deslerein, histrelin, nafarelin, leuprolide sustained release depot (ATRIGEL), leuprolide implant (DUROS), goserelin, somatropin, Eutropin, KP-102 program, somatropin, somatropin, mecasermin (growth failure), enivircide, Org-33408, insulin glargine, insulin glulisine, insulin (in-
INNO-105, oral teriparadine (eligen), GEM-OS1, AC-162352, PRX-302, LF-n.p2 fusion vaccine (Therapore), EP-1043, S. pneumoniae pediatric vaccine, malaria vaccine, Neisseria meningitidis Group B vaccine, neonatal group B streptococcal vaccine, anthrax vaccine, HCV vaccine (gpE1+gpE2+MF-59), otitis media therapy, HCV vaccine (core antigen+ICOMATRIX), hPTH(1-34) (transdermal, ViaDerm), 768974, SYN-101, PGN-0052, avasimibe, BIM-23190, tuberculosis vaccine, multi-epitope tyrosinase peptide, cancer vaccine, enkastim, APC-8024, G1-5005, ACC-001, TIS-CD3, vascular-targeted TNF (solid tumors), desmopressin (buccal controlled-release), oncorept, TP-9201.

[0400] Non-Repetitive URP s (nrURPs)

[0401] The present invention also encompasses non-repetitive URP s (nrURPs). nrURPs are amino acid sequences that are composed mainly of small hydrophilic amino acids and that have a low tendency to form secondary structure in vivo. nrURPs possess the characteristics of URP s including the lack of well defined secondary and tertiary structures under physiological conditions, contributing to their conformational flexibility; high degree of protease resistance; ability to increase the half-life and/or solubility of a biologically active polypeptide upon incorporation of the URP sequences into the biologically active polypeptide. A particular property of nrURPs is their low degree of internal repetitiveness. nrURPs comprise multiple different peptide subsequences. These subsequences have URP-like amino acid composition but differ from each other in their amino acid sequence and length.

[0402] nrURPs tend to have improved solubility as compared to repetitive URP s (rURPs) with similar amino acid composition. In general, repetitive amino acid sequences have a tendency to aggregate as exemplified by natural repetitive sequences such as collagens and leucine zippers. Repetitive sequences can form higher order structures such that identical subsequences from similar contacts resulting in crystalline or pseudocrystalline structures. nrURPs have a much lower tendency to form such pseudo-crystalline structures as they contain multiple different subsequences that prevent the formation of any repetitive higher order structure. The low tendency of non-repetitive sequences to aggregate enables the design URP s with a relatively low frequency of charged amino acids that would be likely to aggregate in repetitive URP s. The low aggregation tendency of nrURPs facilitates the formulation of nrURP-comprising pharmaceutical preparations in particular enabling preparations containing extremely high drug concentrations exceeding 100 mg/ml.

[0403] (a) nrURPs have low immunogenicity

[0404] The interactions of a repetitive and a non-repetitive URP sequence with B cells that recognize epitopes in said sequences are compared and illustrated in FIG. 74. A rURP is recognized by few B cells in an organism as it contains a relatively small number of different epitopes. However, a rURP can form multivalent contacts with these few B cells and as a consequence it can stimulate B cell proliferation as illustrated in FIG. 74a. In contrast, a nrURP can make contacts with many different B cells as it contains many different epitopes. However, each individual B cell can only make one or a small number of contacts with an individual nrURP due to the lack of repetitiveness as illustrated in FIG. 74b. As a result, nrURPs have a much lower tendency to stimulate proliferation of B cells and thus an immune response.

[0405] An additional advantage of nrURPs relative to rURPs is that nrURPs form weaker contacts with antibodies relative to rURPs. Antibodies are multivalent molecules. For instance, IgGs have two identical binding sites and IgMs contain 10 identical binding sites. Thus antibodies against repetitive sequences can form multivalent contacts with such repetitive sequences with high avidity, which can affect the potency and/or elimination of such repetitive sequences. In contrast, antibodies against nrURPs form mainly monovalent interactions with antibodies as said nrURPs contain few repeats of each epitope.

[0406] (b) Detection of repetitiveness

[0407] The repetitiveness of a gene can be measured by computer algorithms. An example is illustrated in FIG. 75. Based on the query sequence, a pair wise comparison of all subsequences of a particular length can be performed. These subsequences can be compared for identity or homology. The example in FIG. 75 compares subsequences of 4 amino acids for identity. In the example, most 4-mer subsequences occur once in the query sequence and 3 4-mer subsequences occur twice. The repetitiveness in a gene can be averaged. The length of the subsequences can be adjusted. The length of the subsequences reflects the length of sequence epitopes that can be recognized by the immune system. Thus analysis of subsequences of 4-15 amino acids may be most useful.

[0408] (c) Design of nrURP sequences

[0409] Genes encoding nrURPs can be assembled from oligonucleotides using standard techniques of gene synthesis. The gene design can be performed using algorithms that optimize codon usage and amino acid composition. In addition, one can avoid amino acid sequences that are protease sensitive or that are known to contain epitopes that can be easily recognized by the human immune system. Computer algorithms can be applied during sequence design to minimize the repetitiveness of the resulting amino acid sequences. One can evaluate the repetitiveness of large numbers of gene designs that match preset criteria such as amino acid composition, codon usage, avoidance of protease sensitive subsequence, avoidance of epitopes, and chose the least repetitive sequences for synthesis and subsequent evaluation.

[0410] An alternative approach to the design of nrURP genes is to analyze the sequences of existing collections of nrURPs that show high level expression, low aggregation tendency, high solubility, and good resistance to proteases. A computer algorithm can design nrURP sequences based on such pre-existing nrURP sequences by re-assembly of sequence fragments as illustrated in FIG. 76. The algorithm generates a collection of subsequences from these nrURP sequences and then evaluates multiple ways to assemble nrURP sequences from such subsequences. These assembled sequences can be evaluated for repetitiveness to identify nrURP sequences that are only composed of subsequences of previously identified nrURPs but that have reduced repetitiveness compared to all parent nrURPs.

[0411] (d) Construction of nrURP sequences from libraries

[0412] nrURP-encoding genes can be assembled from libraries of short URP segments as illustrated in FIG. 77.

[0413] One can first generate large libraries of URP segments. Such libraries can be assembled from partially randomized oligonucleotides. The randomization scheme can be optimized to control amino acid choices for each position as well as codon usage and sequence length. In one embodiment, the library of URP segments is cloned into an expres-
sion vector. In another embodiment, the library of URP segments is cloned into an expression vector fused to an indicator gene like GFP. Subsequently, one can screen library members for a number of properties such as level of expression, protease stability, binding to serum proteins. One can screen URP segments for binding to antisera to eliminate segment with high affinity for said serum. In particular one can screen library members to identify and avoid binding to antisera with reactivity to URP sequences. The amino acid sequence of the library members can be determined to identify segments that have a particularly desirable amino acid composition, segment length, or to identify segments that have a low frequency of internal repeats. Subsequently, nURP sequences may be assembled from the collections of URP segments by random dimerization or multimerization. Dimerization or multimerization can be achieved by ligation or PCR assembly. This process results in a library of nURP sequences that can be evaluated for a number of properties to identify the nURP sequences with the most desirable properties. The process of dimerization or multimerization can be repeated to further increase the length of nURP sequences.

[0414] Design of Crosslinked Accessory Polypeptides

[0415] The present invention also relates to polypeptides with enhanced properties (such as increased hydrodynamic radius or extended serum half-life) comprising crosslinked accessory polypeptides. A crosslinked accessory polypeptide can be generated by conjugating one or more non-cross-linking components and one or more cross-linking components.

[0416] The advantage of this approach is that one can use an accessory polypeptide of moderate length, which is highly expressed, to efficiently generate larger molecules with desired properties. For example, using chemical coupling one can create a molecule comprising five 200 amino acid long units much more efficiently than a single 1000 amino acid long polypeptide expressed as a single protein.

[0417] Any number of non-crosslinking components, such as 2, 3, 4, 5, 6, 7, 8, 9, 10 or more components can be linked together. These components can be identical or of 2, 3, 4, 5, 6, 7, 8, 9 or 10 or more different kinds. In a preferred embodiment, each component has a determined binding specificity, which can be the same for each component or of 2, 3, 4, 5, 6, 7, 8, 9, or 10 or more different types. The sequences of the non-crosslinking components can also be the same or may comprise 1-10 different sequences.

[0418] A preferred embodiment of the present invention provides for reacting 1, 2, 3, 4, 5, 6, 7, 8 or more copies of a monoactive non-crosslinking component with one copy of a multireactive cross-linking component, which optionally contains polyethylene glycol, an accessory polypeptide or another water-soluble polymer, resulting in a pre-defined polymer containing exactly (for example) four copies of the non-crosslinking component, each copy being linked to the cross-linking agent. The non-crosslinking component may optionally comprise a domain with binding specificity.

[0419] A variety of linkage chemistries can be used for conjugation. In a preferred embodiment, standard amino-carboxyl coupling, and especially linking via the amino group of a lysine group or of the N-terminus, or linking via the carboxyl group of glutamate or of the C-terminus, is especially useful for cross-linking of crosslinked accessory polypeptides.

[0420] In some embodiments, the cross-linking component can be a synthetic polypeptide. For example, such a polypeptide may comprise 5 carboxy residues (i.e. 4 glutamates plus the C-terminal carboxy), optionally spaced by sequences inserted between the carboxyl groups (linkage peptide). The amino-terminus of this linkage peptide can be blocked, for example by amilation, to prevent the formation of additional variants (FIG. 27). The second reactive group is the amino-terminus of the protein that contains accessory polypeptides. Optionally, one can reserve one or more lysines for coupling to the carboxyls in the linkage peptide. After exhaustive chemical linkage, one can obtain a homogeneous single product, which is a molecule that contains 5 accessory polypeptides (optionally containing binding domains), as well as the linkage peptide. A variation is to have the linkage peptide contain the amino groups and use carboxyls on the other protein, which typically carries the binding domain.

[0421] In addition to such branched structures, it is also possible to create linear polymers of 2, 3, 4, 5, 6, 7, 8 or more separately expressed polypeptides by linking the amino-terminus of one protein to the carboxy-terminus of another protein. Again, these polypeptides may be the same or different, as described above.

[0422] The preferred linkage is amino-to-carboxy. The amino group that is used for coupling is located on the recombinant protein and the carboxyl group that is used is located on the chemical crosslinker. Alternatively, the amino group that is used for coupling is located on the chemical cross-linker if the carboxyl group that is used is located on the recombinant protein.

[0423] The number of coupling sites that is used on the crosslinker determines whether the product will contain 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10 or more recombinant proteins, each typically containing 1, 2, 3, 4, 5 or more binding domains. The crosslinking component is typically a small, FDA-approved chemical but can also be a recombinant polypeptide and optionally contains at least 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 75, 100 units of a repeated motif, and at most 10, 20, 30, 40, 50, 75, 100, 150, 200, 250, 300, 350, 400, 450 or 500 units of a repeated motif.

[0424] Using protection groups that can be differentially removed by different conditions, it is possible to have several conjugation steps that each add a different protein to the conjugate. This allows the creation of conjugates with multiple different protein chains in a pre-defined stoichiometry. Conjugation of divalent crosslinker with two protein chains having one linkable position (like an N-terminus) creates dimers. Crosslinking of proteins with two linkage sites creates a linear polymer. Crosslinking of a trivalent crosslinker with a protein containing linkage sites at both ends creates a dendrimer (FIG. 26).

[0425] In some embodiments, non-crosslinking components may comprise one or more biologically active polypeptides with affinity to a target receptor. These biologically active polypeptides can bind to different target receptors, allowing the generation of crosslinked accessory polypeptides that bind several copies of several different target receptors. Alternatively, non-crosslinking components can comprise multiple biologically active polypeptides that bind several different epitopes of the same target receptor. The resulting crosslinked accessory polypeptide can bind multiple copies of a target receptor while making multiple binding contacts with each target receptor resulting in very high avidity. Another option is to use non-crosslinking components that contain multiple identical binding elements in
order to construct crosslinked accessory polypeptides with a very large number of identical binding sites.

[0426] In some embodiments, non-cross-linking components can comprise affinity tags. Examples for useful affinity tags are Flag, HA-tag, hexa-histidine (SEQ ID NO: 1). These affinity tags facilitate the purification of the non-cross-linking components as well as the resulting crosslinked accessory polypeptides. In addition, affinity tags facilitate the detection of crosslinked accessory polypeptides in biological samples. In particular, affinity tags are useful to monitor the serum half-life and/or the tissue distribution of a crosslinked accessory polypeptide in a patient or in animals.

[0427] In yet other embodiments, non-cross-linking components can contain binding elements that increase the serum half-life of the resulting crosslinked accessory polypeptides. Such binding elements can bind to one or multiple serum components like HSA, IgG, red blood cells, or other serum component that is found in high abundance.

[0428] In still other embodiments, non-cross-linking components can be conjugated to one or more small molecule drug molecules. Examples for useful drug molecules are doxorubicin, melphalan, paclitaxel, maytansines, docucarmycines, calicheamycin, auristatin and other cytotoxic, cytostatic, antiinfective drugs.

[0429] In some embodiments, non-cross-linking components can comprise affinity tags. Examples for useful affinity tags are Flag, HA-tag, hexa-histidine. These affinity tags facilitate the purification of the non-cross-linking components as well as the resulting crosslinked accessory polypeptides. In addition, affinity tags facilitate the detection of crosslinked accessory polypeptides in biological samples. In particular, affinity tags are useful to monitor the serum half-life and/or the tissue distribution of a crosslinked accessory polypeptide in a patient or in animals.

[0430] In other embodiments, non-cross-linking components can comprise protease sites that allow the slow release of binding domains, active drugs, or other subsequences with biological activity.

[0431] Of particular utility are non-cross-linking components that are free of lysine residues. Such sequences contain a single amino group at their N-terminus, which can be utilized for conjugation to the cross-linking component. Non-cross-linking components that contain a single free cysteine residue are also of utility as there are many chemistries available that allow the controlled conjugation to the side chain of free cysteine residues. Another approach is to utilize the C-terminal carboxyl group of the non-cross-linking component as reactive group.

[0432] Many molecules that comprise multiple reactive groups can serve as useful cross-linking components.

[0433] Many useful cross-linking agents are commercially available from companies like Sigma-Aldrich, or Pierce. Of particular utility are cross-linking components that are available in activated form and can be directly used for conjugation. Examples are shown in FIG. 22. Cross-linking components can comprise multiple reactive groups with similar or identical chemical structure (FIG. 23). Such reactive groups can be simultaneously activated and coupled to multiple identical non-cross-linking components resulting in the direct formation of homomultimeric products. Examples for cross-linking components with multiple similar reactive groups are citric acid, EDTA, TSSAT. Of particular interest are branched PEG molecules containing multiple identical reactive groups.

[0434] There are a large number of specific chemical products that work based on the following small number of basic reaction schemes, all of which are described in detail at www.piercenet.com. Examples of useful crosslinking agents are imidoesters, active halogens, maleimide, pyridyl disulfide, and NHS-esters. Homobifunctional crosslinking agents have two identical reactive groups and are often used in a one-step chemical crosslinking procedure. Examples are BS3 (a non-cleavable water-soluble DSS analog), BSOCOES (base-reversible), DMA (Dimethyl adipimidate-2HCl), DMP (Dimethyl pimelimidate-2HCl), DMS (Dimethyl suberimidate-2HCl), DSG (5-carbon analog of DSS), DSP (Lomani’s reagent), DSS (non-cleavable), DST (cleavable by oxidizing agents), DTBP (Dimethyl 3,3’dithiobispropionimidate-2HCl), DTSSP, EGs, Sulfo-EGs, ThPP, TSAT, DFDNB (1,5-Difluoro-2,4-dinitrobenzene) is especially useful for crosslinking between small spatial distances (Korbblatt, J. A. and Lake, D. F. (1980). Cross-linking of cytochrome oxidase subunits with difluoro-nitrobenzene. Can J. Biochem. 58, 219-224).

[0435] Sulphydryl-reactive homobifunctional crosslinking agents are homobifunctional protein crosslinkers that react with sulphydryls and are often based on maleimides, which react with —SH groups at pH 6.5-7.5, forming stable thioether linkages. BM[PEO]3 is an 8-atom polymer spacer that reduces potential for conjugate precipitation in sulphydryl-to-sulphydryl cross-linking applications. BM[PEO]4 is similar but with an 11-atom spacer. BM5 is a non-cleavable crosslinker with a four-carbon spacer. BMDB makes a linkage that can be cleaved with periodate. BMH is a widely used homobifunctional sulphydryl-reactive crosslinker. BMOE has an especially short linker DPDDB and DTME are cleavable crosslinkers. HVBS does not have the hydrolysis potential of maleimides. TMEA is another option. Hetero-bifunctional crosslinking agents have two different reactive groups. Examples are NHS-esters and amines/hydrazines via EDC activation, AEDP, ASBA (photoreactive, iodinatable), EDC (water-soluble carbodiimide). Amine-Sulphydryl reactive bifunctional crosslinkers are AMAS, APDP, BMPS, EMCA, EMCS, GMBS, KMUA, LC-SMCC, LC-SPDH, MBS, SBAP, SIA (extra short), SIAB, SMCC, SMPB, SMPH, SMPT, SPD, Sulfo-EMCS, Sulfo-GMBS, Sulfo-KMUS, Sulfo-LC-SMPT, Sulfo-LC-SPDH, Sulfo-MBS, Sulfo-SIAB, Sulfo-SMCC, Sulfo-SMPB. Amino-group reactive heterobifunctional crosslinking agents are ANB-NOS, MSA, NHS-ASA, SAPD, SAEED, SAND, SANPAH, SAD, SAFD, Sulfo-HSAB, Sulfo-NHS-LC-ASA, Sulfo-SAPD, Sulfo-SANPAH, TFCS. Arginine-reactive crosslinking agents are, for example APG, which reacts specifically with arginines at pH 7-8.

[0436] Polypeptides can be designed to serve as cross-linking components. Such polypeptides can be generated by chemical synthesis or using recombinant techniques. Examples are polypeptides that contain multiple aspartate or glutamate residues. The side chains of these residues as well as the C-terminal carboxyl group can be used for coupling to the non-crosslinking component. By adding one or several amino acids between the aspartate or glutamate residues one can control the distance between reactive groups, which can affect the efficiency of conjugation as well as the overall properties of the resulting crosslinked accessory polypeptide. Of particular utility are polypeptides that contain multiple aspartate or glutamate residues and that carry a protection group at their N-terminal amino group. Examples for suitable
protection schemes are acetylation, succinylation, and other modifications that reduce the reactivity of the N-terminal amino group of the peptide.

[0437] Of particular utility as cross-linking components are dendrimeric constructs. Many dendrimeric structures are known in the art and they can be designed to contain a large number of reactive groups. Examples of crosslinked accessory polypeptides are illustrated in FIG. 24.

[0438] Additional Modifications of Accessory Polypeptides

[0439] An additional mechanism may be incorporated into the design of accessory polypeptides as well as crosslinked accessory polypeptides mediated by peptides with binding affinity to serum-exposed molecules. By binding to such a target, the half-life of the polypeptide of the present invention is further increased. For example, a crosslinked accessory polypeptide may comprise a non-crosslinking unit that comprises a polypeptide with binding affinity to a serum-exposed target. Alternatively, an accessory polypeptide may comprise a sequence coding for a polypeptide with such binding affinity. Preferred serum-exposed targets that peptides or protein domains can be made to bind to for half-life extension are (human, mouse, rat, monkey) serum albumin, Immunoglobulins such as IgG (IgG1, 2, 3, 4), IgM, IgA, IgE, as well as red blood cells (RBC), or endothelial cells. Accessory polypeptides may also comprise, by way of example, sequences that target the extracellular matrix, insert into membranes, or other targeting peptides and domains (FIG. 28).

[0440] In another embodiment, accessory polypeptides or crosslinked accessory polypeptides may comprise several biologically active polypeptides separated as well as sequences that comprise specific cleavage sites for serum proteases (FIG. 29). Following administration or exposure to serum, serum proteases act on the cleavage sites leading to gradual proteolysis and release of biologically active polypeptides or accessory polypeptides into the blood.

[0441] Accessory polypeptides or crosslinked accessory polypeptides may also be modified postsynthetically. In one embodiment, accessory polypeptides are expressed comprising one or more lysine residues (FIG. 30). Following expression, the polypeptides are reacted with a lys-reactive moiety that is attached to at least one second functional unit, which may be for example a biologically active polypeptide. In a related embodiment, the functional unit is a polypeptide with binding affinity for serum-exposed targets, such as serum albumin, Immunoglobulins such as IgG (IgG1, 2, 3, 4), IgM, IgA, IgE, as well as red blood cells (RBC) or endothelial cells.

[0442] Accessory Polypeptides Linked to an Antigen-Binding Unit

[0443] The present invention embodies an accessory polypeptide linked to an antigen-binding unit. The term “antigen-binding units” collectively refers to immunoglobulin molecules and any form of immunologically active portions of immunoglobulin molecules, i.e., molecules that contain an antigen-binding site which specifically binds or immunoreacts with an antigen. Structurally, the simplest naturally occurring antibody (e.g., IgG) comprises four polypeptide chains, two heavy (H) chains and two light (L) chains interconnected by disulfide bonds. The immunoglobulins represent a large family of molecules that include several types of molecules, such as IgG, IgA, IgM and IgE. The term “immunoglobulin molecule” includes, for example, hybrid antibodies, or altered antibodies, and fragments thereof. An antibody binding unit can be broadly divided into “single-chain” (“Sc”) and “non-single-chain” (“Nsc”) types, which include, but not limited to, Fab, scFv, diFv, dAb, diabody, triabody, tetrabody, domain Ab, Fab fragment, Fab', (Fab')2, bispecific Ab and multispecific Ab.

[0444] Also encompassed within the term “antigen binding unit” are immunoglobulin molecules of a variety of species origins including invertebrates and vertebrates. The term “human” as applied to an antigen binding unit refers to an immunoglobulin molecule expressed by a human gene or fragment thereof. The term “humanized” as applied to a non-human (e.g., rodent or primate) antibodies are hybrid immunoglobulins, immunoglobulin chains or fragments thereof which contain minimal sequence derived from non-human immunoglobulin. For the most part, humanized antibodies are human immunoglobulins (recipient antibody) in which residues from a complementary determining region (CDR) of the recipient are replaced by residues from a CDR of a non-human species (donor antibody) such as mouse, rat, rabbit or primate having the desired specificity, affinity and capacity. In some instances, Fv framework region (FR) residues of the human immunoglobulin are replaced by corresponding non-human residues. Furthermore, the humanized antibody may comprise residues which are found neither in the recipient antibody nor in the imported CDR or framework sequences. These modifications are made to further refine and optimize antibody performance and minimize immunogenicity when introduced into a human body. In general, the humanized antibody will comprise substantially all of at least one, and typically two, variable domains, in which all or substantially all of the CDR regions correspond to those of a non-human immunoglobulin and all or substantially all of the FR regions are those of a human immunoglobulin sequence. The humanized antibody may also comprise at least a portion of an immunoglobulin constant region (Fc), typically that of a human immunoglobulin.

[0445] “Non-single-chain antigen-binding unit” are heterotrimers comprising a light-chain polypeptide and a heavy-chain polypeptide. Examples of the non-single-chain antigen-binding unit include but are not limited to (i) a cFv fragment, which is a dimeric protein composed of VLF and VH regions, which dimerize via the pairwise affinity of the first and second heterodimerization sequences fused in-frame with the VL and VH regions; (ii) any other monovalent and multivalent molecules comprising at least one cFv fragment; (iii) an Fab fragment consisting of the VL, VH, CL, and CH1 domains; (iv) an Fd fragment consisting of the VH and CH1 domains; (v) an Fv fragment consisting of the VL and VH domains of a single arm of an antibody; (vi) an F(ab')2 fragment, a bivalent fragment comprising two Fab fragments linked by a disulfide bridge at the hinge region; (vii) a diabody; and (viii) any other non-single-chain antigen-binding units that have been described and known in the art.

[0446] As noted above, a non-single-chain antigen-binding unit can be either “monovalent” or “multivalent.” Whereas the former has one binding site per antigen-binding unit, the latter contains multiple binding sites capable of binding to more than one antigen of the same or different kind. Depending on the number of binding sites, a non-single-chain antigen-binding unit may be bivalent (having two antigen-binding sites), trivalent (having three antigen-binding sites), tetravalent (having four antigen-binding sites), and so on. Multivalent non-single-chain antigen-binding unit can be further classified on the basis of their binding specificities. A “monospecific” non-single-chain antigen-binding unit is a
molecule capable of binding to one or more antigens of the same kind. A “multispecific” non-single-chain antigen-binding unit is a molecule having binding specificities for at least two different antigens. While such molecules normally will only bind two distinct antigens, antibodies with additional specificities such as trispecific antibodies are encompassed by the present invention.

[0447] “Single-chain antigen-binding unit” refers to monomeric antigen-binding unit. Although the two domains of the Fv fragment are coded for by separate genes, a synthetic linker can be made that enables them to be made as a single protein chain (i.e., single chain Fv (“scFv”) as described in Bird et al. (1988) Science 242:423-426 and Huston et al. (1989) PNAS 85:5879-5883) by recombinant methods. Other single-chain antigen-binding units include antigen-binding molecules stabilized by the subject heterodimerization sequences, and dAb fragments (Ward et al. (1989) Nature 341:544-546) which consist of a VH domain and an isolated complementarity determining region (CDR). A preferred single-chain antigen-binding unit contains VL and VH regions that are linked together and stabilized by a pair of subject heterodimerization sequences. The scFv can be assembled in any order, for example, VH—(first heterodimerization sequence)—VL or V.sub.L—(first heterodimerization sequence)—VH. The antibody fragment provides an AFBT with specificity for a target antigen (also generally illustrated in FIG. 21). The rPEG domain provides a variety of benefits to the antibody fragment as well as to the payload. These benefits include, but are not limited to, prolonged half-life in vivo, increased solubility, increased thermal stability, increased protease stability, improved protein folding, reduced chain reassociation, reduced immunogenicity of the payload, and avoidance of preexisting immune responses to chemical PEG. The rPEG domain also facilitates production and purification. The high solubility of the rPEG domain renders AFBTs high solubility that can be formulated at high concentration with a low tendency to form aggregates. It should be understood that an AFBT may contain additional components not illustrated in particular in this figure.

[0453] vH/vL Domain-Based Structures

[0454] In one embodiment of the present invention, an AFBT also comprises one or more antibody-derived immunoglobulin (Ig) domains or fragments, including a single-chain variable fragment (scFv). scFv consists of a VH domain linked to a VL domain via a peptide linker between the VH and VL domains. The linker in the scFv is chosen such that it forms a single molecular species, which includes a scFv, diabody, triabody, or tetrabody (FIGS. 53, 54, 55), as compared to the full-length, i.e., whole antibody (FIG. 52). Typically the valency of the resulting AFBT is between one and four although a higher valency is not excluded. Designs that predominantly form a single, homogeneous species are preferred. An Fv fragment may include a disulfide bond between contacting VH and VL domains to reduce the risk of domain reassociation. The fraction of the desired species that may be achieved ranges from less than 1% to 100% of the antibody fragment mix. The primary controls are the linker length, which directs the format, and the rPEG, which reduces antibody fragment chain reassociation. A preferred embodiment includes the formation of monomeric scFv from a single vH-vL chain employing linkers of at least 12 amino acids. More preferred embodiments include a linker length of at least 15, at least 20, at least 30, at least 50, at least 100, at least 200, or at least 288 amino acids. Of particular utility are vH-vL chains that preferentially form diabodies, which require linkers of less than 10-20 amino acids, preferably 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10 amino acids (FIG. 84). A diabody has two protein chains and can have an rPEG at one or both C-terminal ends, and/or at one or both N-terminal ends. The diabody has two binding sites, of which zero, one or two may bind to a pharmaceutical target, or to a cell surface target (e.g., HSA, IgG, Red Blood Cells, Collagen, etc.) or to no target. The diabody may contain zero, one or more drug modules located at the N-terminal or C-terminal end of zero, one or both protein chains. AFBTs containing diabodies have increased molecular weight due to their dimeric structure, which slows down renal elimination. In one embodiment, the degree of antibody fragment chain reassociation from one
species to another species is less than 50%, 40%, 30%, 20% or 10% of the mass of protein per day or per week at a fixed temperature (e.g., 4°C, 25°C, or 37°C), preferably less than 5%, 2%, 1% or 0.1%.

[0455] In another embodiment, the AFBTs include triabodies, which contain three polypeptide chains, each containing a VH and a VL domain connected via a linker of less than 10 amino acids, preferably less than 5 amino acids. The frequency of triabodies can be increased by eliminating one or a few amino acids from one or both joining ends of the VH and VL domains, shortening the connecting sequence so that triabodies are favorably formed. The number of residues removed from one or both of the fused ends of the antibodies can range from 1 to 10 amino acids.

[0456] In yet another embodiment, the AFBTs include tetra- bodies, which contain four polypeptide chains, each having one VH domain and a VL domain connected via a short linker of less than 10 amino acids, or as a result of removal of 1 to 10 residues from one or both of the fused ends of the antibody. The number of amino acids to eliminate from one or both joining ends of the VH and VL domains can be adjusted to ensure the most desirable outcome.

[0457] Examples of various types of single chain (scFv) combination consisting of a single copy of a polypeptide chain include but are not limited to scFv-scFv, scFv-fab-scFv, dAb-scFv, scFv-fab-scFv, scFv-scFv, and scFv-dAb (FIG. 57). A scFv fragment can be fused at one or both of the N- and/or C-terminal ends to a drug module such as IFNα, IgH, etc. (FIG. 85). The scFv has one binding site, which may or may not bind to a pharmaceutical target, or to a half-life target, e.g., HSA (FIG. 85). IgG, red blood cells, etc.

[0458] AFBTs that contain Ig domains can have a variety of architectures. Constructs of particular utility include, but are not limited to, the following: VH-linker-VH-DPEG, VH-linker-VH-DPEG-payload, VH-linker-VH-DPEG-payload, VH-linker-VH-DPEG-payload-DPEG, VH-linker-VH-payload-DPEG, VH-linker-VH-payload-DPEG, VH-linker-VH-payload-DPEG, VH-linker-VH-payload-DPEG, VH-linker-VH-payload-DPEG. These AFBTs can contain additional domains that can be inserted between domains or anywhere into an HPEG domain. There can also be several payload modules.

[0459] The linker sequence joining VH and VL can be optimized to achieve optimal protein folding and stability as well as high level expression and a large fraction of the desired species. A preferred embodiment includes linker sequences that are rich (e.g., greater than 50%) in glycine and other small hydrophilic amino acids such as serine, threonine, glutamic acid, aspartic acid, lysine, arginine, and alanine HPEG is particularly suitable as a linker between VH and VL domains. Linkers with improved properties can be obtained by selection or screening of libraries.

[0460] scFv with HPEG Linkers

[0461] In yet another embodiment, scFv contain DPEG sequences as the linker between the VH and VL domains. A preferred embodiment includes HPEG linkers that contain a significant negative net charge, which results in improved solubility and folding of the scFv domains. Preferred embodiments contain linkers with at least 15, at least 20, at least 30, at least 50, at least 100, at least 200, or at least 288 residues.

[0462] Methods to Generate Stable Antibody Fragments of AFBTs

[0463] The present invention also relates to methods of generating and engineering an antigen binding unit of one AFBT. Many methods are known to generate antibodies with specificity for a target antigen. Examples include monoclonal antibodies, in particular in transgenic animals that produce human antibodies; phage display of Fab or scFv libraries; ribosomal display; and humanization of monoclonal antibodies. Multiple methods to engineer the stability of scFvs have also been described [Worn, A., et al. (2001) J Mol Biol, 305: 989]. It has been shown that adding a disulfide bond between the VH and VL domains of scFv can lead to significant stabilization [Dooley, H., et al. (1998) Biotechnol Appl Biochem 28 (Pt 1), 77, #2002]. An alternative is the introduction of consensus mutations. The amino acid frequencies at various positions in antibody framework residues have been analyzed. It has been shown that the Boltzmann equation can predict the stabilizing effect of some consensus mutations [Steipe, B., et al. (1994) J Mol Biol 240, 188, #2006]. A combinatorial approach that allows the simultaneous introduction of multiple consensus mutations into single chain antibody fragments has been described [Robe, M., et al. (2006) Protein Eng Des Sel, 19: 114]. Producing more stable antibody fragments has resulted in improved in vivo targeting [Worn, A., et al. (2000) J Biol Chem, 275: 2795].

[0464] Some scFv have been expressed in soluble form in the cytosol of E. coli. In general, disulfide bonds are not formed in the cytosol but can form spontaneously after cell lysis [Tavladoraki, P., et al. (1999) Eur J Biochem, 262: 617]. In general, cytosolic expression of an antibody is well correlated with the antibody stability [Worn, A., et al. (2001) J Mol Biol, 305: 989]. Mutant libraries of antibody fragments can be subjected to selection for improved cytosolic expression [Martineau, P., et al. (1998) J Mol Biol, 280: 117]. Redox engineered strains of E. coli can be used to improve cytosolic expression of Fab fragments [Levy, R., et al. (2007) J Immunol Methods, 321: 164]. The culture conditions have been optimized to improve the expression of soluble scFv in the cytosol of E. coli resulting in expression levels of up to 35 mg/L of culture [Padiouelle-Lefèvre, S., et al. (2007) Mol Immunol, 44: 1888]. Another approach to improve the cytosolic expression of scFvs is the screening or selection of genomic libraries with the goal to identify chaperones or other factors that facilitate expression. This approach has been evaluated using lambda phage. Disulfide bonds in scFv have been removed successfully to form intrabodies. Variants of such intrabodies can be identified that result in improved cytosolic expression [der Maur, A., A., et al. (2002) J Biol Chem, 277: 45075]. However, disulfide bonds are important for the overall stability of most antibody fragments and in most cases intrabodies have been of limited utility.

[0465] Complementary Determining Regions (CDR) Grafting

[0466] The binding interactions between antibodies or antibody fragments and their targets are mainly determined by the complementary determining regions (CDRs). It has been shown that CDRs can be grafted between the variable domains of different antibodies [Jones, P. T., et al. (1986) Nature, 321: 522]. In many cases other residues in the antibody framework need to be grafted in addition to CDR residues in order to retain antigen binding. CDR grafting can be useful to improve the stability of an antibody by grafting CDRs from a less stable variable domain to a more stable variable domain. An example is the grafting of CDRs from a fluorescein-binding scFv into a well-expressed scFv that is used as a “scaffold”, resulting in improved expression and increased folding stability [Jung, S., et al. (1997) Protein Eng,

**[0467]** Affinity of the Antigen Binding Unit of AFBT


**[0469]** Various IgG Domains

[0470] AFBTs may contain a variety of immunoglobulin domains. These domains can affect protein expression, multimerization, and can serve as effectors. The following non-exhaustive list, which provides examples for illustrating the variability of Ig domains, is applicable for fusions to any antibody isotype including IgG1, IgG2, IgG3, IgG4, IgE, IgM, IgA, and IgD from any species including humans. Siles for fusion of rPEG to immunoglobulin-family sequences include but are not limited to the following:

<table>
<thead>
<tr>
<th>Light chain</th>
<th>Heavy chain</th>
</tr>
</thead>
<tbody>
<tr>
<td>vL-CL-rPEG</td>
<td>vH-CH1</td>
</tr>
<tr>
<td>vL-CL-rPEG-payload</td>
<td>vH-CH1</td>
</tr>
<tr>
<td>vL-CL-payload-rPEG</td>
<td>vH-CH1</td>
</tr>
<tr>
<td>rPEG-vL-CL</td>
<td>vH-CH1</td>
</tr>
<tr>
<td>payload-rPEG-vL-CL</td>
<td>vH-CH1</td>
</tr>
<tr>
<td>rPEG-payload-vL-CL</td>
<td>vH-CH1</td>
</tr>
<tr>
<td>vL-CL-rPEG</td>
<td>vH-CH1</td>
</tr>
<tr>
<td>vL-CL-payload-rPEG</td>
<td>vH-CH1</td>
</tr>
<tr>
<td>rPEG-vL-CL</td>
<td>vH-CH1</td>
</tr>
<tr>
<td>payload-rPEG-vL-CL</td>
<td>vH-CH1</td>
</tr>
<tr>
<td>rPEG-payload-vL-CL</td>
<td>vH-CH1</td>
</tr>
<tr>
<td>vL-CL-rPEG</td>
<td>vH-CH1</td>
</tr>
<tr>
<td>vL-CL-payload-rPEG</td>
<td>vH-CH1</td>
</tr>
<tr>
<td>rPEG-vL-CL</td>
<td>vH-CH1</td>
</tr>
<tr>
<td>payload-rPEG-vL-CL</td>
<td>vH-CH1</td>
</tr>
<tr>
<td>rPEG-payload-vL-CL</td>
<td>vH-CH1</td>
</tr>
<tr>
<td>vL-CL-rPEG</td>
<td>vH-CH1</td>
</tr>
<tr>
<td>vL-CL-payload-rPEG</td>
<td>vH-CH1</td>
</tr>
<tr>
<td>rPEG-vL-CL</td>
<td>vH-CH1</td>
</tr>
<tr>
<td>payload-rPEG-vL-CL</td>
<td>vH-CH1</td>
</tr>
<tr>
<td>rPEG-payload-vL-CL</td>
<td>vH-CH1</td>
</tr>
<tr>
<td>vL-CL-rPEG</td>
<td>vH-CH1</td>
</tr>
<tr>
<td>vL-CL-payload-rPEG</td>
<td>vH-CH1</td>
</tr>
<tr>
<td>rPEG-vL-CL</td>
<td>vH-CH1</td>
</tr>
<tr>
<td>payload-rPEG-vL-CL</td>
<td>vH-CH1</td>
</tr>
<tr>
<td>rPEG-payload-vL-CL</td>
<td>vH-CH1</td>
</tr>
</tbody>
</table>

**[0471]** N-terminal to the CL1 domain, before the interchain cysteine

**[0472]** N-terminal to the CL1 domain, after the interchain cysteine

**[0473]** C-terminal to the CL1 domain, before the interchain cysteine

**[0474]** C-terminal to the CL1 domain, after the interchain cysteine

**[0475]** N-terminal to the CH1 domain, before the interchain cysteine

**[0476]** N-terminal to the CH1 domain, after the interchain cysteine

**[0477]** C-terminal to the CH1 domain, before the interchain cysteine

**[0478]** C-terminal to the CH1 domain, after the hinge cysteine(s)

**[0479]** C-terminal to the CH1 domain, after the hinge cysteine(s)

**[0480]** N-terminal to the hinge cysteine(s)

**[0481]** C-terminal to the hinge cysteine(s), before CH2

**[0482]** N-terminal to the CH2 domain

**[0483]** C-terminal to the CH2 domain

**[0484]** N-terminal to the CH3 domain

**[0485]** C-terminal to the CH3 domain

**[0486]** N-terminal to the CH4 domain

**[0487]** C-terminal to the CH4 domain

**[0488]** N-terminal to peptides derived from CDRH1-3 and/or CDRL1-3 (lambda and kappa)

**[0489]** N-terminal to peptides derived from CDRH1-3 and/or CDRL1-3 (lambda and kappa)

**[0490]** Fab Domain Based AFBTs

**[0491]** Still another embodiment of the present invention includes a Fab domain-based AFBT (FIG. 56). Fab domains comprise two peptide chains, each of which is derived from the heavy and light chains of an antibody, rPEgs and payloads and other domains can be fused to either chain of a Fab fragment. Alternatively, rPEgs and payloads can be fused to both chains of a Fab. Fab domains can be designed to facilitate the dimerization of the resulting proteins such that the final protein contains four peptide chains.

**[0492]** The following is a list of AFBTs that comprise at least one Fab domain:
Full Length Antibodies

rPEGs and payloads and other domains can be fused to the light chain or heavy chain of an antibody, or to both chains of an antibody. The following table illustrates a few examples of AFBTs that are based on full-length antibodies:

<table>
<thead>
<tr>
<th>Light chain</th>
<th>Heavy chain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Light chain-rPEG</td>
<td>Heavy chain</td>
</tr>
<tr>
<td>Light chain-rPEG-payload-rPEG</td>
<td>Heavy chain</td>
</tr>
<tr>
<td>rPEG-Light chain</td>
<td>Heavy chain</td>
</tr>
<tr>
<td>payload-rPEG-Light chain</td>
<td>Heavy chain</td>
</tr>
<tr>
<td>rPEG-payload-Light chain</td>
<td>Heavy chain</td>
</tr>
<tr>
<td>Light chain-rPEG</td>
<td>Heavy chain-rPEG</td>
</tr>
<tr>
<td>Light chain-rPEG-payload</td>
<td>Heavy chain-rPEG</td>
</tr>
<tr>
<td>rPEG-Light chain</td>
<td>Heavy chain-rPEG</td>
</tr>
<tr>
<td>Light chain-payload-rPEG</td>
<td>Heavy chain-rPEG</td>
</tr>
<tr>
<td>rPEG-Light chain</td>
<td>Heavy chain-rPEG</td>
</tr>
<tr>
<td>Light chain-payload-rPEG</td>
<td>Heavy chain-rPEG</td>
</tr>
<tr>
<td>rPEG-Light chain</td>
<td>Heavy chain-rPEG</td>
</tr>
<tr>
<td>payload-rPEG-Light chain</td>
<td>Heavy chain-rPEG</td>
</tr>
<tr>
<td>rPEG-payload-Light chain</td>
<td>Heavy chain-rPEG</td>
</tr>
<tr>
<td>Light chain-rPEG</td>
<td>Heavy chain-rPEG-payload</td>
</tr>
<tr>
<td>Light chain-rPEG-payload-rPEG</td>
<td>Heavy chain-rPEG-payload</td>
</tr>
<tr>
<td>rPEG-Light chain</td>
<td>Heavy chain-rPEG-payload</td>
</tr>
<tr>
<td>Light chain-payload-rPEG</td>
<td>Heavy chain-rPEG-payload</td>
</tr>
<tr>
<td>rPEG-Light chain</td>
<td>Heavy chain-rPEG-payload</td>
</tr>
<tr>
<td>payload-rPEG-Light chain</td>
<td>Heavy chain-rPEG-payload</td>
</tr>
<tr>
<td>rPEG-payload-Light chain</td>
<td>Heavy chain-rPEG-payload</td>
</tr>
<tr>
<td>Light chain-rPEG</td>
<td>rPEG-Heavy chain</td>
</tr>
<tr>
<td>Light chain-rPEG-payload</td>
<td>rPEG-Heavy chain</td>
</tr>
<tr>
<td>rPEG-Light chain</td>
<td>rPEG-Heavy chain</td>
</tr>
<tr>
<td>Light chain-payload-rPEG</td>
<td>rPEG-Heavy chain</td>
</tr>
<tr>
<td>rPEG-Light chain</td>
<td>rPEG-Heavy chain</td>
</tr>
<tr>
<td>payload-rPEG-Light chain</td>
<td>rPEG-Heavy chain</td>
</tr>
<tr>
<td>rPEG-payload-Light chain</td>
<td>rPEG-Heavy chain</td>
</tr>
<tr>
<td>Light chain-rPEG</td>
<td>payload-rPEG-Heavy chain</td>
</tr>
<tr>
<td>Light chain-rPEG-payload-rPEG</td>
<td>payload-rPEG-Heavy chain</td>
</tr>
<tr>
<td>rPEG-Light chain</td>
<td>payload-rPEG-Heavy chain</td>
</tr>
<tr>
<td>Light chain-payload-rPEG</td>
<td>payload-rPEG-Heavy chain</td>
</tr>
<tr>
<td>rPEG-Light chain</td>
<td>payload-rPEG-Heavy chain</td>
</tr>
<tr>
<td>payload-rPEG-Light chain</td>
<td>payload-rPEG-Heavy chain</td>
</tr>
<tr>
<td>rPEG-payload-Light chain</td>
<td>payload-rPEG-Heavy chain</td>
</tr>
<tr>
<td>Light chain-rPEG</td>
<td>rPEG-payload-Heavy chain</td>
</tr>
<tr>
<td>Light chain-rPEG-payload</td>
<td>rPEG-payload-Heavy chain</td>
</tr>
<tr>
<td>rPEG-Light chain</td>
<td>rPEG-payload-Heavy chain</td>
</tr>
<tr>
<td>Light chain-payload-rPEG</td>
<td>rPEG-payload-Heavy chain</td>
</tr>
<tr>
<td>rPEG-Light chain</td>
<td>rPEG-payload-Heavy chain</td>
</tr>
<tr>
<td>payload-rPEG-Light chain</td>
<td>rPEG-payload-Heavy chain</td>
</tr>
<tr>
<td>rPEG-payload-Light chain</td>
<td>rPEG-payload-Heavy chain</td>
</tr>
</tbody>
</table>

Multispecific AFBTs

The present invention also embodies AFBTs that comprise fragments derived from multiple different antibodies with different binding specificities. An example is shown in FIG. 58b. Such AFBTs combine the binding specificities of two or more parent antibodies. Parent antibodies can be chosen such that the resulting AFBT binds to multiple different target antigens. Alternatively, the parent antibodies can bind to different epitopes of the same target antigen. AFBTs bind the target very tightly if they can form multivalent interactions by binding to multiple sites on the same target antigen as illustrated in FIG. 59. Multispecific AFBTs can form multimers of the same protein chain. For instance, FIG. 58c illustrates a multispecific AFBT that is a dimer of two polypeptide chains that contains two binding sites based on the VH-VL chain A and two additional binding sites based on the VH-VL chain B. One skillful in the art can appreciate the possibility of generating a multispecific AFBT containing many different combinations of binding domains or binding modules. In addition to different variable domains, multispecific AFBTs may include one or more payload domains, rPEG modules and other protein domains that can be chosen to enhance therapeutic utility or production and purification. One embodiment includes multispecific AFBTs that interact with multiple target antigens that are related to the same disease symptoms, the same pathogen or cause of pathogenesis, or the same physiological pathway or process. Examples of such multispecific AFBTs include but are not limited to multispecific AFBTs that block multiple cytokines which are involved in a related biological process. A preferred embodiment includes multispecific AFBTs that block multiple growth factors that are involved in angiogenesis such as VEGF, PDGF, and P1GF. FIG. 59 shows an rPEGs flanked on both sides by a VEGF-receptors. Since VEGF is dimeric, it can be the same receptor on both sides of the rPEG, or a different receptor, preferably VEGF-R1 and VEGF-R2, but VEGF-R3 may also be used. Another preferred embodiment includes multispecific AFBTs that block multiple cytokines that are involved in inflammatory diseases such as TNF-α, IL-1, IL-6, IL-12, IL-13, IL-17, and IL-23. Yet another preferred embodiment includes multispecific AFBTs that bind multiple tumor antigens such as Her1, Her2, Her3, EGFR, TF antigen, CEA, A33, PSMA, MUC1, αvβ3 integrin, αvβ5 integrin, and α5β1 integrin. Still another preferred embodiment includes multispecific AFBTs that bind multiple antigens that are related to an infectious disease. Said multispecific AFBTs can form multivalent interactions with an infectious agent resulting in improved therapeutic efficacy. Multispecific AFBTs can be engineered to comprise a binding site for a tumor antigen and a second binding site for an antigen on an immune cell. Examples include AFBTs that bind tumor antigens and CD3 or CD16, which can recruit and activate natural killer (NK) cells. To further increase potency, a cytokine domain such as IL-2 can be included to activate immune cells in the vicinity of the tumor cells.

Domain Antibody-Based AFBTs

In yet another embodiment, rPEGs and payloads and other domains can be fused to a domain antibody (dAb). In order to generate domain antibodies with suitable binding properties, one can use the naturally nonmoneric VH domains (called vH1) found in the immune repertoire of camels and sharks that naturally lacks a light chain. [Flammers-Casterman, C., et al. (1993) Nature, 363: 446]. Alternatively, one can engineer the vH1-vL interface of a human vH or vL domain in order to improve solubility and reduce dimerization and aggregation. Such mutations carry the risk of increasing immunogenicity of the resulting domain antibody. The present invention describes fusing human vH or vL Ig domains to rPEG, which improves solubility and folding, reduces aggregation, and yet does not induce immune response triggered by the mutagenesis of human framework residues. Examples of AFBTs which are based on dAb domains include, but are not limited to, dAb-rPEG, dAb-rPEG-payload, dAb-payload-rPEG, rPEG-dAb, payload-rPEG-dAb. dAb domain can be derived from the vH or vL domain of an antibody molecule.
AFBTs Containing Multiple Fragments of the Same Antibody

AFBTs can be engineered such that each polypeptide chain contains multiple variable fragments of the same parent antibody. These fragments can be identical in their sequence or they can be engineered to facilitate proper domain assembly. An example is illustrated in FIG. 60a. This AFBT contains a diabody domain and a monoclonal scFv domain based on the same parent antibody. As a result, the AFBT assembles into a dimeric structure that contains a total of 4 equivalent target binding sites. Such multivalent AFBTs can have improved potency due to avidity.

Bispecific AFBTs Based on Diabodies

AFBTs can be constructed to combine one diabody and a variable domain and at least one rPEG domain. The constructs form dimers and contain a total of 4 antigen binding sites. FIG. 58b illustrates an example of a bispecific AFBT. The variable domains A in such a construct can be scFv domains or dAb domains. The variable domains A can be at the C-terminal side of the diabody domain A. Alternatively, the variable domains A can be at the N-terminal side of the diabody domain B. Bispecific AFBTs can contain additional rPEG domains or other domains such as hormones, cytokines, or enzymes. If the variable domain in a bispecific AFBT is a scFv domain, the scFv domain can have the configuration VH-linker-VL or the configuration Vl-linker-VH.

In a preferred embodiment, a bispecific AFBT comprises a diabody B and a scFv A, in which the diabody and scFv domains are optimized to reduce incorrect pairing of the 4 lg domains in these constructs. The diabody can be optimized such that Vn-A and Vn-B as well as Vn-B and Vn-Ab form tighter interactions than the incorrect pairings Vn-A and Vn-B and Vn-B and Vn-Ab. This can be accomplished by choosing frameworks of both VH and VL domains such that the VH/VL contact surface of scFv domain A has significant structural differences compared to the VH/VL contact surface of diabody domain B. One can further enhance these differences by engineering the VH/VL contact regions of scFv domain A and diabody domain B to minimize the chance of undesired contacts. For instance, one can engineer charge differences such that an ion pair is formed for correct VH/VL pairing but the same ion pair cannot be formed during incorrect pairings of VH and VL domains in the bispecific AFBT. Another approach is to construct diabody domains by bonding partners into the desired VH/VL contact surfaces that can not be formed in incorrect pairings of VH and VL domains. Yet another approach is to alter the shape of the contact VH/VL contact surfaces such that incorrect VH/VL pairs are destabilized.

Bispecific AFBTs based on diabodies are of particular utility as they contain two rPEG domains per divalent complex, which results in reduced kidney filtration and improved in vivo half-life. AFBTs can be engineered to contain a diabody domain and two additional variable domains per polypeptide chain. Such a protein can form dimeric complexes that comprise a total of 6 antigen binding sites. Further variable fragments or payload domains can be added to increase potency.

Dimeric AFBTs Containing Payloads

FIG. 60b illustrates a dimeric AFBT that contains a diabody domain and a payload domain. Such proteins form dimeric complexes such that each complex contains 2 target binding sites, 2 rPEGs, and 2 payload domains. Additional protein domains can be added to increase utility. Having two rPEGs per protein complex reduces kidney filtration and increases in vivo half-life. Having two payload domains increases potency. The target binding sites of the diabody domain can be engineered to further increase in vivo half-life by binding to a component of blood such as red blood cells, human serum albumin, IgG, collagen or other protein or cell in the blood.

Combining Antibody Fragments and Payloads

The present invention also embodies AFBTs which comprise one or more payloads. One preferred embodiment includes payloads that are protein domains and can be directly fused to the other domains comprising an AFBT. Examples of such payload domains include, but are not limited to, cytokines, hormones, growth factors, and enzymes. Such AFBTs combine the specificity of antibodies with the efficacy of the payload while the rPEG domain provides half-life and facilitates production and formulation. Another preferred embodiment includes AFBTs that combine an antibody fragment with specificity for a particular tissue and a payload that exerts its activity in the same tissue. One example includes antibody fragments with specificity for a tumor in combination with cytostatic or cytotoxic payloads. Another example comprises antibody fragments with specificity for infected cells or infectious agents in combination with anti-infective payloads. Yet another useful combination comprises antibody fragments with specificity against inflammation in combination with payloads that have anti-inflammatory activity. Antibodies that can be linked to an accessory polypeptide include, but are not limited to, abxizumab, alemtuzumab, basiliximab, bevacizumab, cetuximab, daclizumab, ecclizumab, eflazumab, ibritumomab, tiuxetan, infliximab, muromonab-CD3, natalizumab, omalizumab, palivizumab, pantitumumab, ranibizumab, gemtuzumab ozogamicin, rituximab, tositumomab, trastuzumab, and any antibody fragments specific for antigens including complement C5, CBL, CD147, II8, gp120, VL-A4, CD11a, CD18, VEGF, CD40, anti-IlD, ICAM1, CD2, EGF, TGF-132, TNFx, E-selectin, FucetII, Her2/neu, F gp, CD11/18, CD14, CD80, ICAM3, CD4, CD23, [2D-integrin, α4β7], CD52, CD22, HLA-DR, CD64 (FcR), TCR αβ, CD3, Hep B, CD125, EpCAM, gp120IIIa, Igβ, CD20, IL5, IL4, CD25, CD33, HLA, F gp, and VRIntegrin.

Enzymes can be used as payloads for tumor-specific AFBTs. Enzymes can be chosen in order to eliminate a required nutrient or metabolite from the tumor environment, such as asparaginase, arginase, histidinase, or methioninase. Alternatively, one can utilize enzymes that exert cytotoxic activity. An example includes AFBTs that comprise a tumor specific antibody fragment and RNAse which induces apoptosis upon internalization into cells.

Payloads that are useful in anti-cancer, anti-microbial, and/or anti-inflammatory therapies include toxins such as Pseudomonas exotoxin, ricin, botulinum toxin, and other plant or bacterial toxins. Other biological toxins include, but are not limited to, abrin, aerolysin, botulinum toxin A, B, C1, C2, D, E, F, F-hungaroxin, Caerulescotoxin, Ceratiolysin, Cholera toxin, Clostridium difficile enterotoxin A and B, Clostridium perfringens lecithinase, Clostridium perfringens kappa toxin, Clostridium perfringens perfringolysin O, Clostridium perfringens enterotoxin, Clostridium perfringens beta toxin, Clostridium perfringens deltoxin, Clostridium perfringens epsilon toxin, Conotoxin, Crotoxin, Diphtheria toxin, Listeriolysin, Leucocidin, Modeccin, Nematocyst toxins, Notexin, Pertussis toxin, Pseudolysin, Pseudomonas aeruginosa toxin A, Saxitoxin, Shiga toxin,
Shigella dysenteriae neurotoxin, Streptolysin O, Staphylococcus enterotoxins B and F, Streptolysin S, Taipoxin, Tetanus toxin, Tetrotoxin, Viscumum, Volkensin, and Yersinia pestis murine toxin.

Payloads can be chosen to eliminate a toxic metabolite. Examples are urate oxidase for the treatment of gout and phenylalanine ammonia lyase for the treatment of phenylketonuria. Payloads can also comprise chemically conjugated small molecules. Such payloads can be conjugated to an AFBT resulting in a semisynthetic AFBT. The protein portion of a semisynthetic AFBT can be engineered to facilitate controlled chemical conjugation via exhaustive coupling as illustrated in FIG. 61. The protein portion can be engineered to have a defined number of coupling sites. This enables the use of a coupling reagent in excess to the concentration of coupling sites such that coupling efficiency can be close to completion, which results in a defined coupling product. Useful coupling sites can be amino groups. The protein portion of such semisynthetic AFBTs can be engineered such that all or most lysine residues in the antibody fragments are replaced with other residues that are compatible with folding and target binding. In many proteins one can replace lysine residues with arginine, glutamate, aspartate, serine, threonine or another amino acid. Designated coupling sites can be incorporated into the rPEG domain or into any other protein portion of the protein. In addition, the N-terminus of each protein chain can serve as a conjugation site. Cysteine residues can also serve as conjugation sites. Example payloads that can be conjugated to AFBTs include cytotoxic drugs such as doxorubicin, auristatin, maytansine and related molecules that can be fused to AFBTs with tumor-specific antibody fragments. Other payloads of interest for conjugation include antiviral compounds, imaging reagents, and chemotherapeutic agents that can be labeled with radionuclides to generate imaging agents or AFBTs for radiotherapy.

Thiols in rPEG Tail

Another embodiment of the present invention includes AFBTs comprising rPEG sequences that contain one or multiple cysteine residues. These cysteine side chains can form disulfide bridges with other proteins after injection into a patient. These disulfide bridges can result in increased in vivo half-life. In other embodiments disulfide bond formation can result in prolonged retention of AFBTs at the injection site resulting in a slow-release PK profile. AFBTs that contain free cysteins can also be engineered for improved bioavailability for oral, intranasal, and intradermal administered AFBTs. This can be achieved by forming disulfide bridges with proteins at the surface of epithelial cells resulting in enhanced uptake of the AFBT.

RGD-Peptides in rPEG

AFBTs may also contain one or multiple RGD sequences or related sequences that are known to interact with integrins as well as components of the extracellular matrix. These RGD-related sequences can be flanked by cysteine residues to result in disulfide-mediated cyclization. Alternatively, the RGD-related sequences can be flanked by additional amino acids that can be selected to enhance the affinity and/or specificity of interaction with a particular integrin. One preferred embodiment includes AFBTs that contain RGD sequences and interact with integrins α3β1, α5β1, that are overexpressed on a variety of tumor cells.

Antibody Fragments that Increase Half-Life

The present invention also embodies AFBTs that contain antibody fragments that increase the in vivo half-life of the AFBTs. This can be achieved by incorporating antibody fragments that bind to targets that have long in vivo half-lives. Examples of such targets that increase the in vivo half-life include but are not limited to serum proteins, in particular, serum albumin, immunoglobulins, and other highly abundant proteins. AFBTs can also incorporate antibody fragments with specificity for blood cells or vessel walls. Of particular interest are red blood cells (RBCs), which are extremely abundant, have an average life span of approximately 4 months, and are characterized by minimal metabolic activity. AFBTs can be engineered to bind any protein on the surface of an RBC. A preferred embodiment includes AFBTs that bind to glycophorin A, which is expressed in high abundance on the surface of RBCs. AFBTs can be engineered to bind to any cell surface target that is in contact with an AFBT in vivo resulting in a prolonged retention of the AFBT. Another embodiment includes AFBTs that bind to components of the extracellular matrix (ECM). The ECM contains many proteins including, but not limited to, agrin, alpha-elastin, amysin, bestrophin, collagen, contactin 1, CRIP1, drebrin, entactin, fettu A, HAS3, HCAP-G, syndecan, KALI, 1 Aafatin, laminin, Mint3, MMP24, NCAM, neurocan, nidogen 2, optineurin, procollagen type II, PSCDBP, reelin, SIRP, synaptogamin, synCAM, syndecan, syntrophin, TAG1, tenascin C, and zyxin. Yet another embodiment includes AFBTs which comprise antibody fragments that bind the FeRn receptor, which results in recycling of endocytosed AFBTs. Examples include antibody fragments that show pH-dependent binding to FeRn such that the antibody fragment binds FeRn with low affinity at around neutral pH but binds with high affinity at lower pH, e.g., pH 5, which is within the range of pH predominantly found in lysosomal compartments. AFBTs that provide increased half-life are illustrated in FIGS. 58a and 60b. It should be noted that many other configurations can be designed that comprise a payload domain and an antibody fragment that provides half-life extension.

The present invention also embodies fusion proteins comprising an Fc fragment fused to an rPEG. FIG. 83 shows a construct with a drug module at the N-terminus, followed by rPEG, fused to an antibody Fe fragment, with or without the hinge. The Fc fragment provides a long half-life and the rPEG allows the Fe fragment to be expressed in the E. coli cytoplasm in a soluble and active form. In another embodiment, an antibody Fe fragment, with or without a hinge region, is optionally fused to a drug module (e.g., IFNa, hGH, etc.) on one end and optionally fused to rPEG on the other end. The sequence between CH2 and CH3 mediates binding to FeRn, the neonatal Fc receptor (FIG. 90). Yet another embodiment includes a protein construct comprising a pair of CH3 domains (FIG. 91). Zero, one or both of the two polypeptide chains may be fused to rPEG on the N-terminal and/or C-terminal end, and fused to zero, one or more drug modules at the other end. The FcRn binding sequence can either be retained or deleted. Retention of the FcRn binding sequence yields a longer serum half-life. Still another embodiment describes a protein that is a full Fc, including CH2 and CH3 domains (with or without a hinge), fused at the C-terminus to an rPEG with the drug/pharmacophore located at the C-terminus (FIG. 92). There molecules are capable of polypeptide chain swapping, resulting in heterodimers. Yet another embodiment describes a partial Fc without a hinge and with a CH2 domain that is truncated but retains FeRn binding and with a drug/pharmacophore located at the C-terminus (FIG. 93a, FIG. 93b). FIG. 93c shows another example of a fusion protein comprising an Fc fragment fused to an rPEG.
936 illustrates a partial Fe without hinge and CH2 domain, but retaining the CH3 domain and having a drug/pharma-
cophore located at the C-terminus. Such Fe fragment does not bind FeRn but can dimerize via the CH3 domain.

[0520] Still another embodiment employs an N-terminal drug module followed by rPEG and a C-terminal Fe fragment with hinge (FIG. 101). This is a useful format for half-life extension of drug modules that can be manufactured in the E. coli cytoplasm. An alternative format for a pro-drug containing an Fe fragment is described herein (FIG. 102). The format is similar to that described in FIG. 101, with the addition of an inhibitory sequence that binds to and inhibits the drug sequence. The drug is separated from the inhibitory sequence by a cleavage site. The N-terminal inhibitory binding sequence is followed by a cleavage site, which is followed by the drug sequence. Before cleavage, the pro-drug is bound to the inhibitory sequence and thus it is inactive. Upon cleavage, the inhibitory binding sequence is gradually released and cleared, gradually increasing the amount of time that the drug is active. Assays for assessing correct folding of an Fe fragment fused to an rPEG, including SDS-PAGE on hinge disulfide formation and size exclusion chromatography on CH3 dimerization, are depicted in FIG. 104.

[0521] Antibody Fragments that Result in Slow Release

[0522] AFBTs can be engineered to release slowly from the injection site resulting in long-term drug exposure. One embodiment of the present invention includes the incorporation of an antibody fragment that binds to a molecule expressed in high abundance at the injection site. For example, such antibody fragments may bind to target anti-
gens including but not limited to collagen, hyaluronic acid, heparan sulfate, laminins, elastins, chondroitin sulfate, ker-
tane sulfate, fibronectin, and integrins. By engineering the affinity and/or avidity of the antibody fragment for its target antigen, the rate of AFBT release from the injection site can be controlled. Another embodiment includes the introduction of one or several protease sites that can be cleaved by pro-
teases at the injection site in order to control the rate of AFBT release at the injection site.

[0523] Antibody Fragments that Affect Tissue Distribution

[0524] The present invention also includes AFBTs that incorporate antibody fragments that bind to a target antigen present in a particular cell or tissue, or a particular set of tissues. These constructs can increase the therapeutic window of an active drug by achieving a local tissue-specific accumu-
lation of the AFBT. Examples include AFBTs that contain antibody fragments with specificity for tumor antigens that are overexpressed in tumor tissues or tumor microenvironment including tumor vasculature. One can choose tumor anti-
gens that are effectively internalized by cells as targets for AFBTs that include a payload with intracellular activity. For instance, AFBTs comprise an antibody fragment with specific-
ficity for a tumor antigen capable of being internalized upon binding and a cytotoxic payload. Other examples include AFBTs with specificity for viral targets.

[0525] Collagen Binding Domains (CBDs)

[0526] Another embodiment of the present invention includes the use of CBDs as domains in AFBTs and other protein drugs. Collagen is highly abundant in many tissues in particular in the extracellular space. Protein pharmaceuticals that comprise CBDs can bind to collagen at the injection site or in the vicinity of the injection site, forming a depot from which the AFBT is then slowly released. The release rate can be controlled by introducing protease sites or by choosing CBDs with a suitable affinity to collagen. By choosing a CBD with low affinity to collagen, the rate of release of the AFBT is increased. Alternatively, the rate of AFBT release can be slowed down by including CBDs that bind to collagen with very high affinity or by including multiple CBDs into an AFBT to achieve avidity. CBD sequences can be obtained from naturally occurring CBDs. Examples of proteins that bind to collagen and comprise CBDs include, but are not limited to, integrins, in particular α1β1, integrin, α1β2 integrin, angiogenesis inhibitor, collagen V, C-proteinase, decorin, fibronectin, interlinkin-2, matrix metallopro-
teases 1, 2, 9, and 13, phosphorylase, thrombospondin, biglycan, bilirubin, BM40/SPARC, MRP8, MRP-14, calin from leeches, DDR1, DDR2, fibromodulin, Gla protein, glycoprotein 46, heat shock protein 47, lumican, myelin associ-
ated glycoprotein, platelet receptors, staphylococcus aureus surface molecules and other microbial adhesion molecules, sydecan-1, tenasin-C, vitronectin, von Willebrand factor, and factor XII. Additional examples of proteins that bind collagen and contain CBDs are listed in [Di Lullo, G.A., et al. (2002) J Biol Chem, 277: 4223]. CBDs from natural proteins can be further engineered to increase their therapeutic utility and improve their stability. Immunogenicity of the CBD-containing proteins can be reduced by removing epitopes recognized by B and/or T cells. CBD sequences can also be optimized to maximize protein production and/or to improve protein solubility.

[0527] HSA-Binding Peptides in Tail

[0528] AFBTs comprising peptide sequences that increase in vivo half-life can also be achieved by utilizing peptide sequences that bind to a serum protein or to the surface of a blood cell. Examples include peptide sequences that bind to human serum albumin (HSA). Such sequences can be obtained by phage display of random peptide sequences or similar selection of screening approaches. AFBTs can contain one or more, either identical or different, copies of such peptide sequences.

[0529] Target Antigens

[0530] Yet another embodiment of the invention encompasses an antibody fragment that binds to a target antigen which is of therapeutic or diagnostic relevance. FIG. 87 illustrates a Fab fragment binding to a cell-surface target. Extension of the length of the natural linkers from the usual 2-6 amino acids to 4 to 100 or more amino acids, between the VH and the CH domains, and between the VL and the CL domains, increases the ability of one Fab to crosslink to another Fab by domain swapping, thereby forming a binding complex with higher valency, resulting in higher apparent affinity (avidity). The linker may be an rPEG or a different composition. The extended linker format allows binding with increased affinity specifically at sites with a higher density of target. The antibody fragment of an AFBT can bind to a blood component to increase the half-life of the AFPT in circulation. The antibody fragment of the AFBT can also bind to a receptor that facilitates lysosomal recycling. An example is the FcRn receptor that can re-export proteins after lysosomal uptake. Of particular interest are antibody fragments that bind with spatially or temporally-dependent affinity to a receptor that can facilitate lysosomal recycling such that the antibody fragment binds with high affinity under conditions found in a lysosome but it binds with lower affinity to the same receptor under conditions found on the cell surface. The antibody fragment of an AFBT can bind to a target antigen that is predominantly found in a disease-relevant tissue. As a result
such AFBTs can accumulate in a particular disease relevant tissue. Examples include AFBTs that bind to tumor tissue or virally-infected tissues. The antibody fragment of an AFBT can bind to a target antigen that facilitates cellular internalization in a disease-relevant tissue. Antibody fragments of an AFBT can also bind to a target antigen that facilitates uptake of the AFBT into a particular compartment of the body, for example, target antigens that facilitate oral, intranasal, mucosal, or lung uptake of an AFBT, and target antigens that facilitate the transport of the AFBT across the blood brain barrier. Examples of target antigens that are of particular interest include, but are not limited to, ILL IL4, IL6, 1112, IL13, IL17, IL23, CD22, BAFF, and TNEc.

Advantages of rPEG in AFBTs

AFBTs combine valuable properties of rPEG and antibody fragments. The rPEG portion of an AFBT results in a low overall immunogenicity of an AFBT. This is achieved by sterical shielding of the antibody fragment and other potentially immunogenic portions of an AFBT by rPEG. rPEGs are highly flexible and as a result they lack conformational epitopes. Due to their high hydrophilicity and high entropy, rPEGs have a very low inherent immunogenic potential.

The rPEG portions of an AFBT also result in a stabilization of other AFBT domains. Due to their hydrophilic nature, rPEG domains reduce aggregation of AFBTs. This greatly simplifies the formulation development for AFBTs. In addition, steric shielding by rPEG protects other portions of AFBTs from proteolysis. This is of particular importance for payloads and antibody fragments that are prone to proteolytic degradation.

Another advantage of using AFBT over a full length antibody is the minimization or elimination of undesirable effector function associated with a full length antibody. Full length antibody molecules have a number of effector functions such as antibody-dependent cell-mediated cytotoxicity (ADCC) as well as complement activation (CDC) that significantly limit their therapeutic use for indications where effector function is undesirable. For instance, many indications require an agent that binds and sequesters a molecule such as a cytokine or hormone. In general it is not desirable to utilize antibodies for such indications as their effector functions results in undesirable toxicity. Most of the antibody dependent effector function is mediated through the Fc portion. In many embodiments described herein, the AFBTs utilize the variable domains of antibodies that are responsible for target binding while replacing the Fc portion that is responsible for effector function. AFBTs can be engineered to bind and activate cell surface receptors such as death receptors DR4 and DR5. Activation can be achieved by receptor multimerization. Although full length antibodies are able to activate such receptors, they also induce toxicity caused by antibody binding to healthy cells that express the same target receptor. AFBTs can activate cell surface receptors without eliciting effector function that would cause toxicity.

Yet another advantage of rPEG is that it helps associate the two proteins that belong to the same complex, as illustrated in FIG. 86. The affinity between such proteins is often insufficient to keep them associated, but the addition of rPEG stabilizes their interaction and reduces their tendency to form polymers.

Manufacture/Production of AFBTs

The present invention also relates to the production of the AFBTs. The rPEG domain in AFBT facilitates protein folding and reduces protein aggregation. This property facilitates microbial production of AFBTs. Most antibody fragments described in the literature require refolding from inclusion bodies or secretion into the periplasmic space, which results in low production yields. In contrast, most AFBTs can be produced in soluble form at high concentration in the cytosol of microbial expression hosts. A preferred expression host for AFBTs is E. coli (FIG. 45). However, the properties of AFBTs make them suitable for expression in most microbial as well as eukaryotic expression systems. The N-terminal sequence of AFBTs can be optimized to control posttranslational processing. In particular the amino acid following the start codon can determine the subsequent processing of the N-terminal methionine [Hire, P. H., et al. (1989) Proc Natl Acad Sci USA, 86: 8247]. One embodiment includes N-terminal sequences that result in uniform products. By choosing gly, ala, pro, ser, thr, or val as amino acid following the N-terminal met, efficient processing and removal of the N-terminal met can be achieved. Another embodiment includes his, gln, glu, phe, met, lys, tyr, trp, or arg as amino acid following the N-terminal met, which prevents removal of the N-terminal met and results in heterogeneous products. AFBTs also facilitate refolding under conditions of high protein concentration where most unmodified proteins yield aggregates. The advantage of rPEG during manufacturing of AFBTs is crucial as AFBTs contain multiple protein domains that have a tendency to form aggregates. Such protein domains can be separated by rPEG sequences in the AFBTs to minimize aggregation between individual protein domains during folding.

Generation and Production of Disease-Associated and/or Patient-Specific AFBTs

The present invention also embodies the generation and production of disease-associated AFBTs, i.e. antibody fragments fused to an accessory polypeptide such as rPEG. Antibody genes can be directly isolated from infected or otherwise exposed patients [Wrammert, J., et al. (2008) Nature]. Various formats of antibody fragments fused to rPEG can be rapidly generated from such antibody genes. The resulting fusion proteins can be produced and purified using standardized protocols, enabling rapid generation of the disease-associated AFBTs. An example of the process is illustrated in FIG. 86. The rapid discovery process enables discovery and development of specific treatments in response to an acute disease outbreak such as a bacterial or viral infection. The rapid generation of fusion proteins between antibody fragments and rPEG also enables one to produce patient-specific treatments, which encompass but are not limited to isolation of immune cells from a patient; cloning of disease-specific antibody genes from the immune cells; construction and subsequent manufacturing of antibody fragment-rPEG fusions (i.e. disease-associated AFBTs); and treatment of the patient with the disease.

Polyclonal and Multiclonal AFBTs

The present invention also relates to a pharmaceutical composition comprising more than one AFBT. Such composition of AFBT mixture may have improved performance relative to the individual AFBTs. AFBT-based product can be multiclonal such that they contain two, three, or more defined AFBTs. Alternatively, AFBTs can be polyclonal containing multiple AFBTs. Such polyclonal AFBTs can be generated by cloning antibody fragments from a source that is enriched for antibodies or antibody fragments with a useful specificity. One example is cloning of antibody fragment
repertoires from an infected patient. Another example includes display libraries that have been enriched by panning against a target of interest.

[0542] rPEG Fusion Products

[0543] In one embodiment, an rPEG sequence is genetically fused to the N- or C-terminus of the human growth hormone (hGH) or human growth hormone receptor (hGH-R) gene under control of appropriate transcription and translation sequences for high level protein expression in a biological system (e.g. Escherichia coli, Pichia pastoris, CHO—S, etc). Protein expression is induced using standard techniques well known in the art for the expression system employed and purified using standard procedures (e.g. ion exchange chromatography, size exclusion chromatography, affinity chromatography, differential precipitation, phase extraction, etc). The purified protein can then be administered to human patients for therapeutic treatment of indications including, but not limited to: adult growth hormone deficiency, pediatric growth hormone deficiency, Turner syndrome, chronic renal failure, idiopathic short stature, post-transplant growth failure, hypophosphatemic rickets, inflammatory bowel disease, Noonan syndrome, pediatric Coeliac disease, AIDS wasting, obesity, aging, or other indications for which the unmodified protein has been shown to provide therapeutic benefit. The addition of the rPEG sequence confers the properties of extended serum half-life, improved patient exposure/efficacy, and/or improved manufacturing efficiency. Due to the sensitivity of the N-terminus of exenatide to maintaining in vivo efficacy, special considerations may be required to maintain the native N-terminal structure upon recombinant expression and purification, and preferred embodiments would comprise fusions of rPEG to the C-terminus of the exenatide sequence. N-terminal leader sequences which can be cleaved by proteases either in vitro or in vivo can be employed to improve manufacturing yield and/or improve delivery of active molecules in vivo. An alternative strategy would comprise mutating the internal methionine of exenatide to a compatible amino acid (eg leucine, which is present at the homologous position in the GLP-1 sequence) and use cyanoethyl bromide or similar chemical methods to remove the N-terminal leader sequence to generate the native exenatide N-terminus.

[0546] In one embodiment, an rPEG sequence is genetically fused to the N- or C-terminus of the GLP-1 gene under control of appropriate transcription and translation sequences for high level protein expression in a biological system (e.g. Escherichia coli, Pichia pastoris, CHO—S, etc). Protein expression is induced using standard techniques for the expression system employed and purified using standard procedures (e.g. ion exchange chromatography, size exclusion chromatography, affinity chromatography, differential precipitation, phase extraction, etc) well known to those skilled in the art. The purified protein can then be administered to human patients for therapeutic treatment of the following indications: type II diabetes, or other indications for which the unmodified protein has been shown to provide therapeutic benefit. The addition of the rPEG sequence confers the properties of extended serum half-life, improved patient exposure/efficacy, and/or improved manufacturing efficiency. Due to the sensitivity of the N-terminus of GLP-1 to maintaining in vivo efficacy, special considerations may be required to maintain the native N-terminal structure upon recombinant expression and purification, and preferred embodiments would comprise fusions of rPEG to the C-terminus of the GLP-1 sequence. N-terminal leader sequences which can be cleaved by proteases either in vitro or in vivo can be employed to improve manufacturing yield and/or improve delivery of active molecules in vivo.

[0547] In one embodiment, an rPEG sequence is genetically fused to the N- or C-terminus of the IL-1 RA gene under control of appropriate transcription and translation sequences for high level protein expression in a biological system (e.g. Escherichia coli, Pichia pastoris, CHO—S, etc). Protein expression is induced using standard techniques for the expression system employed and purified using standard procedures (e.g. ion exchange chromatography, size exclusion chromatography, affinity chromatography, differential precipitation, phase extraction, etc) well known to those skilled in the art. The purified protein can then be administered to human patients for therapeutic treatment of indications including, but not limited to: rheumatoid arthritis, psoriatic arthritis, psoriasis, inflammatory bowel disease, Crohn’s disease, or other indications for which the unmodified protein has been shown to provide therapeutic benefit. The addition of the rPEG sequence confers the properties of extended serum half-life, improved patient exposure/efficacy, and/or improved manufacturing efficiency.
[0548] In one embodiment, an rPEG sequence is genetically fused to the N- or C-terminus of the interferon alpha, beta, or gamma gene under control of appropriate transcription and translation sequences for high level protein expression in a biological system (e.g., *Escherichia coli*, *Pichia pastoris*, CHO—S, etc.). Protein expression is induced using standard techniques for the expression system employed and purified using standard procedures (e.g. ion exchange chromatography, size exclusion chromatography, affinity chromatography, differential precipitation, phase extraction, etc.) well known to those skilled in the art. The purified protein can then be administered to human patients for therapeutic treatment of indications including, but not limited to: hairy cell leukemia, AIDS-related Kaposi’s syndrome, pH chromosome positive CML, chronic hepatitis C, condylomata acuminate, chronic hepatitis B, malignant melanoma, follicular lymphoma, multiple sclerosis, non-Hodgkins lymphoma, osteoporosis, chronic granulomatous disease-associated infections, pulmonary multi-drug resistant tuberculosis, or other indications for which the unmodified protein has been shown to provide therapeutic benefit. The addition of the rPEG sequence confers the properties of extended serum half-life, improved patient exposure/efficacy, and/or improved manufacturing efficiency.

[0549] In one embodiment, an rPEG sequence is genetically fused to the N- or C-terminus of the G-CSF gene under control of appropriate transcription and translation sequences for high level protein expression in a biological system (e.g., *Escherichia coli*, *Pichia pastoris*, CHO—S, etc.). Protein expression is induced using standard techniques for the expression system employed and purified using standard procedures (e.g. ion exchange chromatography, size exclusion chromatography, affinity chromatography, differential precipitation, phase extraction, etc.) well known to those skilled in the art. The purified protein can then be administered to human patients for therapeutic treatment of indications including, but not limited to: chemotherapy-induced febrile neutropenia, bone-marrow transplantation, congenital neutropenia, cyclic neutropenia, idiopathic neutropenia, AIDS-associated neutropenia, myelodysplastic syndrome, or other indications for which the unmodified protein has been shown to provide therapeutic benefit. The addition of the rPEG sequence confers the properties of extended serum half-life, improved patient exposure/efficacy, and/or improved manufacturing efficiency.

[0550] In one embodiment, an rPEG sequence is genetically fused to the N- or C-terminus of the FGF21 gene under control of appropriate transcription and translation sequences for high level protein expression in a biological system (e.g., *Escherichia coli*, *Pichia pastoris*, CHO—S, etc.). Protein expression is induced using standard techniques for the expression system employed and purified using standard procedures (e.g. ion exchange chromatography, size exclusion chromatography, affinity chromatography, differential precipitation, phase extraction, etc.) well known to those skilled in the art. The purified protein can then be administered to human patients for therapeutic treatment of indications including, but not limited to: diabetes, obesity, or other indications for which the unmodified protein has been shown to provide therapeutic benefit. The addition of the rPEG sequence confers the properties of extended serum half-life, improved patient exposure/efficacy, and/or improved manufacturing efficiency.

[0551] In one embodiment, an rPEG sequence is genetically fused to the N- or C-terminus of the calcitonin gene under control of appropriate transcription and translation sequences for high level protein expression in a biological system (e.g., *Escherichia coli*, *Pichia pastoris*, CHO—S, etc.). Protein expression is induced using standard techniques for the expression system employed and purified using standard procedures (e.g. ion exchange chromatography, size exclusion chromatography, affinity chromatography, differential precipitation, phase extraction, etc.) well known to those skilled in the art. The purified protein can then be administered to human patients for therapeutic treatment of indications including, but not limited to: postmenopausal osteoporosis, Paget’s disease, hypercalcemia or other indications for which the unmodified protein has been shown to provide therapeutic benefit. The addition of the rPEG sequence confers the properties of extended serum half-life, improved patient exposure/efficacy, and/or improved manufacturing efficiency.

[0552] In one embodiment, an rPEG sequence is genetically fused to the N- or C-terminus of the parathyroid hormone (PTH) gene under control of appropriate transcription and translation sequences for high level protein expression in a biological system (e.g., *Escherichia coli*, *Pichia pastoris*, CHO—S, etc.). Protein expression is induced using standard techniques for the expression system employed and purified using standard procedures (e.g. ion exchange chromatography, size exclusion chromatography, affinity chromatography, differential precipitation, phase extraction, etc.) well known to those skilled in the art. The purified protein can then be administered to human patients for therapeutic treatment of the following indications: osteoporosis, or other indications for which the unmodified protein has been shown to provide therapeutic benefit. The addition of the rPEG sequence confers the properties of extended serum half-life, improved patient exposure/efficacy, and/or improved manufacturing efficiency.

[0553] In one embodiment, an rPEG sequence is genetically fused to the N- or C-terminus of the human chorionic gonadotropin (hCG) gene under control of appropriate transcription and translation sequences for high level protein expression in a biological system (e.g., *Escherichia coli*, *Pichia pastoris*, CHO—S, etc.). Protein expression is induced using standard techniques for the expression system employed and purified using standard procedures (e.g. ion exchange chromatography, size exclusion chromatography, affinity chromatography, differential precipitation, phase extraction, etc.) well known to those skilled in the art. The purified protein can then be administered to human patients for therapeutic treatment of indications including, but not limited to: infertility, Kaposi’s sarcoma, asthma, artherosopathy, thalassemia, osteopenia, glucocarcoma, obesity, or other indications for which the unmodified protein has been shown to provide therapeutic benefit. The addition of the rPEG sequence confers the properties of extended serum half-life, improved patient exposure/efficacy, and/or improved manufacturing efficiency.

[0554] In one embodiment, an rPEG sequence is genetically fused to the N- or C-terminus of the Fuzon (enfurvidine) gene under control of appropriate transcription and translation sequences for high level protein expression in a biological system (e.g., *Escherichia coli*, *Pichia pastoris*, CHO—S, etc.). Protein expression is induced using standard techniques for the expression system employed and purified
using standard procedures (e.g. ion exchange chromatography, size exclusion chromatography, affinity chromatography, differential precipitation, phase extraction, etc) well known to those skilled in the art. The purified protein can then be administered to human patients for therapeutic treatment of the following indications: HIV-1 infection, or other indications for which the unmodified protein has been shown to provide therapeutic benefit. The addition of the rPEG sequence confers the properties of extended serum half-life, improved patient exposure/efficacy, and/or improved manufacturing efficiency.

[0555] In one embodiment, an rPEG sequence is genetically fused to the N- or C-terminus of the leptin or leptin receptor gene under control of appropriate transcription and translation sequences for high level protein expression in a biological system (e.g. Escherichia coli, Pichia pastoris, CHO—S, etc). Protein expression is induced using standard techniques for the expression system employed and purified using standard procedures (e.g. ion exchange chromatography, size exclusion chromatography, affinity chromatography, differential precipitation, phase extraction, etc) well known to those skilled in the art. The purified protein can then be administered to human patients for therapeutic treatment of indications including, but not limited to: breast cancer, osteoarthritis, osteoporosis, septic arthritis, obesity, or other indications for which the unmodified protein has been shown to provide therapeutic benefit. The addition of the rPEG sequence confers the properties of extended serum half-life, improved patient exposure/efficacy, and/or improved manufacturing efficiency.

[0556] In one embodiment, an rPEG sequence is genetically fused to the N- or C-terminus of the TNF Binding protein 1 (TNF-BP1; p55) gene under control of appropriate transcription and translation sequences for high level protein expression in a biological system (e.g. Escherichia coli, Pichia pastoris, CHO—S, etc). Protein expression is induced using standard techniques for the expression system employed and purified using standard procedures (e.g. ion exchange chromatography, size exclusion chromatography, affinity chromatography, differential precipitation, phase extraction, etc) well known to those skilled in the art. The purified protein can then be administered to human patients for therapeutic treatment of indications including, but not limited to: rheumatoid arthritis, psoriatic arthritis, psoriasis, inflammatory bowel disease, Crohn’s disease, or other indications for which the unmodified protein has been shown to provide therapeutic benefit. The addition of the rPEG sequence confers the properties of extended serum half-life, improved patient exposure/efficacy, and/or improved manufacturing efficiency.

[0557] In one embodiment, an rPEG sequence is genetically fused to the N- or C-terminus of the glucagon gene under control of appropriate transcription and translation sequences for high level protein expression in a biological system (e.g. Escherichia coli, Pichia pastoris, CHO—S, etc). Protein expression is induced using standard techniques for the expression system employed and purified using standard procedures (e.g. ion exchange chromatography, size exclusion chromatography, affinity chromatography, differential precipitation, phase extraction, etc) well known to those skilled in the art. The purified protein can then be administered to human patients for therapeutic treatment of indications including, but not limited to: type II diabetes, juvenile diabetes, or other indications for which the unmodified protein has been shown to provide therapeutic benefit. The addition of the rPEG sequence confers the properties of extended serum half-life, improved patient exposure/efficacy, and/or improved manufacturing efficiency. Due to the sensitivity of the N-terminus of glucagon to maintain in vivo efficacy, special considerations may be required to maintain the native N-terminal structure upon recombinant expression and purification, and preferred embodiments would comprise fusions of rPEG to the C-terminus of the GLP-1 sequence. N-terminal leader sequences which can be cleaved by proteases either in vitro or in vivo can be employed to improve manufacturing yield and/or improve delivery of active molecules in vivo.

[0558] In one embodiment, an rPEG sequence is genetically fused to the N- or C-terminus of the IGF-1 gene under control of appropriate transcription and translation sequences for high level protein expression in a biological system (e.g. Escherichia coli, Pichia pastoris, CHO—S, etc). Protein expression is induced using standard techniques for the expression system employed and purified using standard procedures (e.g. ion exchange chromatography, size exclusion chromatography, affinity chromatography, differential precipitation, phase extraction, etc) well known to those skilled in the art. The purified protein can then be administered to human patients for therapeutic treatment of indications including, but not limited to: IGF-1 deficiency, bGH deficiency caused by gene deletion or anti-GH antibody formation, or other indications for which the unmodified protein has been shown to provide therapeutic benefit. The addition of the rPEG sequence confers the properties of extended serum half-life, improved patient exposure/efficacy, and/or improved manufacturing efficiency.

Depot Modules

[0559] The compositions of the present invention may optionally include a depot module. The depot module may be a naturally occurring polypeptide, an artificial polypeptide or one selected by phage display. In one embodiment, the depot module will bind directly to the polymeric matrix referred to below. The depot module can be incorporated at any position within the modified polypeptide and can be present once or in multiple copies as indicated in FIGS. 2 and 3.

[0560] The depot module can be attached to the modified polypeptide in a variety of ways. For example, in one embodiment (FIG. 4), the modified polypeptide comprises repeating units as follows: accessory polypeptide-biologically active polypeptide-depot module, biologically active polypeptide-accessory polypeptide-depot module, depot module-accessory polypeptide-biologically active polypeptide, or depot module-biologically active polypeptide-accessory polypeptide.

[0561] In another aspect of the invention, the depot module comprises a polypeptide that is specifically sensitive to serum proteases (FIG. 8). Protease cleavage of the depot module releases biologically active polypeptide. The protease sites can be engineered to be sensitive to specific proteases, such as to a serum protease, or to display different rates of protease cleavage. Thus the rate or site of release can be controlled through engineering of the protease cleavage site of the depot module. The modified polypeptide so engineered can be formulated with a polymeric matrix as described herein.

[0562] The depot module can also include the use of a “hot cysteine” to ensure site-specific modification. A “hot cysteine” is flanked by lysine residues, for example (KCKK) (SEQ ID NO. 446), where K is lysine and C is cysteine. The
proximal lysine residues shift the pKa of the cysteine, increasing its nucleophilicity and making this residue more reactive. Several groups have shown that a “hot cysteine” can be preferentially modified (greater than 90%) even in the background of 23 other cysteine residues present on the same protein. [Okten, Z., et al. (2004) Nat Struct Mol Biol, 11:884-7]. Thus, the depot module can yield site-specific, efficient modification of the accessory polypeptide or the accessory polypeptide-biotinylated active polypeptide fusion in vitro. Biotin conjugated to either of these reactive groups is commercially available.

[0563] In yet another aspect of the invention, the depot module is designed to provide a tetraivalent accessory protein-biotinylated active polypeptide fusion protein, for example, to increase target avidity and/or for slow release applications. The depot module is designed to contain an amino acid or amino acids for the site-specific conjugation of the small molecule biotin. Biotin is a common vitamin found in other counter nutritional supplements. It serves as a “co-factor” for several enzymes including those involved in the biosynthesis of fatty acids. Biotin is also extensively used in biotechnology applications because it forms a very high affinity complex with the proteins avidin, neutraavidin, and streptavidin. In this embodiment, avidin, streptavidin, or neutraavidin, which each bind to four molecules of biotin, can be used to form highly stable accessory polypeptide-biotinylated active polypeptide fusion protein tetramers (FIG. 5).

[0564] Lysine (K) and cysteine (C) residues can be modified by chemical reaction with succinimidyl esters or maleimides, respectively, under mild conditions with high yield and specificity. When the accessory polypeptide does not contain any lysine (K) or cysteine (C) residues, these can be easily incorporated into the depot module. The depot module can comprise one, two, or more lysine or cysteine residues.

[0565] The depot module can also include the use of a “hot cysteine” to ensure site-specific modification. A “hot cysteine” is flanked by lysine residues, for example (KCKK), where K is lysine and C is cysteine. The proximal lysine residues shift the pKa of the cysteine, increasing its nucleophilicity and making this residue more reactive. Several groups have shown that a “hot cysteine” can be preferentially modified (greater than 90%) even in the background of 23 other cysteine residues present on the same protein. [Okten, Z., et al. (2004) Nat Struct Mol Biol, 11:884-7]. Thus, the depot module can yield site-specific, efficient modification of the accessory polypeptide or the accessory polypeptide-biotinylated active polypeptide fusion in vitro. Biotin conjugated to either of these reactive groups is commercially available.

[0566] The addition of biotin-binding proteins such as avidin, streptavidin, or neutraavidin can induce the formation of a very stable accessory polypeptide-binding protein polypeptide tetramer. The accessory polypeptide-binding protein polypeptide tetramer can then be formulated with polymeric matrix (e.g., encapsulated into microspheres) as described below. An accessory polypeptide-binding protein polypeptide tetramer exhibits a very large hydrodynamic radius, ensuring slow release from the polymeric matrix, e.g., microspheres. An accessory polypeptide-binding protein polypeptide tetramer will also have an increased avidity towards its biological target. Because the accessory polypeptide-binding protein polypeptide tetramer can interact with four target molecules, for example, on the plasma membrane of a cell, the off-rate of the accessory polypeptide-binding protein polypeptide will be dramatically reduced. Increased avidity may enhance the biological activity or reduce the required dose of the accessory polypeptide-binding protein polypeptide.

[0567] In a further aspect of the invention, the depot module with the same active residues can be modified with polyethylene glycol instead of the reactive biotin. Of particular interest are four- and eight-armed PEG molecules. These PEG molecules can be covalently attached to depot module described herein, thus generating homogeneous tetramer and octamer species. Protein therapeutics conjugated in this manner will have a significantly enhanced avidity towards their biological targets, particularly toward cell surface proteins.

[0568] Countersions for Making Protein Precipitate

[0569] The present invention also relates to the use of countersions for regulating the solubility of the protein of interest, i.e. making protein precipitate for a depot formulation. A countersion is an ion, the presence of which allows the formation of an overall neutral or charged species. For example, in the (neutral) species NaCl the sodium cation is balanced by the chloride anion and vice versa. The mechanism of poorly water-soluble salt formation with a cation exchanger is depicted by the following formula: rPEG™+nC°→rPEG-C°m (insoluble) in which rPEG™ represents the positively charged peptide ion, whereas C° represents a negatively charged counterion. The participating amino acid residues in this reaction include Arg, Lys and the N-terminus. The mechanism of poorly water-soluble salt formation with an anion exchanger is depicted by the following formula: rPEG™+nC°→rPEG-C°m (insoluble) in which rPEG™ represents the negatively charged peptide ion, whereas C° represents a positively charged counterion. The participating amino acid residues in this reaction include Asp, Glu and the C-terminus.

[0570] In a preferred embodiment, the countersion displays mixed hydrophilic and ionic character. Thus, once the charge of the countersion is neutralized by complex formation with the protein of interest, the hydrophilic nature of the counterion dominates the resultant complex, causing its aqueous solubility to decrease significantly. In addition, the countersions must be compatible with in vivo administration within the clinical indications intended for the protein of interest in terms of acute and chronic toxicity, carcinogenicity, reproductive effects, etc. Non-limiting examples of mixed countersions suitable for this application are provided below:

[0571] Anions:

[0572] Behenate
[0573] Cholesteryl sulfate
[0574] Deoxycholate
[0575] Dodecane sulfonate
[0576] Epigallocatechin gallate
[0577] Hexadecane sulfonate
[0578] Pumatoe
[0579] Pentagulloyl Glucose
[0580] Stearate
[0581] Tannate

[0582] Cations:

[0583] Choline derivatives
[0584] Peptide countersions: eg H-Lys-(Leu)m-NH2; H-(Leu)m-NH2
[0585] Lipids:
[0586] Phosphatidylcholine
[0587] Polymeric materials:
[0588] Chitosan
[0589] Collagen
[0590] Hyaluronic Acid
[0591] Poly β-amino esters
[0592] PLA/PLGA
[0593] Poly(ethylene glycol) bis(2-aminoethyl)

[0594] In one embodiment, a protein of interest is mixed at a defined ratio with a counterion comprising both hydrophobic and charged character as described above. Upon interaction, the protein and counterion form an insoluble complex which precipitates from the solution. In a preferred embodiment, greater than or equal to 20%, 40%, 60%, or 80% of the total protein is precipitated under these conditions, which can be assessed by quantitative assay of the protein remaining in solution. Optimization of the protein:counterion ratio, inclusion of organic solvents, pH adjustment, ionic strength, and/or temperature adjustment may be employed to modulate the efficiency of the precipitation reaction. The precipitate can be separated from the liquid phase using standard methods (i.e., filtration, centrifugation), and can be stored in a dry form or as a suspension in an inert buffer. For a pharmaceutical composition, protein stability upon storage is a critical parameter for determining the viability of a given formulation. In one embodiment, the protein is stable under the defined storage conditions and formulation for greater than 1, 2, 3, 6, 12, 18, or 24 months.

[0595] The present invention also embodies the method of administering the above described protein complex into a subject in vivo. Compounds of the invention may be administered as pharmaceutical formulations including those suitable for oral (including buccal and sub-lingual), rectal, nasal, topical, transdermal patch, pulmonary, vaginal, suppository, or parenteral (including intramuscular, intraarterial, intravenous, intradermal, intrapertoneal, subcutaneous and intravenous) administration or in a form suitable for administration by aerosolization, inhalation or insufflation. In a preferred embodiment, the protein complex is administered to a subject via parenteral injection. As used herein, the term "parenteral" refers to introduction of the complex into the body not through the intestines, but rather by injection through intravenous (i.v.), intraarterial (i.a.), intraperitoneal (i.p.), intramuscular (i.m.), intravenous, intracranial, and subcutaneous (s.c.) routes. To be administered via parenteral injection (e.g., bolus injection or continuous infusion), the precipitate is suspended in a buffer compatible with the route of administration. In the preferred embodiment, the precipitate is suspended as a homogeneous suspension capable of passing through a 18, 22, 25, 26, 27, or 28 gauge needle with minimal occlusion. Milling or similar processing can be performed in order to improve the suspensions properties as well as reducing the size of the particles to enable efficient passage through larger gauge needles. Detergents or other excipients capable of modifying the surface tension, viscosity, or wetting properties of the solution can also be useful in improving the homogeneity of the precipitate suspension for injection.

[0596] The present invention also relates to the protein release rate in a depot formulation upon introduction of the precipitate into an in vivo environment via, for example, parenteral injection. The protein release rate can be approximated in vitro by suspension of the protein:counterion precipitate in an isotonic buffer (e.g. phosphate buffered saline) and measuring the concentration of soluble protein over time. A preferred embodiment uses physiological temperatures in order to better mimic the in vivo conditions, although a higher or lower temperature may be employed to modify the solubilization rate depending on the experimental setup. The optimal release rate for a given protein is dependent upon its in vivo clearance rate and mechanism, as well as the required exposure for in vivo efficacy. In order to achieve significant accumulation of the protein, the resolubilization rate should be faster than the natural clearance rate. Serum concentration of the protein is expected to be proportional to the ratio of the resolubilization rate to the clearance rate. The kinetics of the protein complex between its soluble and precipitates states is depicted in the following equation:

\[
\text{Protein}^{\text{c}}(t) = \frac{k_{\text{solubilization}}}{k_{\text{precipitation}}} \left( \frac{\text{Protein}^{\text{c}}(0)}{k_{\text{clearance}}} \right)
\]

[0597] Assuming that the rate of precipitation into the complex is negligible: \([\text{Protein}^{\text{c}}](t)\) is approximately equal to \(k_{\text{solubilization}} / k_{\text{clearance}}\) (i.e. the ratio of the resolubilization rate to the clearance rate).

[0598] The actual serum concentration achieved in vivo is also dependent upon a number of other factors including total amount of complex injected, surface area of the precipitate particles, protein absorption rate, binding of the protein to its cognate receptors, and recycling mechanisms.

[0599] The resolubilization properties of the precipitate may be modified by various treatments of the precipitate. For example, heat treatment or ultraviolet crosslinking of the counterion can be used to modify the chemical and resolubilization properties of the precipitate. The precipitate may also be formed by direct removal of solvent (e.g. spray drying, lyophilization), followed by treatment with a counterion or coating material to achieve the desired depot characteristics.

[0600] Excipients may be included in the complex formation reaction to control the rate and efficiency of complex formation, as well as to modulate the rate of resolubilization of the protein:counterion complex upon transfer to an in vivo environment. Excipients are typically unchanged, inert molecules which are included in the complex formation reaction buffer and comprise a varying degree of the final precipitate mass. Excipients may also comprise a coating applied to the surface of the precipitate particle which serves to modulate the surface area of the precipitate particle to solvent and hence modulate the resolubilization rate and/or stability of the protein. Examples of excipients include, but are not limited to, the following:

- Polymers
  - Polyethylene glycol 500
  - Polyethylene glycol 2000
  - Poloxamer 188
  - Poloxamer 8000
  - Polyethylene glycol 20000
- Polysine
- PLA/PLGA
- Detergents
  - Polysorbate 20 (TWEEN 20)
  - Polysorbate 80 (TWEEN 80)
  - Triton X-100
- Sugars/Polyalcohols
  - Glucose
  - Glycerin
  - Glycerol
  - Mannitol
Another embodiment of the present invention includes a formulation such that depot formation occurs in situ upon injection. For example, the protein and counterion are chosen such that a precipitate is formed close to physiological pH (i.e. pH 7.4). The protein and counterion are formulated at an optimal concentration ratio relative to one another, but at a pH sufficiently different from physiological pH (e.g. pH 4 or pH 10) such that no complex formation occurs. Upon parenteral injection, preferably subcutaneous or intramuscular injection, the inherent buffering capacity of the tissue causes the solution to adjust to pH 7.4, resulting in the precipitation of the protein-counterion complex at the site of injection and the resultant slow release thereof. Temperature change upon injection and complex formation of the injected protein with a natural counterion found in vivo are also methods by which a slow releasing protein depot may be formed in situ.

Production of Accessory-Linked Polypeptides

The present invention provides methods of producing biologically active polypeptide, comprising a) providing a polynucleotide sequence coding for a modified polypeptide comprising the biologically active polypeptide linked with an accessory polypeptide such that expression of the modified polypeptide in a host cell yields a higher quantity of soluble form of biologically active polypeptide as compared to expression of the biologically active polypeptide by itself; and b) causing the modified polypeptide to be expressed in said host cell, thereby producing the biologically active polypeptide. Expression of the modified biologically active polypeptides may yield at least about 100%, 200%, 500% or 1000% more soluble form of biologically active polypeptide as compared to expression of the biologically active polypeptide by itself. In some embodiments, the expression of the modified biologically active polypeptides may yield at least between 100%, and 1000% more soluble form of biologically active polypeptide as compared to expression of the biologically active polypeptide by itself.

Methods of the invention may involve culturing a cell transformed with a chimeric DNA molecule encoding an accessory polypeptide under conditions whereby the DNA is expressed, thereby producing the accessory-linked polypeptide; and extracting an expression product of the chimeric DNA molecule from the cell or culture medium.

Standard recombinant techniques in molecular biology can be used to make the accessory-linked polypeptides of the present invention. In one embodiment, a construct is first prepared containing the DNA sequence corresponding to the accessory polypeptide. For example, a gene or polynucleotide encoding the biologically active protein can be first cloned into a construct, which can be a plasmid or other vector. In a later step, a second gene or polynucleotide coding for the accessory polypeptide is cloned into the construct adjacent and in frame with the gene coding for the biologically active polypeptide. This second step can occur through a ligation or multimerization step.

In this manner, a chimeric DNA molecule coding for a modified polypeptide is generated within the construct. Optionally, this chimeric DNA molecule may be transferred or cloned into another construct that is a more appropriate expression vector. At this point, a host cell capable of expressing the chimeric DNA molecule is transformed with the chimeric DNA molecule. The transformation may occur with or without the utilization of a carrier, such as an expression vector. Then, the transformed host cell is cultured under conditions suitable for expression of the chimeric DNA molecule, resulting in the encoding of the accessory polypeptide. Methods of ligation or multimerization useful in the present invention are well known. See, Joseph Sambrook, et al., Molecular Cloning: A Laboratory Manual, 2nd ed., 1.53 (Cold Spring Harbor Laboratory Press 1989).

Several cloning strategies are envisioned to be suitable for performing the present invention, many of which can be used to generate a construct that comprises a gene coding for the accessory polypeptide of the present invention.

The vectors containing the DNA segments of interest can be transferred into the host cell by well-known methods, depending on the type of cellular host. For example, calcium chloride transfection is commonly utilized for prokaryotic cells, whereas calcium phosphate treatment, lipofection, or electroporation may be used for other cellular hosts. Other methods used to transform mammalian cells include the use of polybren, protoplast fusion, liposomes, electroporation, and microinjection (see, generally, Sambrook et al., supra). Prokaryotic or eukaryotic cells are envisioned as hosts. Accessory polypeptides can be produced in a variety of expression systems including prokaryotic and eukaryotic systems. Suitable expression hosts are for instance yeast, fungal, mammalian cell culture, and insect cells.

Useful expression vectors that can be used include, for example, segments of chromosomal, non-chromosomal and synthetic DNA sequences. Suitable vectors include, but are not limited to, derivatives of SV40 and pCDNA and known bacterial plasmids such as col E1, pCR1, pBR322, pMak1-C2, pET, pGEX as described by Smith, et al., Gene 57:31-40 (1988), pMB9 and derivatives thereof, plasmids such as RP4, plasmid DNAs such as the numerous derivatives of plasmid 1 such as NM98 9, as well as other plasmid DNA such as M13 and filamentous single stranded phage DNA; yeast plasmids such as the 2 micron plasmid or derivatives of the 2 micron plasmid, as well as centromeric and integrative yeast shuttle vectors; vectors useful in eukaryotic cells such as vectors useful in insect or mammalian cells; vectors derived from combinations of plasmids and phage DNAs, such as plasmids that have been modified to employ phage DNA or the expression control sequences; and the like. The requirements that the vectors are replicable and viable in the host cell of choice. Low- or high-copy number vectors may be used as desired.

For example in a baculovirus expression system, both non-fusion transfer vectors, such as, but not limited to pVL941 (BamHI cloning site, available from Summers, et al., Virology 84:390-402 (1978)), pVL1393 (BamHI, SmaI, XbaI, EcoRI, IvoI, XmaIII, BglII and PstI cloning sites; Invitrogen), pVL1392 (BglII, PstI, NotI, XmaIII, EcoRI, XbaI, SmaI and BamHI cloning site; Summers, et al., Virology 84:390-402 (1978) and Invitrogen) and pBlueBacIII (BamHI, BglII, PstI, Ncol and HindII cloning site, with blue/white recombinant screening, Invitrogen), and fusion transfer vectors such as, but not limited to, pAc700 (BamHI and KpnI cloning sites, in which the BamHI recognition site begins with the initiation codon; Summers, et al., Virology 84:390-402 (1978), pAc701 and pAc702-2 (same as pAc700, with different reading frames), pAc360 (BamHI cloning site 36 base pairs downstream of a polyhedrin initiation codon; Invitrogen (1995)) and pBlueBacHis A, B, C (three different
reading frames with BamHI, BglII, PstI, Nco I and Hind III cloning site, an N-terminle peptide purified with ProBond resin and cleaved with enterokinase; Invitrogen), pRS vectors and the like.

[0635] In addition, the expression vector containing the chimeric DNA molecule may include drug selection markers. Such markers aid in cloning and in the selection or identification of vectors containing chimeric DNA molecules. For example, genes that confer resistance to neomycin, puromycin, hygromycin, dihydrofolate reductase (DHFR), guanine phosphoribosyl transferase (GPT), zoxacin, and histidinol are useful selectable markers. Alternatively, enzymes such as herpes simplex virus thymidine kinase (tk) or chloramphenicol acetyltransferase (CAT) may be employed. Immunologic markers also can be employed. Any known selectable marker may be employed so long as it is capable of being expressed simultaneously with the nucleic acid encoding a gene product. Further examples of selectable markers are well known to one of skill in the art and include reporters such as enhanced green fluorescent protein (EGFP), beta-galactosidase (beta-gal) or chloramphenicol acetyltransferase (CAT).

[0636] Consequently, mammalian and typically human cells, as well as bacterial, yeast, fungi, insect, nematode and plant cells can be used in the present invention as host cells and may be transformed by the expression vector as defined herein.

[0637] Examples of suitable cells include, but are not limited to, VERO cells, HELA cells such as ATCC No. CCL2, CHO cell lines, COS cells, WI38 cells, BHK cells, HepG2 cells, 3T3 cells, A549 cells, PC12 cells, K562 cells, 293 cells, SP cells and C6 cells.

[0638] Other suitable cells that can be used in the present invention include, but are not limited to, prokaryotic host cells such as Escherichia coli, (e.g., strain DH5-alpha), Bacillus subtilis, Salmonella typhimurium, or strains of the genera of Pseudomonas, Streptomyces and Staphylococcus. Non-limiting examples of suitable prokaryotes include those from the genera: Actinoplanes; Archaeglobus; Bdellovibrio; Borelia; Chloroflexus; Enterococcus; Escherichia; Lactobacillus; Listeria; Oceanobacillus; Paracoccus; Pseudomonas; Staphylococcus; Streptococcus; Streptomyces; Thermoplasma; and Vibrio. Non-limiting examples of specific strains include: Archaeglobus fulgidus; Bdellovibrio bacteriovorus; Borelia burgdorferi; Chloroflexus aurantiacus; Enterococcus faecalis; Enterococcus faecium; Lactobacillus johnsonii; Lactobacillus plantarum; Lactococcus lactis; Listeria innocua; Listeria monocytogenes; Oceanobacillus iheyensis; Paracoccus zeaxanthinifaciens; Pseudomonas mevalonii; Staphylococcus aureus; Staphylococcus epidermidis; Staphylococcus haemolyticus; Streptococcus agalactiae; Streptomyces griseolaceus; Streptococcus mutans; Streptococcus pneumoniae; Streptococcus pyogenes; Thermoplasma acidophilum; Thermoplasma volcanii; Vibrio cholerae; Vibrio parahaemolyticus; and Vibrio vulnificus.

[0639] Further suitable cells that can be used in the present invention include yeast cells such as those of Saccharomyces such as Saccharomyces cerevisiae.

[0634] A key advantage of using bacterial expression to perform the present invention is the absence of glycosylation. While glycosylation of the accessory polypeptide increases its molecular weight and generally increases its serum half-life, quality control of glycosylated products is notoriously difficult to perform. When many glycosylation sites are present and the expression level of the protein is high, the glycosylation machinery may not be able to keep up and glycosylation is likely to be incomplete due to incomplete
processing, resulting in carbohydrate structures that are heterogeneous, which greatly complicates purification, characterization, quality control and reproducibility.

[0641] Depending on how the protein is expressed in bacteria (secreted to media, to periplasm, soluble in cytoplasm or as insoluble inclusion bodies in the cytoplasm), the product or intermediate may contain a formylated N-terminus.

[0642] Additional post-translational modifications to which accessory polypeptides or the accessory-modified polypeptides of the invention may be subjected include, but are not limited to acetylation, acylation, alkylaition, demethylation, amidation, biotinylation, formylation, gamma-carboxylation, glutamyltation, glycosylation, glycylation, attachment of heme moiety, hydroxylation, iodination, iso-prenylation, lipoylation, prenylation, myristoylation, farnesylation, geranylgeranylation, ADP-ribosylation, flavin attachment, oxidation, pegylation, attachment of phosphati-dylinositol, phosphoproteinethylation, phosphorylation, pyroglutamate formation, racemization of proline by prolyl isomerase, tRNA-mediation addition of amino acids such as arginylation, selenylation and selenocysteinylation.

[0643] Host cells containing the polynucleotides of interest can be cultured in conventional nutrient media (e.g., Ham’s nutrient mixture) modified as appropriate for activating promoters, selecting transformants or amplifying genes. The culture conditions, such as temperature, pH and the like, are those previously used with the host cell selected for expression, and will be apparent to the ordinarily skilled artisan. Cells are typically harvested by centrifugation, disrupted by physical or chemical means, and the resulting crude extract retained for further purification. Microbial cells employed in expression of proteins can be disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents, all of which are well known to those skilled in the art. Embodiments that involve cell lysis may entail use of a buffer that contains protease inhibitors that limit degradation after expression of the chimeric DNA molecule. Suitable protease inhibitors include leupeptin, pepstatin or aprotinin. The supernatant then may be precipitated in successively increasing concentrations of saturated ammonium sulfate.


[0645] Production of Crosslinked Accessory Polypeptides Crosslinked accessory polypeptides can be produced by a variety of methods. Both the non-cross-linking and the cross-linking components can be generated by chemical synthesis or using recombinant techniques. Of particular utility is the recombinant manufacture of the non-cross-linking component, which can be achieved in a variety of microbial as well as eukaryotic expression systems, for example as described above. The non-cross-linking component can be purified to remove interfering or contaminating by-products prior to cross-linking. Of particular utility are chemical crosslinkers that can be activated for coupling. Examples are shown in FIG. 22. The resulting coupling products can be further purified by a variety of methods, in particular size exclusion chromatography and ion exchange chromatography.

[0646] Multiple different non-crosslinking components can be conjugated to a crosslinking component using methods that allow the control of product structure. For instance one can use cross-linking components that carry several different reactive groups that allow different conjugation chemistries. Alternatively, one can use crosslinking components that carry protecting groups on some of their reactive groups. Such partially-protected crosslinking components can be coupled to one or more non-crosslinking components. Subsequently, one can remove the protecting groups from the crosslinking components and conjugate additional non-crosslinking components to the crosslinking component. This process can be repeated by using multiple different protecting groups that allow selective removal.

[0647] In another embodiment of the present invention, a recombinant cross-linking component may be used. The cross-linking component can be amino acids sequences that can be manufactured by recombinant technology using a variety of expression systems. For example, D-Met amino acids incorporated in the sequence of a non-cross-linking component may be conjugated to amino groups in a recombinant cross-linking component.

[0648] One preferred embodiment provides for cross-linking components that comprise one or more glutamate and/or aspartate residues, which contain side chains that can serve as reactive groups and can be effectively conjugated to non-cross-linking components that have a free amino group as reactive group. A variety of carbodiimides can be used to activate free carboxyl groups but many more chemistries are suitable. Free amino groups in the recombinant cross-linking component may be blocked by acetylation or succinylation.

[0649] Alternately, the cross-linking component can be a protein that has multiple high-affinity binding sites.

[0650] Examples are avidin, streptavidin, IgGs or IgMs. For instance one can form Crosslinked accessory polypeptides by contacting biotinylated non-cross-linking components with streptavidin, which will lead to the formation of a tetravalent complex. The process is illustrated in FIG. 25. In a similar way one could use for instance an IgM or IgG with specificity for a peptide epitope in conjunction with non-cross-linking components that comprise said peptide epitope.

[0651] The accessory polypeptides of the present invention may be assayed in order to determine the effect of which to a biologically active polypeptide. Methods of assaying biologically active polypeptides are commonly known in the art. For example, serum half-life can be measured by combining the protein with human (or mouse, rat, monkey, as appropriate) serum or plasma, typically for a range of days (i.e. 0.25, 0.5, 1, 2, 4, 8, 26 days) at 37°C. The samples for these timepoints can then be run on a Western assay and the protein is detected with an antibody. The antibody can be to a tag in the protein. If the protein shows a single band on the western, where the protein’s size is identical to that of the injected protein, then no degradation has occurred. The timepoint where 50% of the protein is degraded, as judged by Western Blots or equivalent techniques, is determined to be the serum degradation half-life or “serum half-life” of the protein.
The accessory polypeptides of the present invention may be used to modulate the expression or activity of a variety of cellular targets, including without limitation those named in the section “Biologically active polypeptides”. In some embodiments, the expression of a target will be reduced by administration of accessory polypeptides, while in other embodiments it will be increased. The accessory polypeptide may interfere with the activity of a cellular target by interaction with functional sites on the target.

Slow Release Agents

The modified polypeptides of the invention may be incorporated, encapsulated, formulated or otherwise included into compositions which allow for controlled release of the polypeptides in desired applications. Generally, the modified polypeptides of the invention may interact with the slow release agents of the invention in various manners, including not limited to covalent attachment, ionic interaction, or encapsulation within a polymer or a formulation.

Various types of slow release agents suitable for use in the present invention are described below.

Polymer Matrices

In general, microspheres are substantially spherical colloidal structures having a size ranging from about one or greater up to about 1000 microns. Microcapsules are generally described as structures in which a substance, such as a polymeric formulation, is covered by a coating of some type. The term “microsphere” may be used to describe structures that may not be readily placed into either of the above two categories or as a generic term for both. For structures that are less than about one micron in diameter the corresponding terms “nanosphere,” “nanocapsule,” and “nanoparticle” may be utilized, but these are encompassed in the terms “microsphere,” microcapsule “and microsphere,” respectively. In certain embodiments, nanospheres, nanocapsules or nanoparticles have a size of about 300, 200, 100, 50 or 10 nm.

The slow release formulations of the invention may also take the form of microcapsules, which may comprise microcapsules or microspheres.

In a microcapsule, the modified polypeptides may be centrally located within a membrane formed by the polymer molecules, or may be dispersed throughout the microcapsule. The internal structure may comprise a matrix of the modified polypeptide and a polymer excipient. Typically, the polymer which is comprised of a microsphere or microcapsule, and a membrane, respectively. In some embodiments, the polymer may comprise alginate polymers, (hydroxyethyl)methacrylated dextran polymers, or chitosan polymers may be used.

The modified polypeptides of the invention may be mixed with physiologically acceptable carriers, excipients, or stabilizers (Remington’s Pharmaceutical Sciences, 16th edition, Oslo, A., Ed., 1980), in the form of lyophilized cake or aqueous solutions. Acceptable carriers, excipients, or stabilizers for the preparation of microparticles are non-toxic to recipients at the dosages and concentrations employed, and include buffers such as phosphate, citrate, and other organic acids; antioxidants including ascorbic acid; low molecular weight (less than about 10 residues) polypeptides; proteins, such as serum albumin, gelatin, or immunoglobulins; hydrophilic polymers such as oleylvinylpyrrolidone; amino acids such as glycine, glutamine, asparagine, arginine or lysine; monosaccharides, disaccharides, and other carbohydrates including glucose, mannose, or fructose; chelating agents such as EDTA; sugar alcohols such as mannitol or sorbitol; salt-forming countersuch as sodium; and/or non-ionic surfactants such as Tween, Pluronics, or polyethylene glycol (PEG).

The microspheres of this invention are manufactured by standard techniques. For example, in one embodiment, volume exclusion is performed by mixing the active agent in solution with a polymer or mixture of polymers in solution in the presence of an energy source for a sufficient amount of time to form particles as disclosed in U.S. Pat. No. 6,268,053. The pH of the solution is adjusted to the desired pH. Next, the solution is exposed to an energy source, such as heat, radiation, or ionization, alone or in combination with sonication, vortexing, mixing or stirring, to form microparticles. The resulting microparticles are then separated from any unincorporated components present in the solution by physical separation methods well known to those skilled in the art and may then be washed.
In some embodiments, a suspension of microparticles is prepared by vigorously mixing an aqueous solution containing the modified polypeptide and an organic solution (typically dichloromethane) in which the polymer is dissolved. This water-in-oil suspension is then diluted into aqueous buffer containing an emulgent (typically poly-vinylalcohol). Finally, the microspheres are removed from this water-in-oil-in-water (W/O/W) emulsion and freeze-dried. This well-known and tested W/O/W process generally yields microspheres that are 0.1-100 µm in diameter. Microspheres of these dimensions are readily prepared as suspensions for subcutaneous injection. Alternatively, microspheres can be prepared by the single-emulsion solvent extraction/evaporation (O/W), the solid/oil/oil methods (S/O/O), and all variants of these methods described in the literature.


Microparticles are also well known and readily available to one of ordinary skill in the art from companies experienced in providing such technologies for extended release drug delivery. For example, Epic Therapeutics, a subsidiary of Baxter Healthcare Corp., developed PROMAXX®, a protein-matrix drug delivery system that produces biodegradable protein microspheres in a totally water-based process; OctoPlus developed OctoDEXTM, crosslinked dextran microspheres that release active ingredients based on bulk degradation of matrix rather than based on surface erosion; and Brookwood Pharmaceuticals advertises the availability of its microparticle technologies for drug delivery.

A search of patents, published patent applications and related publications will also provide those skilled in the art reading this disclosure with significant possible microparticle technologies. For example, U.S. Pat. Nos. 6,669,961; 6,517,859; 6,458,387; 6,395,302; 6,303,148; 6,268,053; 6,090,925; 6,024,983; 5,942,252; 5,981,719; 5,578,709; 5,554,730; 5,407,609; 4,897,268; and 4,542,025, the contents of which are incorporated by reference in their entirety, describe microspheres and methods for their manufacture. One skilled in the art, considering both the disclosure of this invention and the disclosures of these other patents could make and use microparticles for the extended release of the modified polypeptides of the invention.

Further modifications are provided by the invention. Because microparticles such as PLGA beads still release significant levels of drug immediately after administration, the present invention provides ways of ameliorating this bolus effect by including accessory polypeptides and optional depot modules as part of the modified polypeptide, as described hereinabove.

If desired, release of the therapeutic protein can be further controlled if microparticles with two or more layers are used. In one embodiment, the microspheres have an inner layer as well as an outer layer. The composition or the thickness of the outer layer may be modified to introduce differences in the time it takes to expose the modified-polypeptide-containing center of the bead. In one embodiment, microspheres may have an inner layer containing the modified polypeptide at high concentration, while the outer layer may contain a lower concentration of the modified polypeptide or no modified polypeptide. Alternatively, the outer layer varies in thickness between different microspheres. The microspheres with a thin outer layer will release modified polypeptide earlier (for example, from day 1-5), while the beads with a medium thickness of outer layer release modified polypeptide at a later time (for example, from day 4-8), and the beads with a thicker outer layer release modified polypeptide even later (for example, from day 7-11). Thus, a more constant rate of release is obtained in this embodiment.

The rate of drug release from polymeric matrix formulations can be dependent on the accessory polypeptide released either from the biologically active peptide. The accessory polypeptide significantly increases the hydrodynamic radius of the modified polypeptide. Thus the accessory polypeptide module provides means to control the rate of drug release from the microspheres. Any of the accessory polypeptides described herein can be formulated with a polymeric matrix to achieve beneficial effects in controlled-release, serum half-life stability, and other desirable properties described herein.

In a further aspect of the invention, the depot modules described herein can be designed to enhance the non-covalent interactions between the accessory polypeptide-biologically active polypeptide and the polymer matrix and to slow down the rate of release of the modified polypeptide from the matrix beads. For example, alginate is a polymer consisting of mannuronic and guluronic acid and alginate microspheres can be prepared via water/oil emulsion methods [Srivastava, R., et al. (2005) J Microencapsul, 22: 397-411], similar to the preparation of PLGA microspheres. Unlike PLGA microspheres, alginate forms highly porous microspheres from which protein release is usually complete in days. This present invention provides the use of a depot module in conjunction with the volume enhancing module and biologically active polypeptide to increase the retention of the fusion protein within alginate microspheres.

Each unit of the alginate polymer matrix contains a carboxyl group that has a –1 charge at physiological pH.

Thus alginate polymers have a large net negative charge under physiological conditions. The depot module is designed to have a basic isoelectric point (that is positively charged at physiological pH) and will therefore be retained much longer within alginate microspheres (FIG. 6). This depot module comprises a human polypeptide containing multiple lysine (K) and/or arginine (R) residues, for example. At physiological pH the lysine amino acids will carry a net positive charge, thus increasing its non-covalent binding to the alginate polymer. The depot module may include naturally occurring polypeptides or designed/engineered or selected polypeptides. Potential depot modules can be rapidly evaluated for their ability to interact with alginate. Additionally, polypeptides that bind only weakly to alginate can be combined to form repeating depot module units in order to strengthen the interactions with the polymer.

In a further embodiment of the invention, a divalent cation chelating polymer matrix (e.g. hydrogel; Lin, C. C. and Metters, A. T. (2007) J Biomed Mater Res A) is used in conjunction with a depot module that binds to divalent cations. For example, both the depot module and the chelating
polymer matrix binds to Cu^{2+}, Co^{2+} and Ni^{2+} cations and the strong non-covalent interactions between the depot module and the divalent cations serve as an efficient mechanism to achieve sustained release of the therapeutic protein from the hydrogel (FIG. 7). FIG. 46 illustrates the sustained release of accessory-modified polypeptides. For example, the depot module can incorporate poly-histidine tagged protein. Poly-histidine sequences are routinely used as purification tags, because such sequences bind tightly to Ni^{2+} cations on solid support. Alternative depot modules can be similarly designed in light of the teachings hereinabove. The depot module can be attached directly to the accessory polypeptide, instead of the biologically active polypeptide, if the poly-histidine sequence is otherwise likely to interfere with the biological activity of the therapeutically polypeptide.

Thus, any number of variations and choice of polymer matrix, accessory polypeptide, depot module and/or biologically active polypeptide can be combined to achieve the desired effect in a patient.

The present invention provides pharmaceutical compositions comprising the modified polypeptide. They can be administered orally, intranasally, parenterally or by inhalation therapy, and may take the form of tablets, lozenges, granules, capsules, pills, ampolules, suppositories or aerosol form. They may also take the form of suspensions, solutions and emulsions of the active ingredient in aqueous or nonaqueous diluents, syrups, granulates or powders. In addition, the pharmaceutical compositions can also contain other pharmaceutically active compounds or a plurality of compounds of the invention.

The compositions of the invention also can be combined with various liquid phase carriers, such as sterile or aqueous solutions, pharmaceutically acceptable carriers, suspensions and emulsions. Examples of non-aqueous solvents include propyl ethylene glycol, polyethylene glycol and vegetable oils.

More particularly, the present pharmaceutical compositions may be administered for therapy by any suitable route including oral, rectal, nasal, topical (including transdermal, aerosol, buccal and sublingual), vaginal, parental (including subcutaneous, intramuscular, intravenous and intradermal) and pulmonary. It will also be appreciated that the preferred route will vary with the condition and age of the recipient, and the disease being treated.

Extended release formulations useful in the present invention may be oral formulations comprising a matrix and a coating composition. Suitable matrix materials may include waxes (e.g., carnauba, bees wax, paraffin wax, cerasine, shellac wax, fatty acids, and fatty alcohols), oils, hardened oils or fats (e.g., hardened rapeseed oil, castor oil, beef tallow, palm oil, and soya bean oil), and polymers (e.g., hydroxypropyl cellulose, polyvinylpyrrolidone, hydroxypropyl methyl cellulose, and polyethylene glycol). Other suitable matrix tabletting materials are microcrystalline cellulose, powdered cellulose, hydroxypropyl cellulose, ethyl cellulose, with other carriers, and fillers. Tablets may also contain granulates, coated powders, or pellets. Tablets may also be multi-layered. Multi-layered tablets are especially preferred when the active ingredients have markedly different pharmacokinetic profiles. Optionally, the finished tablet may be coated or uncoated.

The coating composition may comprise an insoluble matrix polymer and/or a water soluble material. Water soluble materials can be polymers such as polyethylene glycol, hydroxypropyl cellulose, hydroxypropyl methyl cellulose, polyvinylpyrrolidone, polyvinyl alcohol, or monomeric materials such as sugars (e.g., lactose, sucrose, fructose, mannitol and the like), salts (e.g., sodium chloride, potassium chloride and the like), organic acids (e.g., fumaric acid, succinic acid, lactic acid, and tartaric acid), and mixtures thereof. Optionally, an enteric polymer may be incorporated into the coating composition. Suitable enteric polymers include hydroxypropyl methyl cellulose, acetate succinate, hydroxypropyl methyl cellulose, phthalate, polyvinyl acetate phthalate, cellulose acetate phthalate, cellulose acetate trimellitate, shellac, zein, and polyacrylates containing carboxyl groups. The coating composition may be plasticised by adding suitable plasticisers such as, for example, diethyl phthalate, citrate esters, polyethylene glycol, glycerol, acetylated glycerides, acetylated citrate esters, dibutylosebacate, and castor oil. The coating composition may also include a filler, which can be an insoluble material such as silicon dioxide, titanium dioxide, talc, kaolin, alumina, starch, powdered cellulose, MCC, or pellacrilin potassium. The coating composition may be applied as a solution or latex in organic solvents or aqueous solvents or mixtures thereof. Solvents such as water, lower alcohol, lower chlorinated hydrocarbons, ketones, or mixtures thereof may be used.

The modified polypeptides of the invention may be formulated using a variety of excipients. Suitable excipients include microcrystalline cellulose (e.g., Avicel PH102, Avicel PH101), polymeric acid, poly(ethylene acrylate, methyl methacrylate, trimethylaminooethyl methacrylate chloride) (such as Eudragit RS-30D), hydroxypropyl methylcellulose (Methocel K100M, Premium CR Methocel K100M, Methocel E5, Opadry®), magnesium stearate, talc, triethyl citrate, aqueous ethylcellulose dispersion (Surelease®). The slow release agent may also comprise a carrier, which can comprise, for example, solvents, dispersion media, coatings, antibacterial and anti fungal agents, isotonic and absorption delaying agents. Pharmaceutically acceptable salts can also be used in these slow release agents, for example, mineral salts such as hydrochlorides, hydrobromides, phosphates, or sulfates, as well as the salts of organic acids such as acetates, proprionates, malonates, or benzoates. The composition may also contain liquids, such as water, saline, glycerol, and ethanol, as well as substances such as wetting agents, emulsifying agents, or pH buffering agents. Liposomes may also be used as a carrier.

Administration via transdermal formulations can be performed using methods also known in the art, including those described generally in, e.g., U.S. Pat. Nos. 5,186,938 and 6,183,770, 4,861,800, 6,743,211, 6,945,952, 4,284,444, and WO 89/09051, incorporated herein by reference in their entirety. A transdermal patch is particularly useful embodiment with polypeptides having absorption problems. Patches can be made to control the release of skin-permeable active ingredients over a 12 hour, 24 hour, 3 day, and 7 day period. In one example, a 2-fold daily excess of a polypeptide of the present invention is placed in a non-volatile fluid. The compositions of the invention are provided in the form of a viscous, non-volatile liquid. The penetration through skin of specific formulations may be measured by standard methods in the art (for example, Franz et al., J. Invest. Derm. 64:194-195 (1975)). Examples of suitable patches are passive transfer skin patches, iontophoretic skin patches, or patches with microneedles such as Nicoderm.
In other embodiments, the composition may be delivered via intranasal, buccal, or sublingual routes to the brain to enable transfer of the active agents through the olfactory passages into the CNS and reducing the systemic administration. Devices commonly used for this route of administration are included in U.S. Pat. No. 6,715,485. Compositions delivered via this route may enable increased CNS dosing or reduced total body burden reducing systemic toxicity risks associated with certain drugs. Preparation of a pharmaceutical composition for delivery in a subdermally implantable device can be performed using methods known in the art, such as those described in, e.g., U.S. Pat. Nos. 3,992,518; 5,660,848; and 5,756,115.

Osmotic Pumps may be used as slow release agents in the form of tablets, pills, capsules or implantable devices. Osmotic pumps are well known in the art and readily available to one of ordinary skill in the art from companies experienced in providing osmotic pumps for extended release drug delivery. Examples are ALZA’s DUROSTM; ALZA’s OROSTM; Osmotica Pharmaceutical’s OsmojetSTM system; Shire Laboratories’ EnSoTrotSTM system; and AlzetTM. Patents that describe osmotic pump technology are U.S. Pat. Nos. 6,800,180; 6,535,863; 6,447,223; 6,354,000; 6,314,532; 6,341,796; 6,361,796; 6,352,721; 6,294,201; 6,284,276; 6,110,498; 5,735,776; 4,200,998; and 4,088,864, the contents of which are incorporated herein by reference. One skilled in the art, considering both the disclosure of this invention and the disclosures of these other patents could produce an osmotic pump for the extended release of the polyepptides of the present invention.

Syringe Pumps may also be used as slow release agents. Syringe pumps are known to one skilled in the art and readily available. Such devices are described in U.S. Pat. Nos. 4,976,696; 4,933,185; 5,017,378; 6,309,370; 6,254,573; 4,435,173; 4,398,908; 6,572,585; 5,298,022; 5,176,502; 5,492,534; 5,318,540; and 4,988,337, the contents of which are incorporated herein by reference. One skilled in the art, considering both the disclosure of this invention and the disclosures of these other patents could produce a syringe pump for the extended release of the polyepptides of the present invention.

In another embodiment, the modified polyepptides of the present invention are encapsulated in liposomes, which have demonstrated utility in delivering beneficial active agents in a controlled manner over prolonged periods of time. Liposomes are closed bilayer membranes containing an entrapped aqueous volume. Liposomes may also be unilamellar vesicles possessing a single membrane bilayer or multilamellar vesicles with multiple membrane bilayers, each separated from the next by an aqueous layer. The structure of the resulting membrane bilayer is such that the hydrophobic (non-polar) tails of the lipid are oriented toward the center of the bilayer while the hydrophilic (polar) heads orient toward the aqueous phase. In one embodiment, the liposome may be coated with a flexible water soluble polymer that avoids uptake by the organs of the mononuclear phagocyte system, primarily the liver and spleen. Suitable hydrophilic polymers for surrounding the liposomes include, without limitation, PEG, polyvinylpyrrolidone, polyvinylmethylether, polyethyleneoxazoline, polyethyleneoxazoline, polyethyleneoxypropyloxazoline, polyethyleneoxypropyloxazoline, polyethyleneoxazoline, polyethyleneoxazoline, polyethyleneoxypropyloxazoline, polyethyleneoxypropyloxazoline, polyethyleneoxazoline, polyethyleneoxypropyloxazoline, polyelectrolytes, and polyelectrolytes. The size of the polyepitides of the present invention is dependent on the size of the liposome.

Liposomes may be comprised of any lipid or lipid combination known in the art. For example, the vesicle-forming lipids may be naturally-occurring or synthetic lipids, including phospholipids, such as phosphatidylcholine, phosphatidylethanolamine, phosphatidic acid, phosphatidyserine, phosphatidylglycerol, phosphatidylglycerol, and sphingomyelin as disclosed in U.S. Pat. Nos. 6,056,973 and 5,874,104. The vesicle-forming lipids may also be glycolipids, cerebrosides, or cationic lipids, such as 1,2-dioleoyl-sn-(trimethylamino) propane (DOTAP); N-[1-(2,3-dioleoyloxypropyl)-N,N-dimethyl-N-hydroxyethylammonium bromide (DMRIE); N41 [(2,3-dioleoyloxypropyl)-N,N-dimethyl-N-hydroxyethylammonium bromide (DORIE); N-[1-(2,3-dioleoyloxypropyl)-N,N-trimethylammonium chloride (DOTMA); 3 [N-(N,N-dimethylaminoethane) carbamoly] cholesterol (DC-Chol); or dimethyldioctadecylammonium (DDAB) also as disclosed in U.S. Pat. No. 6,056,973. Cholesterol may also be present in the proper range to impart stability to the vesicle as disclosed in U.S. Pat. Nos. 5,916,588 and 5,874,104.

Liposomes are also well known in the art and readily available from companies experienced in providing liposomes for extended release drug delivery. For example, ALZA’s (formerly Seqens Pharmaceuticals’) STEALTHTM liposomal technology for intravenous drug delivery uses a polyethylene glycol coating on liposomes to evade recognition by the immune system; Gilead Sciences (formerly Nexstar’s) liposomal technology was incorporated into AmBisomeTM, and FDA approved treatment for fungal infections; and NOF Corp. offers a wide variety of GMP-grade phospholipids, phospholipid derivatives, and PEG-phospholipids under the tradenames COATSOMETM and SUN-BRIGHTTM.

Additional possible liposomal technologies are described in U.S. Pat. Nos. 6,759,057; 6,406,713; 6,352,716; 6,316,024; 6,204,191; 6,126,966; 6,056,973; 6,043,094; 5,963,156; 5,916,588; 5,874,104; 5,215,680; and 4,684,479, the contents of which are incorporated herein by reference. These describe liposomes and lipid-coated microbubbles, and methods for their manufacture. Thus, one skilled in the art, considering both the disclosure of this invention and the disclosures of these other patents could produce a liposome for the extended release of the polyepptides of the present invention.

Diseases amenable to treatment by administration of the compositions of the invention include without limitation cancer, inflammatory diseases, arthritis, osteoporosis, infections in particular hepatitis, bacterial infections, viral infections, genetic diseases, pulmonary diseases, diabetes, hormone-related disease, Alzheimer’s disease, cardiac diseases, myocardial infarction, deep vein thrombosis, diseases of the circulatory system, hypertension, hypotension, allergies, pain relief, dwarfism and other growth disorders, intoxications, blot clotting diseases, diseases of the innate immune system, embolism, wound healing, healing of burns, Crohn’s disease, asthma, ulcer, sepsis, glaucoma, cerebrovascular ischemia, respiratory distress syndrome, corneal ulcers, renal disease, diabetic foot ulcer, anemia, factor IX deficiency, factor VIII deficiency, factor VII deficiency, mucositis, dysphagia, thromboembolic disorder, lung embolism, infertility,
hypogonadism, leucopenia, neutropenia, endometriosis, Gaucher disease, obesity, lysosome storage disease, AIDS, premenstrual syndrome, Turner syndrome, cachexia, muscular dystrophy, Huntington’s disease, colitis, SARS, Kaposi sarcoma, liver tumor, breast tumor, glioma, Non-Hodgkin lymphoma, Chronic myelocytic leukemia; Hairy cell leukemia; Renal cell carcinoma; Liver tumor; Lymphoma; Malaria, multiple sclerosis, Kaposi sarcoma, papilloma virus, emphysema, bronchitis, periodontal disease, dementia, purpura, non small cell lung cancer, pancreatic tumor, prostate tumor, acromegaly, psoriasis, ovary tumor, Fabry disease, lysosome storage disease.

[0688] Accessory polypeptides may also comprise protease cleavage sites or other sequences that allow the modified polypeptide to be cleaved following expression. Such site or sites may be located anywhere within the modified polypeptide. For example, a protease cleavage site may be introduced between a sequence that improves solubility and another sequence comprising an affinity tag, such that the affinity tag is removed by protease treatment. Alternatively, the cleavage site may be located between the biologically active protein and the accessory polypeptide, such that a specific protease would cleave off the entire accessory polypeptide sequence. Various enzymatic methods for cleaving proteins are known. Such methods include enterokinase (DAAK) (SEQ ID NO: 447), Factor Xa (IDGR) (SEQ ID NO: 448), thrombin (LVPR/GS) (SEQ ID NO: 449), PreScission™ (LEVLFQ/GP) (SEQ ID NO: 450), TEV protease (EQLYFQ/G) (SEQ ID NO: 451), 3C protease (EITELFG/GP) (SEQ ID NO: 452), Sortase A (LPET/G) (SEQ ID NO: 453), Granzyme B (DX/N/X, M/N or S/X), inteins, SUMO, DAPase (TAGZZyme™), Acrimonas anopetidase, Aminopetidase M, and carboxypeptidases A and B. Additional methods are disclosed in Arnau et al, Prot Expr and Purif (2006) 48, 1-13.

[0689] Optimization of Production of Modified Polypeptide

[0690] Additionally, the accessory polypeptides of the invention may comprise additional sequences which allow improved folding or purification during expression. This concept is described generally in FIG. 32. For example, accessory polypeptides may be linked to affinity or solubility tags to aid in purification. Non-limiting examples include His-tag, FLAG, Streptag II, HA-tag, Softag 1, Softag 3, c-myc, 17-tag, 5-tag, elastin-like peptides, chitin-binding domain, Thioredoxin, Xylanase 10A, Glutathione S-transferase (GST), maltose binding protein (MBP), NusA, and cellulose binding protein.

[0691] Accessory polypeptides may also comprise protease cleavage sites or other sequences that allow the modified polypeptide to be cleaved following expression. Such site or sites may be located anywhere within the modified polypeptide. For example, a protease cleavage site may be introduced between a sequence that improves solubility and another sequence comprising an affinity tag, such that the affinity tag is removed by protease treatment. Alternatively, the cleavage site may be located between the biologically active protein and the accessory polypeptide, such that a specific protease would cleave off the entire accessory polypeptide sequence. Various enzymatic methods for cleaving proteins are known. Such methods include enterokinase (DAAK), Factor Xa (IDGR), thrombin (LVPR/GS), PreScission™ (LEVLFQ/GP), TEV protease (EQLYFQ/G), 3C protease (EITELFG/P), Sortase A (LPET/G), Granzyme B (DX/N/X, M/N or S/X), inteins, SUMO, DAPase (TAGZZyme™), Acrimonas anopetidase, Aminopetidase M, and carboxypeptidases A and B. Additional methods are disclosed in Arnau et al, Prot Expr and Purif (2006) 48, 1-13.

[0692] Analysis of Protein Expression

[0693] The activity of the expressed proteins may be measured to ascertain the degree of correct folding. Such assays are well known in the art depending on the specific modified polypeptide expressed. Such assays may include cell-based assays, including assays for proliferation, cell death, apoptosis and cell migration. Other possible assays may determine receptor binding of expressed polypeptides, wherein the assay may comprise soluble receptor molecules, or may determine the binding to cell-expressed receptors. Additionally, techniques such as flow cytometry or surface plasmon resonance can be used to detect binding events. Specific in vivo biological assays may be used to assess the activity of each biologically active polypeptide of the invention. For example, the properties of hGH may be determined using an ESTA bioassay, or alternatively by measuring rHG human dose-related body weight gain and bone growth, or receptor binding. Additional methods are disclosed in Dattani, M. T., et al. (1996) Horm Res 46: 64-73; Alam, K. S., et al. (1998) J Biotechnol, 65: 183-90; Clark, R., et al. (1996) J Biol Chem, 271: 21969-77; Clarg R G et al, (1996) Endocrinology. 137: 4308-15.

[0694] The present invention also relates to the composition and method of engineering the pPEG fusion products for administration into a subject. An association peptide, such as SKVLF/E (SEQ ID NO: 8) or RARADADA (SEQ ID NO: 9), which bind to another copy of the same sequence in an antiparallel orientation, can be used to create a prodrug, as shown in FIG. 88e-c. In one embodiment, the drug is protease-cleaved in the last step of manufacture, but the cleavage does not activate the drug since the two chains are still associated by the association peptides. Only after the drug is injected into a subject and the concentration is greatly reduced, the small, non-pPEG-containing protein chain leaves the complex at a rate that depends on the affinity, and is likely to be cleared via the kidney, thereby activating the p-PEG-containing drug module.

[0695] More specifically, cellular localization of expressed polypeptides of the invention can be determined by any of the methods described above. For example, an exogenous protein obtained from cells expressing the polypeptide of interest may be centrifuged in order to separate soluble expressed protein in the cytosolic fraction from insoluble protein in the inclusion bodies. If desired, the soluble (cytosolic) and insoluble (inclusion body) fractions can then be analyzed by Western Blot or similar techniques to determine the ratio of expression as soluble vs. insoluble protein.

[0696] Soluble protein in the lysate may be purified further by techniques such as anion exchange or size exclusion chromatography, techniques which can be applied preparatively or analytically (FIGS. 35-39, 47, 48, 50 and 51). Confirmation of the purity of the final product may be obtained by techniques known in the art such as SDS-PAGE, HPLC (e.g. reverse phase or size exclusion) or mass spectrometry. The purification steps may be preceded or followed by protease cleavage steps to remove affinity/solubility tags and/or the accessory polypeptide, or both. Further purification steps by any of the methods outlined above may be needed to remove, for example, the used protease from digestion mixtures. Such
steps would be well within the grasp of a person skilled in the art. Several such methods are also described in more detail in the Examples section.

[0697] Formulation, Pharmacokinetics, and Administration of rPEG Fusion Products

[0698] The present invention also relates to the composition and method of engineering the rPEG fusion products for administration into a subject. An association peptide, such as SKVILF(E) or RARADADA, which bind to another copy of the same sequence in an antiparallel orientation, can be used to create a prodrug, as shown in FIG. 88a-c. In one embodiment, the drug is protease-cleaved in the last step of manufacture, but the cleavage does not activate the drug since the two chains are still associated by the association peptides. Only after the drug is injected into a subject and the concentration is greatly reduced, the small, non-rPEG-containing protein chain leaves the complex at a rate that depends on the affinity, and is likely to be cleared via the kidney, thereby activating the r-PEG-containing drug module.

[0699] In another embodiment, the rPEG50 contains a proteolytic site and the proteolytic cleavage converts the manufactured single-chain protein into a complex of two protein chains (FIG. 89a-c). This cleavage can occur as the last manufacturing step before injection into a subject or it can occur after injection into a subject, by proteases present in the subject.

[0700] Another embodiment includes an rPEG flanked by identical receptor domains or domains having the same binding function, or domains that can bind simultaneously to the same target (FIG. 94a-c). If both receptors can bind the target simultaneously, then the binding of one receptor stabilizes binding of the second receptor, resulting in mutual stabilization of the complex, thereby increasing the apparent affinity (avidity) typically by 10 to 100-fold, but at least 3-fold, with the rPEG serving as a valency bridge that increases the effective concentration of the receptors (FIG. 94b). In one embodiment, the rPEG product is pre-loaded with a ligand (FIG. 94c). When administered into a subject, the injected product is inactive for as long as it remains bound to the ligand. When the ligand dissociates, it is likely to be rapidly cleared via the kidney, resulting in activation of the product, which has a long half-life attributed to the rPEG tail. This approach reduces the peak dose toxicity and receptor-mediated clearance, thereby extending the serum secretion halftime, as illustrated in FIG. 99.

[0701] As shown in FIG. 94, some pro-drug formats do not need a cleavage or other activation site. A single protein chain can contain two or more drug modules separated by rPEG. These modules can be of a single type or of two or more different types. This rPEG containing product is complexed with a second, complementary protein to form a receptor-ligand-receptor interaction. In this format the ligand may be dimeric or multimeric, but may also be monomeric, especially if the two drug modules are different. Both modules bind to a third protein. X and Y can be the same or different, and X and Y can be a drug module or bind to a drug module. In each case in FIGS. 94a-c, X and Y (and rPEG) comprise one protein chain, and the molecule they bind to is a separate molecule, typically protein or small molecule. It is possible to have more than two binding proteins combined in a single protein chain.

[0702] It is generally desirable in therapies that the drug be maintained at a concentration that is higher than the therapeutic dose, but lower than the toxic dose. A typical bolus injection (IV, IM, SC, IP or similar) of a drug with a short halflife results in a peak concentration that is much higher than the toxic dose, followed by an elimination phase that causes the drug concentration to rapidly drop below the therapeutic dose. FIG. 100 illustrates the drug concentration changes over time after an i.v. injection of a drug alone as compared to the drug linked to an rPEG. The drug alone is present at therapeutic concentrations for only a short time (blue line). The addition of rPEG to a drug decreases the peak concentration and thereby decreases toxicity, and increases the period of time that the drug is present at a therapeutic, non-toxic dose. The creation of a pro-drug by addition of rPEG plus a drug-binding protein can prevent the “burst release” or toxic peak dose (red line), as the drug is only gradually activated over hours and the length of time between the toxic dose and the therapeutic dose is increased compared to the other formats.

[0703] In another embodiment, the rPEG fusion products are either cleaved before administration into a subject or administered as an inactive pro-drug (i.e. cleaved after administration into a subject and activated in vivo). The process is illustrated in FIG. 96a-h. The inactivation of the drug is mediated by a binding protein that is linked to the drug by rPEG such that all three modules are manufactured as a single protein chain. If the drug is a receptor, then the binding protein may be a ligand of that receptor; if the drug is an antibody fragment, then the binding site may be an antigen. In these examples, the drug is activated by protease cleavage of a site between the two binding domains, herein termed X and Y. If protein Y is the active product, then Y retains the rPEG and the protease cleavage site needs to be close to X. If protein Y is the active product, then X retains the rPEG and the cleavage site is close to X. There can be one or multiple cleavage sites, as shown by the blue crossbars (FIG. 96a-g). The drug module includes, but is not limited to, a receptor, a ligand, one or more Ig domains, an antibody fragment, a peptide, a microprotein, or an epitope for an antibody. The protein that binds to the drug module includes, but is not limited to, a binding protein, a receptor, a ligand, one or more Ig domains, an antibody fragment, a peptide, a microprotein, or an epitope for an antibody. FIGS. 105 and 106 illustrate the conversion of an inactive protein (i.e. pro-drug) to an active protein (i.e. either an active peptide or a dAb or scFv) by a site-specific protease, either present in the serum of a subject or given before administration into a subject.

[0704] The amino acid sequence of hGH used in this experiment is:

```
1   FPTIPLSRLDNGMRLRRHLWQLADFQTQREATIPKPRQKYSPLQRHQT
22  SLCPFESSTPPQNEKETQKNSLRLSSLLLISQNSLPQVLRSYFANS
23  LVSIGSDSHYDLLKKLDLERQIQTMGKLEDGSFRTQQFDQGYSKPDTHS
24  MNDKALKNVGLCYCPRMDEKVEPTFLQVQCRSVEGSCGF.
```

[0705] A single-chain protein drug may also contain multiple bio-active peptides, which can be at the same end of rPEG or at an opposite end of rPEG (FIG. 98). These peptides can have the same activity or different activities. Having multiple peptides in a single chain increases their effective potency through binding avidity without complicating manufacturing.
EXEMPLARY EXAMPLES

Example 1
Design of Human Growth Hormone (hGH) Fused to Accessory Polypeptides

[0706] This example describes the preparation of an rPEG-hGH fusion protein with increased active, cytoplasmic yield and having improved serum half-life. Human growth hormone products typically require daily or twice-daily injections because the half-life of hGH in the serum is only about 30 minutes. Half-life extension through PEGylation is not feasible as hGH contains multiple lysines that are required for therapeutic activity and these cannot be used for conjugation. hGH is typically manufactured by expression in the cytoplasm of E. coli, where it can aggregate and form inclusion bodies containing inactive protein. Typically, these inclusion bodies are solubilized and the protein is refolded to obtain active protein. In this example, rPEG-hGH is expressed in the cytoplasm in soluble and active form, avoiding the step of refolding from inclusion bodies.

[0707] The amino acid sequence of hGH used in this experiment is:

```plaintext
PFPLGLAEPLDNLAMLRHNLQALVPTQPPFPKPIPSLQPGQT
SLCSESIPTPSNRTQMSLNLSSLLIGWLEPVRKVRSPFNS
LVTGAGDVNLQDELQEIQTQWLEDEGSRPQIQFEQTVSSQDNIS
HENQALLKCGTGLYCPQREDMDKEVFTPQLINVCSVECSGCF
```

[0708] hGH contains 191 amino acids, with a pI of 5.27 and a molecular weight of 22,130 KD. hGH contains 13 Glutamate residues, 11 Aspartate residues (24 total negative residues), 8 Lysine residues and 11 Arginine residues (19 total positive residues), for a net charge of +5 and a net charge density of +0.026 (calculated as +5/191 amino acids). This net charge density correlates with the experimental pI value of 5.27.

[0710] Various hGH-rPEG fusion proteins are designed as follows.


[0712] This design describes a polypeptide modified with a short-length accessory polypeptide and a net charge density of -0.1.

[0713] The goal of this design is to produce a protein with a net charge density of -0.1 while adding only a few amino acids. The number of charges needed to create an hGH protein with a -0.1 charge density is 14.1 (19.1 - 5 = 14.1) without accounting for the increase in total length resulting from the added charged amino acids. The addition of 16 negatively charged amino acids brings the net charge density of the modified hGH polypeptide to -0.1 (calculated as (16 + 5) / (191 + 16) amino acids).

[0714] Design 2. Construction of rPEG-Modified hGH with Net Charge Density of -0.2.

[0715] This design describes a polypeptide modified with a short-length accessory polypeptide and net charge density of -0.2.

[0716] This design incorporates an accessory protein with 41 negative charges, for a total of 46 combined negatively charged amino acid residues in the entire polypeptide. The total length of the modified polypeptide is 232 amino acids (calculated as 191 + 41 amino acids). Consequently, a charge density of -0.2 requires a total of 46 negatively charged amino acid residues (calculated as 0.2 x 232 amino acids), which means the accessory protein contains 41 negatively charged residues (calculated as 46 - 5).


[0718] rPEG J288 has the sequence (GGSGGE)14 (SEQ ID NO: 455) and contains 48 E residues (FIG. 17). When rPEG J288 was added to hGH, the total length of the modified polypeptide became 479 amino acids (calculated as 191 + 288) and the net charge became 53 (calculated as 48 + 5), thus yielding a net charge density of (calculated as 53/479 - 0.11). In this design, the accessory polypeptide itself has a net charge density of 16% due to the presence of many Glycine and Serine residues, whereas in Design 1 the accessory polypeptide is entirely composed of charged residues. As the experimental results demonstrated, this design yields highly soluble and active polypeptide. It appears that a net charge density of -0.11 can be sufficient to keep the protein in solution if the charges are spread out by the addition of Serines and/or Glycines.

[0719] This example describes the construction of a fusion gene encoding an accessory polypeptide of 144 amino acids and the sequence (GGSGGE)14 (SEQ ID NO: 455). A stuffer vector pCW00051 is constructed as shown in FIG. 16. The sequence of the expression cassette in pCW0051 is shown in FIG. 18. An insert is obtained essentially as described below for rPEG_1.288 but by annealing a synthetic oligonucleotide encoding the rPEG sequence rPEG J288 (FIG. 11) with a pair of oligonucleotides encoding an adaptor to the Kpn I site. The following oligonucleotides are used as forward and reverse primers:

```
pr_LCW0057f: AGTTAGTGAGCGGAGGAGGCTGGGGAGAG
[SEQ ID NO: 456]
```

```
pr_LCW0057r: ACCTCCCTTCGCGGAGGGCCTCGCGGAGAG
[SEQ ID NO: 457]
```

The following oligonucleotides are used as stopper primers:

```
pr_3PmIstopperFor: AGGTTCGCCCTCTACCGAGGAGAGAG
[SEQ ID NO: 458]
```

```
pr_3PmIstopperRev: CCGCGAGGAGAGAGAGAGAG
[SEQ ID NO: 459]
```

[0720] The following oligonucleotides are used as stopper primers:

```
pr_3PmIstopperFor: AGGTTCGCTCTACGAGGAGAGAG
[SEQ ID NO: 456]
```

[0721] This example describes a polypeptide modified with a long hydrophilic accessory polypeptide of 288 amino acids comprising 25% glutamate residues. rPEG L288 has the sequence (SSESSESEESSE)14 (SEQ ID NO: 40) and contains 72 E residues. When rPEG L288 is added to hGH, the total length of the fusion becomes 479 amino acids (calculated as 191 + 288 amino acids) and the net charge becomes 77
(calculated as 72±5), yielding a net charge density of 0.16 (calculated as 77/479 amino acids). As the experimental results described below demonstrated, this design with a net charge density of -0.16 showed excellent solubility and the protein was active. Some gel formation was observed at low temperatures but this did not appear to be a problem.

[0723] This section describes the construction of a codon optimized gene encoding a accessory polypeptide, rPEG_L.288 with 288 amino acids and the sequence (SSSESSSESSSSE)24 (SEQ ID NO: 460). A stuffer vector pCW0150 which is based on a pET vector and includes a T7 promoter is constructed as shown in FIG. 9. The vector encodes a Flag sequence followed by a stuffer sequence that is flanked by BsaI, BbsI, and Kpnl sites. The stuffer sequence was followed by a His6 tag (SEQ ID NO: 1) and the gene of green fluorescent protein (GFP). GFP was chosen as the biologically active protein and may be used in imaging applications or as a selection marker. The stuffer sequence contains stop codons and thus E. coli cells carrying the stuffer plasmid pCW0150 form non-fluorescent colonies. The stuffer vector pCW0150 was digested with BsaI and Kpnl. A codon library encoding accessory polypeptides of 36 amino acid length was constructed. The accessory polypeptide was designated rPEG_L.36 and had the amino acid sequence (SSSESSSESSSSE) (SEQ ID NO: 461). The insert was obtained by annealing synthetic oligonucleotide pairs encoding the amino acid sequence SSSESSSESSSSE (SEQ ID NO: 462) as well as a pair of oligonucleotides that encode an adapter to the Kpnl site. The following oligonucleotides were used as forward and reverse primers:

```
pr_LCW0148for: TTCTAOGARTCTGARTCTGARTCTGARTCTAOG

pr_LCW0148rev: AGAAGACTTCRTGARTCTGARTCTGARTCTGARTCTG
```

[0724] The following oligonucleotides are used as stopper primers:

```
pr_Kpn1stopfor: TTCTCRTCCTCCTACCTCGAGG

pr_Kpn1stopRev: CCTCGAGG
```

[0725] By varying the ratio of forward/reverse primers to stopper primers, the size of the resulting PCR products can be controlled. The annealed oligonucleotide pairs were ligated, which resulted in a mixture of products with varying length that represents the varying number of (SSSESSSESSSSE) (SEQ ID NO: 362) repeats. The product corresponding to the length of rPEG_L.36 was isolated from the mixture by agarose gel electrophoresis and ligated into the BsaI/Kpnl digested stuffer vector pCW0150. Cells transformed with vector showed green fluorescence after induction which shows that the sequence of rPEG_L.36 had been ligated in frame with the GFP gene. The resulting library was designated LCW0148. Isolates (e.g., 312 isolates) from library LCW0148 were screened for high level of fluorescence. Isolates (e.g., 70 isolates) with strong fluorescence were analyzed by PCR to verify the length of the rPEG_L segment and 34 clones were identified that had the expected length of rPEG_L.36. This process resulted in a collection of 34 isolates of rPEG_L.36 showing high expression and differing in their codon usage. A plasmid mixture was digested with BsaI/NcoI and a fragment comprising the rPEG_L.36 sequence and a part of GFP was isolated. The same plasmid mixture was also digested with BsaI/NcoI and the vector fragment comprising rPEG_L.36, most of the plasmid vector, and the remainder of the GFP gene was isolated. Both fragments were mixed, ligated, and transformed into BL21Gold(DE3) and isolates were screened for fluorescence. This process of dimerization was repeated two more rounds. During each round, the length of the rPEG_L gene was doubled and ultimately a collection of genes that encode rPEG_L.288 were obtained. The rPEG_L.288 module contains segments of rPEG_L.36 that differ in their nucleotide sequence despite having identical amino acid sequence. Thus, internal homology in the gene is minimized and as a result the risk of spontaneous recombination is reduced. E. coli BL21Gold(DE3) harboring plasmids encoding rPEG_L.288 were cultured for at least 20 doublings and no spontaneous recombination was observed.

```
pr_Kpn1stopRev: CCTCGAGG
```

[0726] By varying the ratio of forward/reverse primers to stopper primers, the size of the resulting PCR products can be controlled. The insert was used to generate a plasmid encoding the rPEG J288-modified GFP and cells expressing this plasmid in a fashion similar to rPEG_L.288-modified GFP (FIG. 10). A similar insert was used to generate a plasmid encoding the rPEG J288-modified hGH and rPEG J288-modified GLP1 and cells expressing this plasmid in a fashion similar to rPEG_L.288-modified GFP (FIG. 12).

[0727] This design describes a polypeptide modified with a long hydrophilic accessory polypeptide of 288 amino acids comprising 33% glutamate residues. rPEG_K288 has the sequence (GEGGGEYEGG)23 (SEQ ID NO: 463) and contains 96 E residues. When rPEG_K288 was added to hGH, the total length of the fusion became 479 amino acids (calculated as 191+288) and the net charge became 101 (calculated as 96+5), yielding a net charge density of 0.21 (calculated as 101/479). As predicted and confirmed by the experimental results described below, this design with a net charge density of -0.21 showed the highest degree of solubility and the protein was active. No gel formation was observed at the temperature or salt concentrations tested.

[0728] This section describes the construction of a fusion gene encoding an accessory polypeptide of the sequence (GEGGGEYEGG)23 (SEQ ID NO: 466). An insert is obtained essentially as described for rPEG_L.288 but by annealing a synthetic oligonucleotide encoding the rPEG sequence rPEG_K288 with a pair of oligonucleotides encoding an adapter to the Kpnl site. The following oligonucleotides were used as forward and reverse primers:

```
pr_LCW0147for: AAGTGAAAGTCAAGAGGAGG

pr_LCW0147rev: GCACTCCGAGG
```
[0729] The following oligonucleotides are used as stopper primers:

\[ \text{SEQ ID NO: 456} \]
\[ \text{pr}_3\text{KpnIstopperFor: AGGTTCGCTTCTACGAGGAGGAC} \]
\[ \text{SEQ ID NO: 459} \]
\[ \text{pr}_3\text{KpnIstopperRev: CCTCGAGTAAGAGGCG} \]
\[ \text{pr}_3\text{KpnIstopperRev: CCTCGAGTAAGAGGCGA} \]

[0730] By varying the ratio of forward/reverse primers to stopper primers, the size of the resulting PCR products can be controlled. The annealed oligonucleotide pairs were ligated, which resulted in a mixture of products with varying length that represents the varying number of (SSSSESSSSESSSESSSESSSSE) repeats. The product corresponding to the length of rPEG_L136 was isolated from the mixture by agarose gel electrophoresis and ligated into the Bsal/KpnI digested stuffer vector pCW0150. Cells transformed with vector showed green fluorescence after induction which shows that the sequence of rPEG_L136 had been ligated in frame with the GFP gene. The resulting library was designated LCW0148. Isolates (e.g., 312 isolates) from library LCW0148 were screened for high level of fluorescence. Isolates (e.g., 70 isolates) with strong fluorescence were analyzed by PCR to verify the length of the rPEG_L segment and 34 clones were identified that had the expected length of rPEG_L136. This process resulted in a collection of 34 isolates of rPEG_L136 showing high expression and differing in their codon usage. A plasmid mixture was digested with Bsal/NcoI and a fragment comprising the rPEG_L136 sequence and a part of GFP was isolated. The same plasmid mixture was also digested with Bsal/NcoI and the vector fragment comprising rPEG_L136, most of the plasmid vector, and the remainder of the GFP gene was isolated. Both fragments were mixed, ligated, and transformed into BL21Gold(DE3) and isolates were screened for fluorescence. This process of dimerization was repeated two more rounds. During each round, the length of the rPEG_L gene was doubled and ultimately a collection of genes that encode rPEG_L288 were obtained. The rPEG_L288 module contains segments of rPEG_L136 that differ in their nucleotide sequence despite having identical amino acid sequence. Thus, internal homology in the gene is minimized and as a result the risk of spontaneous recombination is reduced. E. coli BL21Gold(DE3) harboring plasmids encoding rPEG_L288 were cultured for at least 20 doublings and no spontaneous recombination was observed.

[0731] E. coli BL21Gold(DE3) cells harboring plasmids encoding rPEG_L288 were grown overnight in Terrific Broth (TB) and diluted 200-fold into fresh TB the following day. When the culture reached an A600 nm of 0.6, expression of rPEG_L288-GFP was induced with the addition of IPTG to 0.2 mM final concentration. The cells were harvested following 18 hr at 26°C and can be stored at −80°C until further processing. The cells were resuspended in 90 ml of 50 mM Tris-HCl, 200 mM sodium chloride, 0.1% Tween-20, 10% glycerol, pH 8.0 per liter of bacterial culture. Protease inhibitors, lysozyme (final 20 mg/ml), and benzamide nuclease were added to the bacterial suspension prior to lysis. The cells were lysed by sonication on ice for four minutes followed by heat treatment at 80°C for 20 min. The lysate was subsequently cooled on ice and centrifuged for 20 min at 15000 rpm in a Sorvall SS-34 rotor. The soluble recombinant protein was purified by immobilized metal ion affinity chromatography (IMAC) of the supernatant. The protein was further purified by ion exchange chromatography (IEC) and gel filtration chromatography. Optionally, the protein can be further purified by a column with immobilized anti-FLAG antibody using standard techniques. Purity and homogeneity of the protein was assessed using standard biochemical methods including SDS-PAGE, native-PAGE, analytical gel filtration chromatography, light scattering, and mass spectrometry. A purity of at least 90% was obtained. Additionally, the modified polypeptides rPEG_L288-hGH1 and rPEG_L288-GLP1 were obtained in a similar manner.

[0732] The purity of rPEG_L288-modified GFP was confirmed by SDS-PAGE (FIG. 36), analytical reverse phase HPLC (FIG. 38). The apparent molecular weight of rPEG_L288-modified GFP was also measured as previously described (FIG. 41). FIG. 49 illustrates the increase in apparent molecular weight observed upon linking a biologically active polypeptide (GLP1) to rPEG_L288 accessory polypeptide. The in vivo stability in rat and human serum was determined as shown in FIG. 42. rPEG is stable in rat and human serum, and rPEG288 has a half-life of about 10 to 20 hours in rats (FIG. 43). Little immunogenicity in vivo experiments could be observed with this polypeptide (FIG. 44).


[0734] This design describes a polypeptide modified with a long hydrophilic accessory polypeptide of 288 amino acids comprising 33% glutamate residues. rPEG_K288 has the sequence (GEGGGEGGGE)32 and contains 96 E residues. When rPEG_K288 was added to hGH, the total length of the fusion became 479 amino acids (calculated as 191+288) and the net charge became 101 (calculated as 96+5), yielding a net charge density of 0.21 (calculated as 101/479). As predicted and confirmed by the experimental results described below, this design with a net charge density of ~0.21 showed the highest degree of solubility and the protein was active. No gel formation was observed at the temperature or salt concentrations tested.

[0735] This section describes the construction of a fusion gene encoding an accessory polypeptide of the sequence (GEGGGEGGGE)32. An insert is obtained essentially as described for rPEG_L288 but by annealing a synthetic oligonucleotide encoding the rPEG_L sequence rPEG_K288 with a pair of oligonucleotides encoding an adaptor to the KpnI site. The following oligonucleotides were used as forward and reverse primers:

\[ \text{pr}_3\text{LCW0147for: AAGTTGAACGAGGCGGCGGCGGA} \]
\[ \text{pr}_3\text{LCW0147rev: ACCCTCCTCCNCWCCYNCYCCTC} \]

[0736] A plasmid harboring hGH, N-terminally fused to 288 amino acids of rPEG-K288 and, having the repetitive sequence (GEGGGEGGGE)32 (SEQ ID NO: 466) and a TEV protease cleavage site (ENLYFQK) (SEQ ID NO: 469), following the T7 promoter (i.e. T7 promoter-hGH-TEV-rPEG_K288), is transformed into BL21(DE3)-star E. coli strain and is grown as described above. Cells are collected by centrifugation and the cell pellet is resuspended in 50 ml Buffer containing 50 mM Tris pH-8.0, 100 mM NaCl, Protease inhibitors, 10% (v/v) glycerol, 0.1% Triton X-100 and DNase. Cells are disrupted using an ultrasonic sonicator cell disruptor, and cell debris is removed by centrifugation at
having the repetitive sequence (GEGGGEDEGE)_{32} (SEQ ID NO: 466), resulting in a vector containing pPEG-K288-TEV-hGH, is transformed into the BL21(DE3)-star E. coli strain (Novagen) and grown essentially as described in Example 3. Cells are collected and disrupted essentially as described in Example 3. hGH is found in the cell fluid, while pPEG-K288 remains on the column. The pooled flow-through is loaded on the anion-exchange (Q-Sepharose, Pharmacia), washed with buffer A (25 mM Tris pH=8.0) and eluted from the column using a linear gradient of the same buffer with 1M NaCl. The eluted hGH protein is pooled, dialyzed against buffer A, concentrated, and purified by size-exclusion chromatography (SEC) as the final purification. Protein purity is estimated to be above 98%.

[0741] Cells were collected by centrifugation and the cell pellet was resuspended in 50 ml Buffer containing 50 mM Tris pH=8.0, 100 mM NaCl, Proteinase inhibitors, 10% (v/v) glycerol, 0.1% Triton X-100 and DNase. Cells were disrupted using an ultrasonic sonicator cell disruptor, and cell debris was removed by centrifugation at 15000 RPM at 4°C. Cellular supernatant was applied on an anion-exchanger (Q-Sepharose, Pharmacia), washed with buffer A (25 mM Tris pH=8.0) and eluted from the column using a linear gradient of the same buffer with 1M NaCl. The eluted fusion protein was pooled, dialyzed and loaded on the anion-exchanger (Q-Sepharose, Pharmacia), washed with buffer A (25 mM Tris pH=8.0) and eluted from the column using a shallow linear gradient of the same buffer with 1M NaCl. The eluted fusion protein was pooled, dialyzed against buffer A, concentrated, and purified by size-exclusion chromatography (SEC) as the final purification. Protein purity was estimated to be above 98%, which was unexpected considering only ion exchange and SEC had been used to purify the protein in an pPEG-specific manner from whole cells. The quantity of eluted fusion protein was determined by SDS-PAGE analysis and by measurement of total protein concentration. A high quantity of eluted fusion protein reflects higher solubility of the fusion protein relative to hGH alone.

Testing of Accessory Polypeptide-Modified hGH in an hGH Receptor Binding Assay

[0742] A plasmid harboring hGH, N-terminally fused to 288 amino acids of pPEG-K, having the repetitive sequence (GEGGGEDEGE)_{32} (SEQ ID NO: 466) following the T7 promoter, is prepared essentially as described in Example 1 but replacing the hGH coding sequence with a domain antibody coding sequence. The domain antibody coding sequence is provided in Dumoulin, M. et al., Protein Science 11:500-505 (2002). Amino acid residues 1-113 of clone dAb-Lys3 are incorporated into the pPEG construct. This sequence is a domain antibody that binds to hen egg lysozyme with a Kd of 11 nM. This domain antibody sequence yields only inclusion bodies composed of inactive protein when expressed in the cytoplasm of E. coli in the absence of additional solubility enhancing sequences; alternatively it can be expressed in active form in the periplasm if guided by a leader sequence. The VHH dAb sequence is inserted upstream of the pPEG-K288 sequence and the resulting plasmid is transformed into
BL21(DE3)-star E. coli strain (Novagen). Cells are grown, collected and disrupted essentially as described above. The cellular supernatant is applied on an anion-exchange (Q-Sepharose, Pharmacia), washed with buffer A (25 mM Tris pH 8.0) and protein is eluted from the column using a linear gradient of the same buffer with 1 M NaCl. Protein chutes at about 500 mM NaCl. The eluted fusion protein is pooled, dialyzed and loaded on the anion-exchange (Q-Sepharose, Pharmacia), washed with buffer A (25 mM Tris pH 8.0) and eluted from the column using a shallow linear gradient of the same buffer with 1 M NaCl. The eluted fusion protein is pooled, dialyzed against buffer A, concentrated, and purified by size-exclusion chromatography (SEC) as the final purification. Protein purity is estimated to be above 98%.

Example 3

Expression of Human Growth Hormone (hGH)
Fused to CBD and rPEG_K288

This example describes the preparation of a CBD-TEV-rPEG_K288-hGH fusion protein. After digestion with TEV protease, and purification, the final protein product is rPEG_K288-hGH.

[0748] A pET-series vector was constructed with T7 promoter, which expresses a protein containing cellulose binding domain (CBD) at the N-terminus, followed by a Tomato Etch Virus (TEV) protease cleavage site, followed by the hGH coding sequence, and by the rPEG_K288 coding sequence: CBD-TEV-rPEG_K288-hGH. The rPEG_K288 has the repetitive sequence (GGGGEQGEQGEQGEQGEQ). The CBD sequence used is shown in Swissprot file Q06851 and the purification of CBD fusion proteins is described in Olin, K. et al. (2005) Proteomics 5:1806. The sequence of the TEV cleavage site is ENLYFQGX; G was used in the X position. This construct was transformed into BL21(DE3)-star E. coli strain and grown essentially as described above, except that the CBD sequence was introduced N-terminal to the rPEG sequence. Cells were collected and disrupted essentially as described above. The cellular supernatant was applied on beaded cellulose resin (Perloza 100), washed with buffer A (25 mM Tris pH 8.0) and eluted from the column with 20 mM NaOH, pH 0. Protein purity was estimated to be above 90%.

[0749] rPEG_J288 has the sequence GGSGGE (SEQ ID NO: 62) and contains 48 E residues and can therefore be used to increase the charge density of IFNα2a. When rPEG_J288 is added to IFNα2a, the total length of the fusion protein is 453 amino acids (calculated as 165+288 amino acids) and the net charge is 51 (48+3), yielding a net charge density of 0.11 (calculated as 51/453), which allows expression of IFNα2a in soluble, active form in the cell cytoplasm. The construct, expression and purification methods are prepared and carried out essentially as described in Example 1. The fusion protein proved to be soluble and active, although some tendency towards aggregation could still be observed under some conditions. This can be overcome by increasing the net charge density to keep the protein in solution. rPEGs of the same size but with more charges, such as rPEG_L (288AA, 32% E) and rPEG_K (288AA, 33% E) may be used to make the IFNα2a-rPEG fusion protein completely soluble and actively folded. For IFNα2a-rPEG_K288 the number of negatively charged amino acid residues in the accessory polypeptide is 96, such that the total net charge of the fusion protein is 99 (calculated as 96+3), which means that the net charge density is 0.218 (calculated as 99/(288+165)).

Example 4

Expression of CBD-Human Growth Hormone (hGH)
Fused to rPEG_K288

[0750] This example describes the preparation of CBD-rPEG_K288-TEV-hGH, fusion protein. After TEV protease digestion and purification, the final protein product is pure hGH.
[0751] G-CSF has a length of 174 amino acids, a pl of 5.65 and a molecular weight of 18672.29 corresponding to the sequence:

[TPLGPFSSLPQSFSPGLECLSQKQVQQGGAQKSDKLYCATVHCPEELVLL
GHSFLGPWPLSNESPQALQAGHGSQGGFLQYGLQIALGKQIPEL
PTLTQLDQVADPATTWQQHEGLNAPALQPTQGAMPFACAPFQERAGG
VLVASHIQSFLEYSVFLVHLAQPP.]

Example 5

Expression of rPEG_K288-VHH, a Domain Antibody that Binds Lysozyme

[0752] This example describes the preparation of rPEG_K288 fused to a VHH domain antibody (dAb).

[0753] A plasmid harboring hGHI, N-terminally fused to 288 amino acids of rPEG-K, having the repetitive sequence (GEGGE)_{32} following the T7 promoter, is prepared essentially as described in Example 1 but replacing the hGHI coding sequence with a domain antibody coding sequence. The domain antibody coding sequence is provided in Example 1 and is described in the paper. Amino acid residues 1-113 of clone dAb-Lys3 are incorporated into the rPEG construct. This sequence is a domain antibody that binds to hen egg lysozyme with a Kd of 11 nM. This domain antibody sequence yields only inclusion bodies composed of lysozyme protein when expressed in the cytoplasm of E. coli in the absence of additional solubility enhancing sequences, although it can be expressed in active form in the periplasm if guided by a leader sequence. The VHH dAb sequence is inserted upstream of the rPEG_K288 sequence and the resulting plasmid is transformed into BL21(DE3)-star E. coli strain (Novagen). Cells are grown, collected and disrupted essentially as described above. The cellular supernatant is applied on an anion-exchange (Q-Sepharose, Pharmacia), washed with buffer A (25 mM Tris pH=8.0) and protein is eluted from the column using a linear gradient of the same buffer with 1M NaCl. Protein elutes at about 500 mM NaCl. The eluted fusion protein is pooled, dialyzed and loaded on the anion-exchange (Q-Sepharose, Pharmacia), washed with buffer A (25 mM Tris pH=8.0) and eluted from the column using a shallow linear gradient of same buffer with 1M NaCl. The eluted fusion protein is pooled, dialyzed against buffer A, concentrated, and purified by size-exclusion chromatography (SEC) as the final purification. Protein purity is estimated to be above 98%.

[0754] The resulting VHH-rPEG_K288 protein is assayed by ELISA for the ability to bind to the target, hen egg lysozyme (Sigma). The protein was shown to bind specifically to lysozyme but not to three control proteins, demonstrating that the addition of rPEG_K288 to the VHH caused it to express in soluble and active form in the cytoplasm of E. coli.

Example 6

Expression of IFNα2a-rPEG

[0755] This example describes the preparation of an IFNα2a-rPEG fusion protein.

[0756] Interferon alpha 2a has 165 amino acids, a pl of 5.99, and a molecular weight of 19241.62 corresponding to the sequence:

[CDLPQTHSLGSRRVTMLLAQMKRSILFECILKFIDHPFQREEGHQFQKA
ETIPYHLIQQIHNLFSTFSDSAANRGLTLKQFLLYTQQQLNQDLRACV
OQYVGETPILMKMSILAVRKYQRITLTYLRSKYSPCAMEFVRAHLMS
FILSTHIQSLKKE.

[0757] rPEG_J288 has the sequence GGSGGE (SEQ ID NO: 61) and contains 48 E residues. When rPEG_J288 is added to GCSF, the total length of the fusion becomes 174+288=462 amino acids and the net charge becomes 48+3=61, yielding a net charge density of 0.11 (calculated as 51/462). This charge is expected to be sufficient to switch GCSF from >80% aggregation to >80% soluble protein. A higher charge density of 0.15 or 0.2 can also be used.

[0758] The addition of 15 negatively charged amino acids to interferon alpha brings the net charge density of the fusion protein to -0.1 (calculated as (15+3)/(165+15)), which is desirable for increased solubility. A higher charge density of -0.2 charges/amino acid may be obtained by including about 26 additional negatively charged amino acid residues in the protein, for a total of 41 negatively charged amino acid residues. Since the combined length is 206 amino acids (calculated as 165+41), a charge density of -0.2 requires 41 total negatively charged amino acid residues (calculated as 0.2x205 amino acids), which means the accessory protein may include 38 negatively charged residues (calculated as 41-3).

[0759] By similar reasoning, to reach a net charge density of +0.1, the accessory polypeptide may include 15+6=21 positively charged amino acids.

[0760] rPEG_J288-GFP—represents the protein sequence composed of the repetitive sequence (GGSGGE)_{68} (SEQ ID NO: 472) fused to GFP protein sequence.

[0761] rPEG_K288-GFP—represents the protein sequence composed of the repetitive sequence (GEGGE)_{32} (SEQ ID NO: 473) fused to GFP protein sequence.

[0762] rPEG_L288-GFP—represents the protein sequence composed of the repetitive sequence (SSSESSSSS)_{24} (SEQ ID NO: 474) fused to GFP protein sequence.

[0763] rPEG_0336-GFP—represents the protein sequence composed of the repetitive sequence (SSSESSSSSSS)_{24} (SEQ ID NO: 475) fused to GFP protein sequence.

[0764] rPEG_P320-GFP—represents protein sequence composed of the repetitive sequence (SSSESSSSSSS)_{24} (SEQ ID NO: 476) fused to GFP protein sequence.

[0765] For GCSF, the addition of 14 negatively charged amino acids brings the net charge density of the fusion protein to about -0.1 (calculated as (1+4)/(174+14)), which is typically desirable for solubility. The preferred charge density of -0.2 would require about 26 additional negatively charged amino acid residues, for a total of 41 negatively charged amino acid residues, since combined length is 217 amino acids (calculated as 174+43). An alternatively chosen charge density of -0.2 requires 43 total negatively charged amino acid residues (calculated as 0.2x217 amino acids), which means the accessory protein should contain 39 negatively charged residues (calculated as 43-4).
In another alternative design, an accessory protein with positively charged amino acids to reach a net charge density of +0.1 is desired, which requires a net positive charge of +21. This could be achieved by addition of an accessory protein containing 25 positive charges (calculated as 25±21), resulting in a combined fusion protein length of 209 amino acids.

**Experimental Results:**

rPEG_J288 has the sequence GGSGGE and contains 48 E residues. When rPEG_J288 is added to GCSF, the total length of the fusion becomes 174+288=462 amino acids and the net charge becomes 48+35=83, yielding a net charge density of 0.11 (calculated as 51/462). This charge is expected to be sufficient to switch GCSF from >80% aggregation to >80% soluble protein. A higher charge density of 0.15 or 0.2 can also be used.

Using standard molecular biological techniques any of the examples provided herein may be modified to use a different rPEG module fused to the therapeutic protein. The present inventors have shown that a net charge density of 0.1 provides improved solubility of proteins in the cytoplasm (e.g., with GFP, hG11 and II/Na2a), whereas a net charge density of around 0.2 provides highly soluble proteins with no tendency towards aggregation.

Example 8

Solubility of Different rPEG Sequences Fused to GFP when Recombinant-Expressed in the Cytoplasm of E. coli

The following protein sequences were prepared and tested in this experiment:

rPEG with the sequence (GGSGGE)_{20} (SEQ ID NO: 455) was fused to green fluorescent protein (GFP) yielding clone LCW0066. The fusion protein also carried an N-terminal Flag tag and a His6 tag (SEQ ID NO: 1) between rPEG and GFP. The fusion protein was expressed in E. coli using a standard T7 expression vector. Cells were cultured in LB medium and expression was induced with IPTG. After expression, the cells were lysed by heating the pellet to 70°C for 15 min. Most E. coli proteins denatured during this heat step and could be removed by centrifugation. The fusion protein was purified from the supernatant by IMAC chromatography followed by purification by immobilized anti-Flag (Sigma). The fusion proteins were analyzed by size exclusion chromatography (SEC) using 10/30 Superdex-200 (GE, Amersham). The column was calibrated with globular proteins (diamonds). The fusion protein comprising rPEG_J288 and GFP eluted significantly earlier from the column than predicted based on its calculated molecular weight. Based on the calibration with globular proteins SEC measured an apparent molecular weight of the fusion protein of 243 kDa, which is almost 5 times larger than the calculated molecular weight of 52 kDa. A related fusion protein (LCW0057) contained rPEG36 and had an apparent molecular weight of 55 kDa versus a calculated molecular weight of 32 kDa. Comparison of the LCW0066 and LCW0057 shows a difference in apparent molecular weight of 189 kDa which is caused by the addition of an rPEG chain with a calculated molecular weight of 20 kDa. Thus, one can calculate that the addition of an rPEG tail with a calculated molecular weight of 20 kDa lead to an increase in molecular weight of 189 kDa.

Example 9

Determination of Properties of Accessory-Linked Polypeptides

The fusion protein rPEG-J288-H6-GFP, purified as shown in FIG. 19, containing an N-terminal Flag tag and the accessory sequence rPEG_J288 fused to the N-terminus of green fluorescent protein is incubated in 50% mouse serum at 37°C for 3 days. Samples are withdrawn at various time points and analyzed by SDS PAGE followed by detection using Western analysis. An antibody against the N-terminal flag tag is used for Western detection. FIG. 20 indicates that the accessory protein is stable in serum for at least three days.

Determination of Plasma Half-Life of an Accessory-Linked Polypeptide

The plasma half-life of accessory-linked polypeptides can be measured after i.v. or i.p. injection of the accessory polypeptide into catheterized rats essentially as described by Pepinsky, R. B., et al. (2001) J Pharmacol Exp Ther, 297: 1059-66. Blood samples can be withdrawn at various time points (5 min, 15 min, 30 min, 1 h, 3 h, 5 h, 1d, 2d, 3d) and the plasma concentration of the accessory polypeptide can be measured using ELISA. Pharmacokinetic parameters can be calculated using WinNonlin version 2.0 (Scientific Consulting Inc., Apex, N.C.). To analyze the effect of the rPEG-linked polypeptide one can compare the plasma half-life of a protein containing the rPEG polypeptide with the plasma half-life of the same protein lacking the rPEG polypeptide.
The in vivo half-life or LCW0057 and LCW0066 was studied in rats. Both proteins were injected intravenously into rats. Serum samples were analyzed for the presence of GFP between 5 min and 3 days after injection. For rats injected with LCW0057 no GFP was detectable 24 h after protein injection. This suggests a half-life of the protein of 1-3 h. In contrast, LCW0066 was detectable even 48 h after injection and one rat showed detectable GFP even 3 days after injection. This shows that LCW0066 has a serum half-life in rats of about 10 hours which is much longer than expected for a protein with a calculated molecular weight of 52 kDa.

Solubility testing of accessory-linked polypeptides.

Determination of Serum Concentration of rPEG-GFP Following Subcutaneous Injection of Encapsulated Protein

The serum concentration of rPEG-GFP and GFP can be tested by following a single subcutaneous injection of rPEG-GFP microspheres or GFP microspheres, respectively, in a model laboratory organism. Encapsulated rPEG-GFP or encapsulated GFP is injected into mice, rats, rabbits, or other model organisms (1 mL/kg of body weight) to evaluate in vivo release rates. Serum samples are collected daily for one month. Serum concentrations of rPEG-GFP are measured using the sandwich ELISA assay described above. rPEG-GFP fusion polypeptides are present at a high concentration much longer than GFP due to a slower release from the microspheres and a longer subsequent half-life.

Example 11

Polymer Encapsulated Interferon-Alpha (IFN-Alpha) Linked to an Accessory Polypeptide

This example describes a depot formulation of rPEG-IFN-alpha which can extend the dosing interval of this polypeptide. The rPEG-fused IFN-alpha is constructed essentially as described for the hGHR-pPEG fusion construct in Example 3, except GLP-1 encoding sequences are replaced by IFN-alpha coding sequence. All other methodologies and techniques, including encapsulation methodologies, are essentially as described in Example 10.

This example describes the construction of scFv-rPEG50 fusions. Two scFvs were made, one that binds Her2 and one that binds epidermal growth factor receptor (EGFR). Each scFv was genetically fused to the N-terminus of rPEG50, respectively. The scFv constructs were cloned into an expression vector with T7 promoter and encoding rPEG50-FLAG-tag-hexahistidine (SEQ ID NO: 1), resulting in constructs expressing scFv-rPEG50-FLAG-His6 (SEQ ID NO: 1). The stuffer fragment was removed by restriction digest using Ndel and Bsal endonucleases. The synthetic scFv fragments were amplified by polymerase chain reaction (PCR), which introduced Ndel and Bsal restriction sites that are compatible with the stuffer construct. Restriction digested scFv fragments and stuffer construct were ligated using T4 DNA ligase and electroporated into E. coli BL21 (DE3) Gold. The resulting DNA construct is shown in FIG. 6A, where the light chain (vl) and heavy chain (vH) variable fragments are separated by rPEG50, a 30 amino acid sequence (SHEGSEGEGCGEGSEGPEGSEGEGEGEGSEG)
Example 12

Construction of Non-Repetitive Accessory Polypeptides

This example describes the construction of a library of accessory polypeptide segments from synthetic oligonucleotides. FIG. 78 lists the amino acid sequences of that were encoded by synthetic oligonucleotides. For each amino acid sequence we used two complementary oligonucleotides. The sequences were designed as codon libraries, i.e. multiple different codons were allowed but all sequences encoded just one amino acid sequence. The complementary oligonucleotides were annealed by heating followed by cooling. The oligonucleotides were designed to generate 4 base-pair overlaps during annealing as illustrated in FIG. 79. Two additional annealed oligonucleotides were also added that acted as terminators during the multimerization by ligation reaction. FIG. 79 shows the ligation of annealed oligonucleotides that yielded gene fragments encoding accessory polypeptide segments of varying length. The resulting ligation mixture was separated by electrophoresis as shown in FIG. 79 and the ligation product encoding URP36 was isolated. This ligation product was ligated into an expression vector and the library of URP36 segments was expressed as fusion protein to GFP.

Example 13

Construction of rPEG_Y576

This example describes the construction of a library of URP segments from synthetic oligonucleotides. FIG. 78 lists the amino acid sequences encoded by the synthetic oligonucleotides. For each amino acid sequence we used two complementary oligonucleotides. The sequences were designed as codon libraries, i.e. multiple different codons were allowed but all sequences encoded only one amino acid sequence. The complementary oligonucleotides were annealed by heating followed by cooling. The oligonucleotides were designed to generate 4 base-pair overlaps during annealing as illustrated in FIG. 79. We also added two additional annealed oligonucleotides that acted as terminators during the multimerization by ligation reaction. FIG. 79 illustrates the ligation of annealed oligonucleotides that yielded gene fragments encoding URP segments of varying lengths. The resulting ligation mixture was separated by electrophoresis as shown in FIG. 79 and the ligation product encoding URP36 was isolated. This ligation product was ligated into an expression vector and the library of URP36 segments was expressed as fusion protein to GFP (FIG. 81). Library members with good expression were identified based on their strong fluorescence intensity.
The library members of URP36 were dimerized and the resulting library of URP72 was screened for high level expression. This process of dimerization and screening was repeated one more time to generate URP144. FIG. 80 shows a collection of sequences. The sequences conform to the design of the libraries but most library members differ in their actual sequences. This collection of URP_Y144 was dimerized two more times to generate collections of URP_Y288 and URP_Y576. The amino acid sequence of one isolate of URP_Y576 is shown in FIG. 80. The resulting isolates were evaluated for expression, aggregation, and immunogenicity to identify URP that is most suitable for fusion to a drug protein.

Example 14

Construction of scFv-rPEG50 Fusions

This example describes the construction of scFv-rPEG50 fusions. Two scFv's were made, one that binds Her2 and one that binds epidermal growth factor receptor (EGFR). Each scFv was genetically fused to the N-terminus of rPEG50, respectively. The scFv constructs were cloned into an expression vector with T7 promoter and encoding rPEG50-FLAG-tag-hexahistidine, resulting in constructs expressing scFv-rPEG50-FLAG-His6. The stuffer fragment was removed by restriction digest using Ndel and BsaI endonucleases. The synthetic scFv fragments were amplified by polymerase chain reaction (PCR), which introduced Ndel and BsaI restriction sites that are compatible with the stuffer construct. Restriction digested scFv fragments and stuffer construct were ligated using T4 DNA ligase and electrophoresed into E. coli BL21 (DE3) Gold. The resulting DNA construct is shown in FIG. 64a, where the light chain (VL) and heavy chain (VH) variable fragments are separated by rPEG530, a 30 amino acid sequence (SEQ 1D NO: 1), and/or FLAG tag. The constructs were confirmed by DNA sequencing. The protein sequences for the aHer230-rPEG (M.W. = 80,044 Da) and aEGFR30-rPEG (M.W. = 80,102 Da) constructs are shown in FIGS. 64b and d, respectively.

The anti-Her230-rPEG and aEGFR30-rPEG fusions in E. coli BL21 (DE3) Gold were expressed by inducing with 0.2 mM isopropyl β-D-1 thiogalactopyranoside (IPTG) at 30°C. Cells were harvested by centrifugation and lysed in BugBuster plus Benzonase in phosphate buffered saline. Lysates were clarified by centrifugation and supernatants (soluble fractions) loaded onto 4-12% SDS PAGE gels. The scFv-rPEG fusions are overexpressed and visible in E. coli lysates at approximately 80 kDa (FIG. 64c).

Example 15

Characterization of the scFv-rPEG50 Fusion aHer230-rPEG

Purification

A single-chain fragment variable (scFv) antibody fragment targeting the Her2 receptor and fused to rPEG, to yield aHer230-rPEG, which was expressed and purified from the cytosol of E. coli. The aHer230-rPEG plasmid was transformed into BL21(DE3) Gold and expression of the recombinant antibody fragment was induced with 0.2 mM isopropyl β-D-1 thiogalactopyranoside (IPTG) at 20°C. Cells were harvested by centrifugation and resuspended in 30 mM sodium phosphate, 0.3 M sodium chloride, 10% glycerol, and 20 mM imidazole, pH 7.5. Lysis was accomplished by sonication and the soluble protein was purified by standard chromatographic methods including, immobilized metal affinity chromatography (IMAC), hydrophobic interaction chromatography (HIC), and ion exchange chromatography (IEC).

Binding

To evaluate target (Her2) binding, aHer230-rPEG was expressed in BL21(DE3) Gold as described above. Cells were lysed by resuspension in phosphate buffer saline (PBS) containing BugBuster reagent and 5 U/ml of benzonase (Novagen). The suspension was incubated for 20 minutes at room temperature prior to centrifugation at 10000 rpm for 10 minutes. The soluble fraction was then serially diluted fivefold into PBS containing 1% bovine serum albumin (BSA) and 0.05% Tween-20. Serially diluted aHer230-rPEG was added to the wells of a 96-well plate which had been coated with a Her2-Fc fusion protein (R&D Systems) and blocked with 1% BSA. The binding reaction was incubated at room temperature for 2 hours with gentle agitation. The wells were thoroughly washed with PBS containing 0.05% Tween-20 and the bound aHer230-rPEG was detected with an HRP-conjugated anti-FLAG antibody (Sigma). FIG. 62a shows that aHer230-rPEG binds to Her2-Fc fusion protein and does not non-specifically bind to human IgG. The binding data are presented as a function of the sample dilution. The half maximal binding (EC50) is estimated to be achieved at approximately 10 nM aHer230-rPEG.

This example describes the construction and bacterial expression of a Fab-rPEG fusion protein. The fragment, antigen binding (Fab) of an IgG can be fused to rPEG as a means of improving soluble Fab expression as well as half-life extension. The expression construct was designed as a bicistronic RNA message that is under the control of an inducible arabinose promoter (FIG. 67). The bicistronic message is terminated at a hairpin terminator, such as the T7 terminator sequence. Each cistron or gene has a ribosomal binding site (RBS) to initiate translation and a stop codon (TAA, TAG, or TGA) to stop translation. The light chain (VL/CL) or heavy chain (VH/CH) sequence can be genetically fused to rPEG and followed by an affinity tag such as HA (hemagglutinin), His (hexahistidine) (SEQ 1D NO: 1), and/or FLAG tag. DNA constructs can encode the heavy chain first or light chain first (HL) or light chain first and heavy chain last (LH) as shown in FIG. 67. Protein expression from this type of construct yields two approximately 50 kDa chains that form a full Fab fragment of approximately 100 kDa in size, which includes a total of 50 kDa of rPEG sequence.

SS-Bond Oxidation

The expression of disulfide containing proteins in the cytoplasm of E. coli is often unsuccessful due to the highly reducing nature of the cytoplasm, which inhibits disulfide formation. However, disulfide bonds may form following cell lysis when the proteins are exposed to more oxidizing conditions. As demonstrated above, aHer230-rPEG expressed in E. coli binds to its target, Her2, suggesting that the protein is properly folded. To test whether the two disulfide bonds, one each in the VH and VL domains, of aHer230-rPEG were properly formed in the purified protein, the number of free
sulphydrys in the denatured, purified protein was compared to a fully reduced form of the scFv. Purified aHer230-rPEG was denatured in 6 M urea or in 6 M urea supplemented with 10 mM Tris [2-carboxyethyl] phosphine (TCEP) for 1 hour at room temperature. The samples were then dialyzed on Sephadex G-25 resin to remove the urea and the TCEP. Immediately, Ellman’s reagent (5,5′-dithio-bis[2-nitrobenzoic acid]) was added to a final concentration of 20 mM and the reaction proceeded for 15 minutes. Finally, the absorbance of each solution was measured at 412 nm. Denatured aHer230-rPEG exhibits very little absorbance, which suggests that the purified sample is completely oxidized (FIG. 62c). The denatured and reduced reaction (FIG. 62c) shows the signal expected if all of the cysteines in aHer230-rPEG were in the reduced state. Thus, all of the disulfides within the anti-Her2 scFv were properly formed.

Example 16

Construction of the Diabody aHer203-rPEG

A diabody can be formed by linking the VH and VL domains with a linker less than 10 amino acids. The short linker does not allow scFv formation and as a result the VH and VL domains bind to a complementary, second VH-VL chain, forming a 4-domain, 2 chain 50 kD complex. The diabody was constructed from a single-chain fragment variable (scFv) antibody fragment that binds Her2, which was genetically fused to the N-terminus of rPEG50. Constructs were generated by replacing the Y30 scFv linker sequence from Example 1 with three amino acids (SGE) to allow a diabody format (FIG. 65a). The SGE sequence was introduced by polymerase chain reaction (PCR), also introducing Ndel and Bbsl restriction sites that are compatible with the rPEG linker construct. Diabody-encoding fragments were then cloned as in Example 1. The construct was confirmed by DNA sequencing. The protein sequence for the aHer203-rPEG diabody (M.W.: 156,598 Da as diabody or 78,299 Da monomer sequence, including rPEG) is shown in FIG. 65b. The aHer203-rPEG in BL21 (DE3) Gold was expressed by inducing with 0.2 mM isopropyl β-D-1 thiogalactopyranoside (IPTG) at 20°C. Cells were harvested by centrifugation and lysed in BugBuster/Benzonase in phosphate buffered saline. Lysates were clarified by centrifugation and supernatants (soluble fractions) loaded onto 4-12% SDS PAGE gels. The aHer203-rPEG diabody was detected in E. coli lysates at approximately 90 kDa (FIG. 65c).

Example 17

Characterization of the Diabody-rPEG50 Fusion aHer203-rPEG

Purification

A diabody can be formed by linking the VH and VL domains with a linker comprising fewer than 10 amino acids. The short linker does not allow scFv formation and as a result the VH and VL domains bind to a complementary VH-VL chain. The diabody is a useful format to generate a bivalent, and possibly bispecific, therapeutic lacking effector Fe function.

A diabody that binds to Her2 was designed as described above. To evaluate target (Her2) binding, recombinant aHer203-rPEG diabody was expressed and purified as described for aHer230-rPEG. aHer203-rPEG50 was transformed into BL21(DE3) Gold and expression of the recombinant antibody fragment was induced with 0.2 mM isopropyl β-D-1 thiogalactopyranoside (IPTG) at 20°C. Cells were harvested by centrifugation and resuspended in 30 mM sodium phosphate, 0.3 M sodium chloride, 10% glycerol, an 20 mM imidazole, pH 7.5. Lysis was accomplished by sonication and the soluble protein was purified by standard chromatographic methods including, immobilized metal affinity chromatography (IMAC), hydrophobic interaction chromatography (HIC), and ion exchange chromatography (IEC).

Binding

[0802] Binding of the aHer203-rPEG diabody to its target was performed as described for aHer230-rPEG. Cells were lysed by resuspension in phosphate buffer saline (PBS) containing BugBuster reagent and 5 U/ml of benzonase (Novagen). The suspension was incubated for 20 minutes at room temperature prior to centrifugation at 10000 rpm for 10 minutes. The soluble fraction was then serially diluted fivefold into PBS containing 1% bovine serum albumin (BSA) and 0.05% Tween-20, hereafter referred to as ELISA binding buffer. Serially diluted aHer203-rPEG diabody was added to the wells of a 96-well plate which had been coated with a Her2-Fc fusion protein (R&D Systems) and blocked with 1% BSA. The binding reaction was incubated at room temperature for 2 hours with gentle agitation. The wells were thoroughly washed with PBS containing 0.05% Tween-20 and the bound aHer203-rPEG diabody was detected with an HRP-conjugated anti-FLAG antibody (M2, Sigma). FIG. 65a shows that the aHer203-rPEG diabody binds to the Her2-Fc fusion proteins and does not non-specifically bind to human IgG. The binding data are presented as a function of the sample dilution. The half maximal binding (EC50) is estimated to be achieved at approximately 10 nM aHer203-rPEG diabody. Thus, a functional aHer203 diabody with an rPEG accessory polypeptide can be expressed in the cytosol of E. coli.

SE-HIPLC

[0803] Diabodies have been explored as potential bivalent therapeutics, however, their propensity to reassert into higher order oligomers—trimers, tetramers, etc.—has limited their utility. Reassortment is particularly problematic for manufacturing, because after purifying a monomeric scFv, upon storage in liquid form it will slowly but predictably reassert to yield dimers, and higher multimers. This leads not only to large losses in the amount of protein of the correct format that can finally be obtained, but it also leads to heterogeneity in the product upon storage and heterogeneity in pharmacokinetics and in efficacy. The equilibrium between monomers and multimers of scFv can be affected by the length of the linker between VH and VL domains. In general constructs with linkers of more than 12 to 14 amino acids occur predominantly in monomeric form while scFv with linkers shorter than 12 amino acids occur mostly in multimeric form [Desplanq, D., et al. (1994) Protein Eng. 7: 1027] [Whitlow, M., et al. (1994) Protein Eng. 7: 1017] [Hudson, P. J., et al. (1999) J Immunol Methods, 231: 177]. Increasing the length of the linker between VH and VL to 30 amino acids shifts the equilibrium into the direction of monomers [Desplanq, D., et al. (1994) Protein Eng. 7: 1027]. Linker lengths between 3 and 7 amino acids favor the formation of diabodies [Dolezel, O., et al. (2000) Protein Eng. 13: 565] [Kortt, A. A., et al. (1997) Protein Eng. 10: 423]. Linkers of 5-10 amino acids give rise

The oligomerization state of the αHer203-rPEG diabody by SEC-HPLC has been evaluated and demonstrated that it does not rearrange. FIG. 63b, shows the size-exclusion chromatograms of αHer230-rPEG single chain and the αHer203-rPEG diabody. It demonstrates that the diabody is largely dimeric and, significantly, it contains less than 3% trimer or tetramer forms. The oligomerization state of the αHer203-rPEG diabody has also been monitored during storage at 4°C and reassortment was not observed (FIG. 63c). The rPEG accessory polypeptide helps prevent the reassortment of the diabody, thus enabling the purification and formulation of a homogenous product.

Example 18
Codon Optimization of an Fc Domain for Bacterial Expression

The Human IgG1 constant fragment (Fc) was synthesized and fused to rPEG25-Green Fluorescent Protein (GFP) to yield Fc-rPEG25-GFP, as shown in FIG. 66a. The DNA encoding the Fc sequence was constructed in vitro using E. coli optimized codons. The Fc codon library was assembled using 60-mer oligonucleotides with 20 nucleotide overlap (annealing) regions. Multiple codons were introduced in the non-overlapping regions of the synthetic oligonucleotides. The resulting codon library had a theoretical size of approximately 10.000 such that all nucleotide sequences encode the desired Fc sequence. A total of 18 oligonucleotides were assembled in the presence of dNTPs and DNA polymerase to a final size of 6684 bp. The Fc codon library was amplified by PCR using primers that create Ndel and BbsI compatible ends. The DNA fragment was restriction digested and ligated into an rPEG25-GFP vector at Ndel and Bsal restriction digestion sites. The ligated DNA was transformed into BL21 (DE3) Gold. A total of 1000 clones were isolated, grown in 96-well format, and plated onto plates containing 0.2 mM IPTG to induce expression. Constructs that were well-expressed showed high levels of fluorescence under ultraviolet light. A total of 17 clones were characterized as highly fluorescent. These clones were expressed in 1 ml cultures using 0.2 mM IPTG and lysed by centrifugation.

Example 20
Construction and Bacterial Expression of a Fab-rPEG Fusion Protein

This example describes the construction and bacterial expression of a Fab-rPEG fusion protein. The fragment, antigen binding (Fab) of an IgG1 can be fused to rPEG as a means of improving soluble Fab expression as well as half-life extension. The expression construct was designed a bicistronic RNA message that is under the control of an inducible arabinose promoter (FIG. 67). The bicistronic message is terminated at a hairpin terminator, such as the T7 terminator sequence. Each cistron or gene has a ribosomal binding site (RBS) to initiate translation and a stop codon (TAA, TAG, or TGA) to stop translation. The light chain (vL/eL) or heavy chain (vH/eH) sequence can be genetically fused to rPEG and followed by an affinity tag such as HA (hemagglutinin), H (hexahistidine), and/or FLAG tag. DNA constructs can encode the heavy chain first or light chain last (HL) or light chain first and heavy chain last (LH) as shown in FIG. 67. Protein expression from this type of construct yields two approximately 50 kDa chains that form a full Fab fragment of approximately 100 kDa in size, which includes a total of 50 kDa of rPEG sequence.

Example 19
Expression and Characterization of Fc-rPEG Fusion Proteins

The Fc fragment of IgG1 was fused to rPEG as detailed in Example 5 (and variants are illustrated in FIG. 34), and expressed in the cytoplasm of E. coli. Cells expressing the fusion protein were resuspended in buffer, in this case 20 mM sodium phosphate pH 7.0, and the cells were lysed by sonication. The insoluble material was removed by centrifugation and Fc-rPEG-GFP was purified from the soluble fraction. Intact, folded Fc fragment binds to Protein A and therefore can be conveniently purified by affinity chromatography using immobilized recombinant Protein A. Soluble lysate containing the Fc fusion was applied to a Protein A column (GE Healthcare) and microbial proteins were removed by extensive washing with phosphate buffer. The Fc-rPEG-GFP fusion protein was eluted from the Protein A column using either glycine buffer or sodium citrate buffer pH 3.0. The pH of the elution fractions was immediately adjusted with and equal amount of Tris buffer pH 8.5. The purified protein was analyzed by SDS-PAGE under reducing and oxidizing conditions. A single band of approximately 80 kDa was detected under reducing conditions, while bands at 160 kDa (hinge oxidized) and 80 kDa (hinge reduced) were detected under oxidizing conditions. The addition of either CuSO4, dehydroascorbic acid, or other oxidizing reagents was used to catalyze the complete oxidation of the hinge cysteines.

Example 21
PK Analysis of GFP-rPEG50

The amino acid sequence of GFP-rPEG50 is shown in FIG. 69. The protein was expressed in BL21(DE3) using a T7 promoter similar to example 1. The protein was purified by ion exchange chromatography followed by hydrophobic interaction chromatography. The pharmacokinetics of GFP-rPEG50 was studied in cynomolgous macaques monkeys following s.c. and i.v. injection. Three cynomolgous macaques monkeys were divided into 2 groups, 2 animals dosed i.v. and one dosed s.c. at 0.15 mg/kg with GFP-rPEG50. Serial blood samples were taken from each monkey, the plasma was separated, and the test article plasma concentration was measured by ELISA Assays. The half-life for the iv. dosed animals was 17.4 hours and 13.8 hrs for the s.c. dosed
animals. The bioavailability for the test article was approximately 54.6% as shown in FIG. 70.

Example 22
PK Analysis of Ex4-rPEG50

Ex4-rPEG50 is a fusion protein between exendin-4 and rPEG50. It was produced as a fusion protein with a cellulose binding domain (CBD), which was designed to be removed by cleavage with TEV protease as illustrated in FIG. 71b. The amino acid sequence of the fusion protein is shown in FIG. 71. The expression plasmid and purification protein were similar as in Example 1 with the addition of a step for TEV proteolysis. The cleaved CBD was removed by incubation with beaded cellulose. The pharmacoine of Ex4-rPEG50 was studied in cynomolgous monkeys. Four cynomolgous macaques monkeys were divided into 2 groups, 2 animals per group and dosed s.c. and i.v., at 0.15 mg/kg with Ex4-rPEG50. Serial blood samples were taken from each monkey and the test article plasma concentration was measured by ELISA assay. The half-life was 9.5 hours and 9.1 hours for the s.c. and i.v. dosing, respectively as shown in FIG. 70.

Example 23
PK Analysis of GFP-rPEG50 in Rodents

This example compares the s.c. and i.v. pharmacokinetics of GFP-rPEG25 and GFP-rPEG50. 15 rats were divided into 5 groups, 3 rats per group and dosed both s.c. and i.v. at 1.67 mg/kg with either GFP-rPEG25 and GFP-rPEG50. Y288. GFP-rPEG25 had approximately an 8-t 0.5 when injected s.c versus 11–15 h t 0.5 for GFP-rPEG50. GFP-rPEG25 was approximately 25% s.c bioavailability versus 11% s.c. bioavailability for GFP-rPEG50. In mice, 125I-GFP-rPEG50 was dosed into nude mice. The half-life was 13.4 hours.

Example 24
PK Analysis of Human Growth Hormone Fused to rPEG50

rPEG50 was fused to either the C- or N-terminus of human growth hormone (hGH). Proteins were purified as described in example 8. The pharmacokinetics was studied in cynomolgous monkeys. Two cynomolgous macaques monkeys were divided into 2 groups, 1 animal per group. Each monkey was i.v. dosed at 0.15 mg/kg with the one growth hormone construct, either hGH-rPEG50 or rPEG50-hGH. The two growth hormone constructs had half-life of 7 and 10.5 hrs, respectively.

Example 25
Mouse Immunogenicity and Toxicology Study of Ex4-rPEG50

This example describes the immunogenicity and potential toxicity associated with ten s.c. 50 µg doses of Ex4-rPEG50 (1/week) into a mouse. 20 mice (Swiss Webster) total, each 30-40 g with 10 mice/group, 5 males and 5 females/group, using 2 groups dosed weekly with either Ex4-rPEG50 or ELSPAR that served as control as illustrated in FIG. 72a. Before each dose a blood sample was taken and the IgG was measured by ELISA assay as shown in FIGS. 72b and 72c. ELSPAR resulted in a significant immune response that increased over time. In contrast Ex4-rPEG50 gave a very weak response that showed a maximum after 6 antigen injections and decreased in the sample obtained after 10 antigen injections. All mice gained weight during the study and showed no behavioral signs of toxicity and necropsy revealed no unusual finding with regard to organ morphology. After completion of the in life portion blood samples, blood smears, and plasma and tissue samples were shipped to RADIL (Columbia, Mo.) for toxicology analysis. Histology analysis showed that no distinct cytoplasmic vacuolation was present in the distal or proximal tubules, which is a major concern for chemical conjugates with PEG. Evaluation of liver histology showed mild inflammation in all four analyzed samples. This is a common finding in the livers of apparently healthy animals. Analysis of the spleen showed that all four mice have moderate to marked megakaryocytosis and moderate hematopoiesis. Clinical chemistry revealed ALT and ALP levels that were moderately high for one of the animals indicating hepatocellular damage/necrosis. It is not severe or chronic based on the observation. Hematology revealed that all four mice had at least one slightly elevated blood cell count, hemoglobin, hematocrit percentage or blood total protein concentration. Overall, multiple injections of rPEG fusion protein resulted in very minor immunogenicity and toxicity.

Example 26
Size Exclusion Chromatography of GFP-rPEG Fusion Proteins

GFP fused to rPEG Y25 and rPEG Y50 was expressed as discussed in Example 8. The proteins were analyzed by analytical SEC using a TSK G4000 SWXL (Tosoh, Grove City, Ohio) as shown in FIG. 73. The column was calibrated using a commercial standard of globular proteins and molecular weights of the controls are shown in FIG. 73. GFP-rPEG25 eluted at an apparent molecular weight of 500 kDa whereas GFP-rPEG50 eluted at an apparent molecular weight of 1500 kDa.

Example 27
Formulation and In vivo Administration of GFP-rPEGY Fusion Proteins

A solution of GFP-rPEG at 10 mg/mL in PBS is mixed with an equal volume of 5 mg/mL Chitosan in PBS and incubated at room temperature for 30 minutes. Precipitate is collected by centrifugation at 5,000xg for 10 minutes, and washed quickly one time with 0.1 volume sterile PBS. The precipitate is then lyophilized to remove excess fluid and ground to a fine powder. 15 mg of powder is then resuspended in 1 mL sterile PBS and homogenized by pipetting up and down. The homogenate is stored rotating at 37° C. for 2 weeks, with 10 mL samples removed at regular intervals. Samples are prepared immediately by centrifugation to remove insoluble material, and resolubilized protein is quantitated in the supernatant by GFP fluorescence, optical density, and rPEG ELISA. Supernatant concentration is plotted as a function of time and fit to a single exponential process to determine the resolubilization rate. To determine in vivo release rates, Sprague-Dawley rats are injected subcutaneously with a freshly prepared suspension of 20 mg powder in 1 mL PBS at a dosage of 1 mL/kg (5 mg/kg effective dose). Intravenous and subcutaneous injections of uncomplexed
GFP-pPEGY are injected at 5 mg/kg into independent cohorts of animals in parallel. Blood samples are taken at regular intervals, and serum concentration of protein is determined by GFP and pPEGY ELISAs. Pharmacokinetic parameters including clearance rate, $C_{\text{max}}$, $C_{\text{ss}}$, $V_p$, AUC and serum half-life are determined by standard methods (ie WinNonLin in analysis). Bioavailability and effective dose for subcutaneous and depot formulations are determined by comparison to intravenous dosing.

[0815] Thus, while preferred embodiments of the present invention have been shown and described herein, it will be obvious to those skilled in the art that such embodiments are provided by way of example only. Numerous variations, changes, and substitutions will now occur to those skilled in the art without departing from the invention. It should be understood that various alternatives to the embodiments of the invention described herein may be employed in practicing the invention. It is intended that the following claims define the scope of the invention and that methods and structures within the scope of these claims and their equivalents be covered thereby.

---

**SEQUENCE LISTING**

<160> NUMBER OF SEQ ID NOs: 524

<210> SEQ ID NO 1
<211> LENGTH: 6
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic 6x His tag

<400> SEQUENCE: 1
His His His His His His
1  5

<210> SEQ ID NO 2
<211> LENGTH: 60
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide
<223> OTHER INFORMATION: This sequence may encompass 2 to 10 "Gly Gly Gly Gly Gly Gly序列" repeating units

<400> SEQUENCE: 2
1  5  10  15
20  25  30
35  40  45
Gly Gly Gly Ser Gly Gly Gly Gly Gly Ser Gly
50  55  60

<210> SEQ ID NO 3
<211> LENGTH: 90
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide
<223> OTHER INFORMATION: This sequence may encompass 1 to 10 "Gly Gly Gly Gly Gly Gly Gly Gly Ser" repeating units

<400> SEQUENCE: 3
1  5  10  15
Glu Gly Gly Glu Gly Gly Gly Gly Glu Gly Gly Glu Gly
Gly Gly Glu Gly Gly Gly Gly Glu Gly Gly Glu Gly
Gly Gly Glu Gly Gly Gly Gly Glu Gly Gly Glu Gly
Glu Gly Glu Gly Gly Glu Gly Gly Glu Gly Gly Glu Gly
Glu Gly Gly Gly Glu Gly Gly Glu Gly Gly Glu Gly
Glu Gly Gly Glu Gly Gly Glu Gly Gly Glu Gly
Glu Gly Gly Glu Gly Gly Glu Gly Gly Glu Gly
Glu Gly Glu Gly Gly Glu Gly Gly Glu Gly

<210> SEQ ID NO 4
<211> LENGTH: 50
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide
<220> FEATURE:
<223> OTHER INFORMATION: This sequence may encompass 10, 15, 20 or 50 'Ser' repeating residues

<400> SEQUENCE: 4
Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser 1 5 10 15
Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser 20 25 30
Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser 35 40 45
Ser Ser
50

<210> SEQ ID NO 5
<211> LENGTH: 18
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 5
Gly Ser

<210> SEQ ID NO 6
<211> LENGTH: 10
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 6
Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly 1 5 10

<210> SEQ ID NO 7
<211> LENGTH: 10
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 7

Ser Ser Ser Ser Ser Ser Ser Ser
1 5 10

<210> SEQ ID NO 9
<211> LENGTH: 7
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (7)...
<223> OTHER INFORMATION: May or may not be present

<400> SEQUENCE: 9

Ser Lys Val Ile Leu Phe Glu
1 5

<210> SEQ ID NO 9
<211> LENGTH: 8
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 9

Arg Ala Arg Ala Asp Ala Asp Ala
1 5

<210> SEQ ID NO 10
<211> LENGTH: 224
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 10

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

Gly Gly Pro Ser Val Phe Leu Phe Pro Pro Lys Pro Asp Thr Leu
20 25

Met Ile Ser Arg Thr Pro Glu Val Thr Cys Val Val Val Asp Val Ser
30 35 40

His Glu Asp Pro Glu Val Lys Phe Asn Trp Tyr Val Asp Gly Val Glu
45 50 55 60

Val His Asn Ala Lys Thr Lys Pro Arg Glu Glu Gln Tyr Asn Ser Thr
65 70 75 80

Tyr Arg Val Val Ser Val Leu Thr Val Leu His Gln Asp Trp Leu Asn
85 90 95

Gly Lys Glu Tyr Lys Cys Lys Val Ser Asn Lys Ala Leu Pro Ala Pro
100 105 110

Ile Glu Lys Thr Ile Ser Lys Ala Lys Gly Gln Pro Arg Glu Pro Gln
115 120 125

Val Tyr Thr Leu Pro Pro Ser Arg Asp Glu Leu Thr Lys Asn Glu Val
130 135 140

Ser Leu Thr Cys Leu Val Lys Gly Phe Tyr Pro Ser Asp Ile Ala Val
Glu Trp Glu Ser Aaa Gly Gln Pro Glu Aam Aam Tyr Lys Thr Ile Pro
165 170 175
Pro Val Leu Asp Ser Asp Gly Ser Phe Phe Leu Tyr Ser Lys Leu Thr
180 185 190
Val Asp Lys Ser Arg Trp Gln Gln Gly Aam Val Phe Ser Cys Ser Val
195 200 205
Met His Glu Ala Leu His Aam His Tyr Thr Gin Lys Ser Leu Ser Leu
210 215 220

<210> SEQ ID NO 11
<211> LENGTH: 9
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 11
Gly Glu Gly Ser Gly Gly Gly Ser Glu
1 5

<210> SEQ ID NO 12
<211> LENGTH: 9
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 12
Gly Glu Gly Gly Ser Glu Gly Ser Glu
1 5

<210> SEQ ID NO 13
<211> LENGTH: 9
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 13
Gly Glu Gly Ser Glu Gly Ser Gly Glu
1 5

<210> SEQ ID NO 14
<211> LENGTH: 9
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 14
Gly Glu Gly Ser Glu Gly Gly Ser Glu
1 5

<210> SEQ ID NO 15
<211> LENGTH: 9
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
-continued

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 15
Gly Glu Gly Ser Gly Glu Gly Gly Glu
1  5

<210> SEQ ID NO 16
<211> LENGTH: 9
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 16
Gly Glu Gly Gly Ser Gly Gly Glu
1  5

<210> SEQ ID NO 17
<211> LENGTH: 9
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 17
Gly Glu Gly Gly Glu Gly Gly Ser Glu
1  5

<210> SEQ ID NO 18
<211> LENGTH: 9
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 18
Gly Glu Gly Gly Glu Gly Ser Gly Glu
1  5

<210> SEQ ID NO 19
<211> LENGTH: 9
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 19
Gly Glu Gly Gly Glu Gly Ser Glu
1  5

<210> SEQ ID NO 20
<211> LENGTH: 9
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 20
Gly Glu Gly Ser Gly Gly Gly Gly Glu
<213> ORGANISM: Artificial Sequence
<220>FEATURE:
<223>OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide
<220>FEATURE:
<221>NAME/KEY: MOD_RES
<222>LOCATION: (2)...(2)
<223>OTHER INFORMATION: Ser or Glu

<400>SEQUENCE: 24
Gly Xaa Glu Gly Gly Ser Glu Gly Gly Ser Gly Glu
1 5 10

<210>SEQ ID NO: 25
<211>LENGTH: 12
<212>TYPES: PRT
<213> ORGANISM: Artificial Sequence
<220>FEATURE:
<223>OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide
<220>FEATURE:
<221>NAME/KEY: MOD_RES
<222>LOCATION: (6)...(6)
<223>OTHER INFORMATION: Ser or Glu
<220>FEATURE:
<221>NAME/KEY: MOD_RES
<222>LOCATION: (9)...(9)
<223>OTHER INFORMATION: Ser or Glu
<220>FEATURE:
<221>NAME/KEY: MOD_RES
<222>LOCATION: (11)...(11)
<223>OTHER INFORMATION: Ser or Glu

<400>SEQUENCE: 25
Gly Ser Gly Glu Gly Xaa Glu Gly Xaa Gly Xaa Glu
1 5 10

<210>SEQ ID NO: 26
<211>LENGTH: 12
<212>TYPES: PRT
<213> ORGANISM: Artificial Sequence
<220>FEATURE:
<223>OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide
<220>FEATURE:
<221>NAME/KEY: MOD_RES
<222>LOCATION: (6)...(6)
<223>OTHER INFORMATION: Ser or Glu
<220>FEATURE:
<221>NAME/KEY: MOD_RES
<222>LOCATION: (9)...(9)
<223>OTHER INFORMATION: Ser or Glu
<220>FEATURE:
<221>NAME/KEY: MOD_RES
<222>LOCATION: (11)...(11)
<223>OTHER INFORMATION: Ser or Glu

<400>SEQUENCE: 26
Gly Gly Ser Glu Gly Xaa Glu Gly Xaa Gly Xaa Glu
1 5 10

<210>SEQ ID NO: 27
<211>LENGTH: 12
<212>TYPES: PRT
<213> ORGANISM: Artificial Sequence
<220>FEATURE:
<223>OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide
<220>FEATURE:
<221>NAME/KEY: MOD_RES
<222> LOCATION: (6) (6)
<223> OTHER INFORMATION: Ser or Glu

<400> SEQUENCE: 27

Gly Ser Gly Glu Gly Xaa Glu Gly Gly Ser Gly Glu
1  5  10

<210> SEQ ID NO 28
<211> LENGTH: 12
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (6) (6)
<223> OTHER INFORMATION: Ser or Glu

<400> SEQUENCE: 28

Gly Ser Glu Gly Xaa Glu Gly Gly Ser Gly Glu
1  5  10

<210> SEQ ID NO 29
<211> LENGTH: 20
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (2) (2)
<223> OTHER INFORMATION: Asp, Glu, Thr or Pro
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (4) (4)
<223> OTHER INFORMATION: Asp, Glu, Thr or Pro
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (6) (6)
<223> OTHER INFORMATION: Asp, Glu, Thr or Pro
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (8) (8)
<223> OTHER INFORMATION: Asp, Glu, Thr or Pro
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (10) (10)
<223> OTHER INFORMATION: Asp, Glu, Thr or Pro
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (12) (12)
<223> OTHER INFORMATION: Asp, Glu, Thr or Pro
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (14) (14)
<223> OTHER INFORMATION: Asp, Glu, Thr or Pro
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (16) (16)
<223> OTHER INFORMATION: Asp, Glu, Thr or Pro
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (18) (18)
<223> OTHER INFORMATION: Asp, Glu, Thr or Pro
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (20) (20)
<223> OTHER INFORMATION: Asp, Glu, Thr or Pro

<400> SEQUENCE: 29
Gly Xaa Gly Xaa Gly Xaa Gly Xaa Gly Xaa Gly Xaa Gly Xaa Gly Xaa
1      5      10     15

Gly Xaa Gly Xaa
20

<210> SEQ ID NO 30
<211> LENGTH: 20
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
  <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide
  <220> FEATURE:
  <221> NAME/KEY: MOD_RES
  <222> LOCATION: (2)...(2)
  <223> OTHER INFORMATION: Asp, Glu, Thr or Pro
  <220> FEATURE:
  <221> NAME/KEY: MOD_RES
  <222> LOCATION: (4)...(4)
  <223> OTHER INFORMATION: Asp, Glu, Thr or Pro
  <220> FEATURE:
  <221> NAME/KEY: MOD_RES
  <222> LOCATION: (6)...(6)
  <223> OTHER INFORMATION: Asp, Glu, Thr or Pro
  <220> FEATURE:
  <221> NAME/KEY: MOD_RES
  <222> LOCATION: (8)...(8)
  <223> OTHER INFORMATION: Asp, Glu, Thr or Pro
  <220> FEATURE:
  <221> NAME/KEY: MOD_RES
  <222> LOCATION: (10)...(10)
  <223> OTHER INFORMATION: Asp, Glu, Thr or Pro
  <220> FEATURE:
  <221> NAME/KEY: MOD_RES
  <222> LOCATION: (12)...(12)
  <223> OTHER INFORMATION: Asp, Glu, Thr or Pro
  <220> FEATURE:
  <221> NAME/KEY: MOD_RES
  <222> LOCATION: (14)...(14)
  <223> OTHER INFORMATION: Asp, Glu, Thr or Pro
  <220> FEATURE:
  <221> NAME/KEY: MOD_RES
  <222> LOCATION: (16)...(16)
  <223> OTHER INFORMATION: Asp, Glu, Thr or Pro
  <220> FEATURE:
  <221> NAME/KEY: MOD_RES
  <222> LOCATION: (18)...(18)
  <223> OTHER INFORMATION: Asp, Glu, Thr or Pro
  <220> FEATURE:
  <221> NAME/KEY: MOD_RES
  <222> LOCATION: (20)...(20)
  <223> OTHER INFORMATION: Asp, Glu, Thr or Pro

<400> SEQUENCE: 30
Ser Xaa Ser Xaa Ser Xaa Ser Xaa Ser Xaa Ser Xaa Ser Xaa Ser Xaa
1      5      10     15
Ser Xaa Ser Xaa
20

<210> SEQ ID NO 31
<211> LENGTH: 21
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
  <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide
  <220> FEATURE:
  <221> NAME/KEY: MOD_RES
  <222> LOCATION: (3)...(3)
  <223> OTHER INFORMATION: Asp, Glu, Thr or Pro
  <220> FEATURE:
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (6)...(6)
<223> OTHER INFORMATION: Asp, Glu, Thr or Pro
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (9)...(9)
<223> OTHER INFORMATION: Asp, Glu, Thr or Pro
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (12)...(12)
<223> OTHER INFORMATION: Asp, Glu, Thr or Pro
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (15)...(15)
<223> OTHER INFORMATION: Asp, Glu, Thr or Pro
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (18)...(18)
<223> OTHER INFORMATION: Asp, Glu, Thr or Pro
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (21)...(21)
<223> OTHER INFORMATION: Asp, Glu, Thr or Pro

<400> SEQUENCE: 31

1   5   10   15

Gly Xaa Gly Gly Xaa
20

<210> SEQ ID NO 32
<211> LENGTH: 21
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (3)...(3)
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (6)...(6)
<223> OTHER INFORMATION: Asp, Glu, Thr or Pro
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (9)...(9)
<223> OTHER INFORMATION: Asp, Glu, Thr or Pro
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (12)...(12)
<223> OTHER INFORMATION: Asp, Glu, Thr or Pro
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (15)...(15)
<223> OTHER INFORMATION: Asp, Glu, Thr or Pro
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (18)...(18)
<223> OTHER INFORMATION: Asp, Glu, Thr or Pro
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (21)...(21)
<223> OTHER INFORMATION: Asp, Glu, Thr or Pro

<400> SEQUENCE: 32

Ser Ser Xaa Ser Ser Xaa Ser Ser Xaa Ser Ser Xaa Ser Ser Ser Xaa Ser
1   5   10   15

Ser Xaa Ser Ser Xaa
<210> SEQ ID NO 33
<211> LENGTH: 20
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (4) (4)
<223> OTHER INFORMATION: Asp, Glu, Thr or Pro
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (8) (8)
<223> OTHER INFORMATION: Asp, Glu, Thr or Pro
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (12) (12)
<223> OTHER INFORMATION: Asp, Glu, Thr or Pro
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (16) (16)
<223> OTHER INFORMATION: Asp, Glu, Thr or Pro
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (20) (20)
<223> OTHER INFORMATION: Asp, Glu, Thr or Pro
<400> SEQUENCE: 33
Gly Gly Gya Gly Gya Gly Gya Gly Gya Gly Gya Gly Gya Gly Gya 1 5 10 15
Gly Gly Gya
20
<210> SEQ ID NO 34
<211> LENGTH: 20
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (4) (4)
<223> OTHER INFORMATION: Asp, Glu, Thr or Pro
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (8) (8)
<223> OTHER INFORMATION: Asp, Glu, Thr or Pro
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (12) (12)
<223> OTHER INFORMATION: Asp, Glu, Thr or Pro
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (16) (16)
<223> OTHER INFORMATION: Asp, Glu, Thr or Pro
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (20) (20)
<223> OTHER INFORMATION: Asp, Glu, Thr or Pro
<400> SEQUENCE: 34
Ser Ser Ser Gya Ser Ser Ser Gya Ser Ser Ser Gya Ser Ser Ser Gya 1 5 10 15
Ser Ser Ser Gya
20
1  5  10  15
Gly Gly Gly Xaa
20

Ser Ser Ser Ser Xaa Ser Ser Ser Ser Xaa Ser Ser Ser Ser Xaa Ser
1  5  10  15
Ser Ser Ser Xaa
20
LOCATION: (1) (20)

OTHER INFORMATION: This region may encompass 1 to 20
"Gly" repeating residues

FEATURE:
NAME/KEY: MOD_RES
LOCATION: (21) (21)
OTHER INFORMATION: Asp, Glu, Thr or Pro

FEATURE:
NAME/KEY: MOD_RES
LOCATION: (22) (41)
OTHER INFORMATION: This region may encompass 1 to 20
"Gly" repeating residues

FEATURE:
NAME/KEY: MOD_RES
LOCATION: (42) (42)
OTHER INFORMATION: Asp, Glu, Thr or Pro

FEATURE:
NAME/KEY: MOD_RES
LOCATION: (43) (62)
OTHER INFORMATION: This region may encompass 1 to 20
"Gly" repeating residues

FEATURE:
NAME/KEY: MOD_RES
LOCATION: (44) (63)
OTHER INFORMATION: Asp, Glu, Thr or Pro

FEATURE:
NAME/KEY: MOD_RES
LOCATION: (45) (83)
OTHER INFORMATION: This region may encompass 1 to 20
"Gly" repeating residues

FEATURE:
NAME/KEY: MOD_RES
LOCATION: (84) (84)
OTHER INFORMATION: Asp, Glu, Thr or Pro

FEATURE:
NAME/KEY: MOD_RES
LOCATION: (85) (104)
OTHER INFORMATION: This region may encompass 1 to 20
"Gly" repeating residues

FEATURE:
NAME/KEY: MOD_RES
LOCATION: (105) (105)
OTHER INFORMATION: Asp, Glu, Thr or Pro

FEATURE:
NAME/KEY: MOD_RES
LOCATION: (106) (125)
OTHER INFORMATION: This region may encompass 1 to 20
"Gly" repeating residues

FEATURE:
NAME/KEY: MOD_RES
LOCATION: (107) (146)
OTHER INFORMATION: This region may encompass 1 to 20
"Gly" repeating residues

FEATURE:
NAME/KEY: MOD_RES
LOCATION: (147) (147)
OTHER INFORMATION: Asp, Glu, Thr or Pro

FEATURE:
NAME/KEY: MOD_RES
LOCATION: (148) (167)
OTHER INFORMATION: This region may encompass 1 to 20
"Gly" repeating residues

FEATURE:
NAME/KEY: MOD_RES
LOCATION: (168) (168)
OTHER INFORMATION: Asp, Glu, Thr or Pro

FEATURE:
NAME/KEY: MOD_RES
LOCATION: (169) (188)
OTHER INFORMATION: This region may encompass 1 to 20
"Gly" repeating residues

FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (189) .. (189)
<223> OTHER INFORMATION: Asp, Glu, Thr or Pro

<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (190) .. (209)
<223> OTHER INFORMATION: This region may encompass 1 to 20 "Gly" repeating residues

<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (210) .. (210)
<223> OTHER INFORMATION: Asp, Glu, Thr or Pro

<400> SEQUENCE: 37

1      5      10     15
20     25     30
35     40     45
50     55     60
65     70     75     80
85     90     95
100    105    110
115    120    125
130    135    140
145    150    155    160
165    170    175
180    185    190
195    200    205

Gly Xaa
210

<210> SEQ ID NO: 38
<211> LENGTH: 210
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence

<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide

<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (22) .. (21)
<223> OTHER INFORMATION: Asp, Glu, Thr or Pro

<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (22) .. (41)
<223> OTHER INFORMATION: This region may encompass 1 to 20 "Ser" repeating residues

<220> FEATURE:
<221> NAME/KEY: MOD_RES
OTHER INFORMATION: Asp, Glu, Thr or Pro

LOCATION: (42)...(42)

FEATURE:
NAME/KEY: MOD_RES
LOCATION: (43)...(43)
OTHER INFORMATION: This region may encompass 1 to 20 "Ser" repeating residues

LOCATION: (63)...(63)

FEATURE:
NAME/KEY: MOD_RES
LOCATION: (64)...(64)
OTHER INFORMATION: This region may encompass 1 to 20 "Ser" repeating residues

LOCATION: (83)...(83)

FEATURE:
NAME/KEY: MOD_RES
LOCATION: (84)...(84)
OTHER INFORMATION: This region may encompass 1 to 20 "Ser" repeating residues

LOCATION: (104)...(104)

FEATURE:
NAME/KEY: MOD_RES
LOCATION: (105)...(105)
OTHER INFORMATION: This region may encompass 1 to 20 "Ser" repeating residues

LOCATION: (125)...(125)

FEATURE:
NAME/KEY: MOD_RES
LOCATION: (126)...(126)
OTHER INFORMATION: This region may encompass 1 to 20 "Ser" repeating residues

LOCATION: (146)...(146)

FEATURE:
NAME/KEY: MOD_RES
LOCATION: (147)...(147)
OTHER INFORMATION: This region may encompass 1 to 20 "Ser" repeating residues

LOCATION: (167)...(167)

FEATURE:
NAME/KEY: MOD_RES
LOCATION: (168)...(168)
OTHER INFORMATION: This region may encompass 1 to 20 "Ser" repeating residues

LOCATION: (187)...(187)

FEATURE:
NAME/KEY: MOD_RES
LOCATION: (188)...(188)
OTHER INFORMATION: This region may encompass 1 to 20 "Ser" repeating residues

LOCATION: (197)...(197)

FEATURE:
NAME/KEY: MOD_RES
LOCATION: (198)...(198)
OTHER INFORMATION: This region may encompass 1 to 20 "Ser" repeating residues

LOCATION: (209)...(209)

FEATURE:
NAME/KEY: MOD_RES
LOCATION: (209)...(209)
OTHER INFORMATION: This region may encompass 1 to 20 "Ser" repeating residues

LOCATION: (229)...(229)

FEATURE:
NAME/KEY: MOD_RES
LOCATION: (230)...(230)
OTHER INFORMATION: Asp, Glu, Thr or Pro

SEQUENCE: 38
-continued

Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser  
1     5     10     15
Ser Ser Ser Ser Xaa Ser Ser Ser Ser Ser Ser Ser  
20    25     30
Ser Ser Ser Ser Ser Ser Ser Ser Xaa Ser Ser Ser Ser  
35    40     45
Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Xaa Ser  
50    55     60
Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser  
65    70     75     80
Ser Ser Xaa Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser  
95    90     95
Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser  
100   105    110
Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Xaa Ser  
115   120    125
Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser  
130   135    140
Ser Ser Xaa Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser  
145   150    155    160
Ser Ser Ser Ser Ser Ser Ser Ser Xaa Ser Ser Ser Ser Ser  
165   170    175
Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Xaa Ser  
180   185    190
Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser  
195   200    205
Ser Xaa  
210

<210> SEQ ID NO 39
<211> LENGTH: 288
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:  
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide

<400> SEQUENCE: 39
Ser Ser Gly Ser Ser Glu Ser Ser Gly Ser Ser Glu Ser Ser Gly Ser  
1     5     10     15
Ser Glu Ser Ser Gly Ser Ser Glu Ser Ser Gly Ser Ser Glu Ser Ser  
20    25     30
Gly Ser Ser Glu Ser Ser Gly Ser Ser Gly Ser Ser Glu Ser Ser  
35    40     45
Ser Ser Gly Ser Ser Glu Ser Ser Gly Ser Ser Glu Ser Ser Gly Ser  
50    55     60
Ser Glu Ser Ser Gly Ser Ser Glu Ser Ser Gly Ser Ser Glu Ser Ser  
65    70     75     80
Gly Ser Ser Glu Ser Ser Gly Ser Ser Gly Ser Ser Glu Ser Ser  
85    90     95
Ser Ser Gly Ser Ser Glu Ser Ser Gly Ser Ser Glu Ser Ser Gly Ser  
100   105    110
Ser Glu Ser Ser Gly Ser Ser Glu Ser Ser Gly Ser Ser Glu Ser Ser  
125   130    135
Ser Xaa  
210
Gly Ser Ser Glu Ser Ser Gly Ser Ser Glu Ser Ser Gly Ser Ser Glu
115     120     125                      130     135     140
Ser Ser Gly Ser Ser Glu Ser Ser Gly Ser Ser Glu Ser Ser Gly
145     150     155                      160
Ser Glu Ser Ser Gly Ser Ser Glu Ser Ser Gly Ser Ser Glu Ser
165     170     175                      180     185     190
Gly Ser Ser Glu Ser Ser Gly Ser Ser Glu Ser Ser Gly Ser Ser Glu
195     200     205                      210     215     220
Ser Ser Gly Ser Ser Glu Ser Ser Gly Ser Ser Glu Ser Ser Gly
225     230     235                      240
Gly Ser Ser Glu Ser Ser Gly Ser Ser Glu Ser Ser Gly Ser Ser Glu
245     250     255                      260     265     270
Gly Ser Ser Glu Ser Ser Gly Ser Ser Glu Ser Ser Gly Ser Ser Glu
275     280     285

<210> SEQ ID NO: 40
<211> LENGTH: 288
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide

<400> SEQUENCE: 40
 Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu Ser
  1     5     10                     15
 Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu Ser
  20    25    30                      35    40    45
 Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu Ser
  50    55    60                      65    70    75
 Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu Ser
  85    90    95                      100   105   110
 Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu Ser
 115   120   125                     130   135   140
 Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu Ser
 145   150   155                     160   165   170
 Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu Ser
 175   180   185                     190   195   200
 Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu Ser
 205   210   215                     220   225   230
 Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu Ser
 235   240   245                     250   255   260
 Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu Ser
 265   270   275                     280   285   288

<224> COMPLEMENTS: Black; BACKCORD: Black; TORQUE: Black
| Ser | Ser | Ser | Glu | Ser | Ser | Glu | Ser | Ser | Ser | Glu | Ser | Ser | Ser | Glu | Ser | Ser | Ser | Glu | Ser | Ser | Ser | Glu |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 180 |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| 185 |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| 190 |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| Ser | Ser | Glu | Ser | Ser | Ser | Glu | Ser | Ser | Ser | Glu | Ser | Ser | Ser | Glu | Ser | Ser | Ser | Ser | Glu | Ser | Ser | Ser | Glu |
| 195 |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| 200 |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| 205 |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| Ser | Ser | Glu | Ser | Ser | Ser | Glu | Ser | Ser | Ser | Ser | Glu | Ser | Ser | Ser | Glu | Ser | Ser | Ser | Ser | Glu | Ser | Ser | Ser | Glu |
| 210 |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| 215 |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| 220 |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| Ser | Ser | Glu | Ser | Ser | Glu | Ser | Ser | Ser | Ser | Ser | Glu | Ser | Ser | Ser | Ser | Glu | Ser | Ser | Ser | Ser | Glu | Ser | Ser | Ser | Glu |
| 225 |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| 230 |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| 235 |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| 240 |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| Ser | Ser | Glu | Ser | Ser | Ser | Ser | Glu | Ser | Ser | Ser | Ser | Glu | Ser | Ser | Ser | Ser | Glu | Ser | Ser | Ser | Ser | Glu | Ser | Ser | Ser | Glu |
| 245 |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| 250 |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| 255 |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| Ser | Ser | Glu | Ser | Ser | Ser | Ser | Ser | Glu | Ser | Ser | Ser | Ser | Glu | Ser | Ser | Ser | Ser | Glu | Ser | Ser | Ser | Ser | Glu | Ser | Ser | Ser | Glu |
| 260 |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| 265 |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| 270 |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| Ser | Ser | Glu | Ser | Ser | Glu | Ser | Ser | Glu | Ser | Ser | Ser | Ser | Glu | Ser | Ser | Ser | Ser | Glu | Ser | Ser | Ser | Ser | Glu | Ser | Ser | Ser | Glu |
| 275 |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| 280 |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| 285 |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |

<210> SEQ ID NO 41
<211> LENGTH: 324
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide

<400> SEQUENCE: 41

1   5   10  15
  20  25  30
  35  40  45
  50  55  60
  65  70  75  80
  85  90  95
 100 105 110
 115 120 125
 130 135 140
 145 150 155 160
 165 170 175
 180 185 190
 195 200 205
 210 215 220
 225 230 235 240
245 250 255
260 265 270
275 280 285
290 295 300
305 310 315 320

Glu Gly Gly Glu

<210> SEQ ID NO 42
<211> LENGTH: 5
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 42
Gly Gly Gly Ser Glu
1 5

<210> SEQ ID NO 43
<211> LENGTH: 4
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 43
Gly Gly Ser Glu
1

<210> SEQ ID NO 44
<211> LENGTH: 5
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 44
Glu Glu Glu Glu Glu
1 5

<210> SEQ ID NO 45
<211> LENGTH: 5
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 45
Gly Gly Gly Gly Gly
1 5

<210> SEQ ID NO 46
Ser Ser Ser Ser Ser
1  5

Ala Ala Ala Ala Ala
1  5

Ser Glu Ser Ser Glu Ser Ser Glu
1  5

Ser Ser Glu Ser Ser Ser Glu Ser Ser Glu
1  5  10

Ser Ser Glu Ser Ser Ser Ser Glu Ser Ser Glu
1  5  10  15
Ser Ser Ser Ser Glu Ser Ser Ser Ser Ser Ser Glu
1  5  10

<210> SEQ ID NO 52
<211> LENGTH: 4
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 52
Ser Ser Ser Glu
1

<210> SEQ ID NO 53
<211> LENGTH: 5
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 53
Ser Ser Ser Ser Glu
1  5

<210> SEQ ID NO 54
<211> LENGTH: 6
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 54
Ser Ser Ser Ser Ser Glu
1  5

<210> SEQ ID NO 55
<211> LENGTH: 7
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 55
Ser Ser Ser Ser Ser Ser Glu
1  5

<210> SEQ ID NO 56
<211> LENGTH: 21
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 56
Gly Glu Gly Glu Ser Glu Gly Glu Gly Glu Gly Glu Gly Glu Ser Glu Gly Glu
1  5  10  15
Gly Glu Ser Gly Glu
20

<210> SEQ ID NO 57
<211> LENGTH: 40
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide

<400> SEQUENCE: 57
1  5  10   15
20  25   30
Gly Glu Glu Glu Glu Glu Glu Glu Glu Glu Glu
35  40

<210> SEQ ID NO 58
<211> LENGTH: 5
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 58
Gly Gly Gly Glu Glu
1  5

<210> SEQ ID NO 59
<211> LENGTH: 6
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 59
Gly Gly Glu Gly Gly Ser
1  5

<210> SEQ ID NO 60
<211> LENGTH: 6
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 60
Glu Gly Gly Ser Gly Gly
1  5

<210> SEQ ID NO 61
<211> LENGTH: 6
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 61
**-continued**

Gly Glu Gly Gly Ser Gly
1  5

```
<210> SEQ ID NO 62
<211> LENGTH: 6
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<222> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 62
Gly Gly Ser Gly Gly Glu
1  5
```

Gly Glu Gly Gly Ser Gly
1  5

```
<210> SEQ ID NO 63
<211> LENGTH: 6
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<222> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 63
Ser Gly Glu Gly Glu Gly
1  5
```

Gly Ser Gly Glu Gly Glu
1  5

```
<210> SEQ ID NO 64
<211> LENGTH: 6
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<222> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 64
Gly Ser Gly Glu Gly Glu
1  5
```

Gly Ser Gly Glu Gly Glu
1  5

```
<210> SEQ ID NO 65
<211> LENGTH: 6
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<222> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 65
Gly Glu Glu Gly Ser Ser
1  5
```

Gly Glu Glu Gly Ser Ser
1  5

```
<210> SEQ ID NO 66
<211> LENGTH: 6
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<222> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 66
Gly Ser Ser Gly Glu Glu
1  5
```

Gly Ser Ser Gly Glu Glu
1  5

```
<210> SEQ ID NO 67
<211> LENGTH: 6
```
Ser Gly Ser Glu Gly Glu
1 5

<210> SEQ ID NO 68
<211> LENGTH: 6
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 68
Ser Ser Gly Glu Glu Gly
1 5

Glu Glu Gly Gly Gly Ser Ser Gly Glu Gly Glu Gly Ser Ser Ser
1 5 10 15

Gly Ser Glu Glu
20

<211> LENGTH: 20
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 69
Glu Glu Gly Gly Gly Ser Ser Gly Glu Gly Glu Gly Ser Ser Ser Ser
1 5 10 15

Gly Ser Glu Glu
20

<211> LENGTH: 20
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 70
Glu Ser Gly Ser Ser Glu Gly Ser Ser Glu Gly Ser Ser Glu Ser Ser Ser
1 5 10 15

Glu Gly Ser Glu
20

<211> LENGTH: 9
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 71
Glu Glu Glu Ser Ser Gly Gly Gly
1 5

<211> LENGTH: 6
Glu Glu Ser Ser Gly Gly  
1  5

Glu Ser Gly Ser Glu  
1  5

Glu Ser Gly Ser  
1  5

Glu Ser Gly Ser Gly  
1  5

Glu Ser Gly Glu Ser Gly  
1  5

Glu Ser Gly Glu Ser Gly  
1  5
<400> SEQUENCE: 77
Glu Ser Gly Pro Glu Ser Gly
1 5

<210> SEQ ID NO 78
<211> LENGTH: 11
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 78
Glu Gly Glu Gly Glu Gly Glu Gly Glu Gly Glu
1 5 10

<210> SEQ ID NO 79
<211> LENGTH: 14
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 79
Gly Ser Gly Ser Gly Ser Gly Ser Gly Ser Gly Ser Gly Ser
1 5 10

<210> SEQ ID NO 80
<211> LENGTH: 17
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 80
Ser Glu Ser Glu Ser Glu Ser Glu Ser Glu Ser Glu Ser Glu Ser Glu Ser
1 5 10 15

Ser

<210> SEQ ID NO 81
<211> LENGTH: 5
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 81
Asp Asp Asp Glu Glu
1 5

<210> SEQ ID NO 82
<211> LENGTH: 5
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 82
Asp Asp Asp Gly Glu
1 5
Amp Asp Asp Lys Lys
1   5

Amp Asp Asp Pro Pro
1   5

Amp Asp Asp Arg Arg
1   5

Amp Asp Asp Ser Ser
1   5

Amp Asp Asp Thr Thr
1   5
-continued

<220>  FEATURE:
<223>  OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400>  SEQUENCE: 88
Glu Glu Glu Aep Aep
1  5

<210>  SEQ ID NO 89
<211>  LENGTH: 5
<212>  TYPE: PRT
<213>  ORGANISM: Artificial Sequence
<220>  FEATURE:
<223>  OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400>  SEQUENCE: 89
Glu Glu Glu Gly Gly
1  5

<210>  SEQ ID NO 90
<211>  LENGTH: 5
<212>  TYPE: PRT
<213>  ORGANISM: Artificial Sequence
<220>  FEATURE:
<223>  OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400>  SEQUENCE: 90
Glu Glu Glu Lys Lys
1  5

<210>  SEQ ID NO 91
<211>  LENGTH: 5
<212>  TYPE: PRT
<213>  ORGANISM: Artificial Sequence
<220>  FEATURE:
<223>  OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400>  SEQUENCE: 91
Glu Glu Glu Pro Pro
1  5

<210>  SEQ ID NO 92
<211>  LENGTH: 5
<212>  TYPE: PRT
<213>  ORGANISM: Artificial Sequence
<220>  FEATURE:
<223>  OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400>  SEQUENCE: 92
Glu Glu Glu Arg Arg
1  5

<210>  SEQ ID NO 93
<211>  LENGTH: 5
<212>  TYPE: PRT
<213>  ORGANISM: Artificial Sequence
<220>  FEATURE:
<223>  OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400>  SEQUENCE: 93
<table>
<thead>
<tr>
<th>Glu  Glu  Glu  Ser  Ser</th>
</tr>
</thead>
<tbody>
<tr>
<td>1  5</td>
</tr>
</tbody>
</table>

```
<210> SEQ ID NO 94  
<211> LENGTH:  5  
<212> TYPE:     PRT  
<213> ORGANISM: Artificial Sequence  
<220> FEATURE:  
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide  
<400> SEQUENCE:  94  
Glu Glu Glu Thr Thr  
1  5  
```

<table>
<thead>
<tr>
<th>Gly  Gly  Gly  Asp  Asp</th>
</tr>
</thead>
<tbody>
<tr>
<td>1  5</td>
</tr>
</tbody>
</table>

```
<210> SEQ ID NO 95  
<211> LENGTH:  5  
<212> TYPE:     PRT  
<213> ORGANISM: Artificial Sequence  
<220> FEATURE:  
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide  
<400> SEQUENCE:  95  
Gly Gly Gly Asp Asp  
1  5  
```

<table>
<thead>
<tr>
<th>Gly  Gly  Gly  Glu</th>
</tr>
</thead>
<tbody>
<tr>
<td>1  5</td>
</tr>
</tbody>
</table>

```
<210> SEQ ID NO 96  
<211> LENGTH:  5  
<212> TYPE:     PRT  
<213> ORGANISM: Artificial Sequence  
<220> FEATURE:  
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide  
<400> SEQUENCE:  96  
Gly Gly Gly Glu Glu  
1  5  
```

<table>
<thead>
<tr>
<th>Gly  Gly  Gly  Lys  Lys</th>
</tr>
</thead>
<tbody>
<tr>
<td>1  5</td>
</tr>
</tbody>
</table>

```
<210> SEQ ID NO 97  
<211> LENGTH:  5  
<212> TYPE:     PRT  
<213> ORGANISM: Artificial Sequence  
<220> FEATURE:  
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide  
<400> SEQUENCE:  97  
Gly Gly Gly Lys Lys  
1  5  
```

<table>
<thead>
<tr>
<th>Gly  Gly  Gly  Pro  Pro</th>
</tr>
</thead>
<tbody>
<tr>
<td>1  5</td>
</tr>
</tbody>
</table>

```
<210> SEQ ID NO 98  
<211> LENGTH:  5  
<212> TYPE:     PRT  
<213> ORGANISM: Artificial Sequence  
<220> FEATURE:  
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide  
<400> SEQUENCE:  98  
Gly Gly Gly Pro Pro  
1  5  
```

```
<210> SEQ ID NO 99  
<211> LENGTH:  5  
```
<210> SEQ ID NO 100
<211> LENGTH: 5
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 100

Lys Lys Lys Asp Asp
1  5

<210> SEQ ID NO 101
<211> LENGTH: 5
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 101

Lys Lys Lys Glu Glu
1  5

<210> SEQ ID NO 102
<211> LENGTH: 5
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 102

Lys Lys Lys Gly Gly
1  5

<210> SEQ ID NO 103
<211> LENGTH: 5
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 103

Lys Lys Lys Pro Pro
1  5

<210> SEQ ID NO 104
<211> LENGTH: 5
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide
Lys Lys Lys Arg Arg  1  5

Lys Lys Lys Ser Ser  1  5

Lys Lys Lys Thr Thr  1  5

Pro Pro Pro Asp Asp  1  5

Pro Pro Pro Glu Glu  1  5

Pro Pro Pro Gly Gly  1  5
Pro Pro Pro Lys Lys
1 5

Pro Pro Pro Arg Arg
1 5

Pro Pro Pro Ser Ser
1 5

Pro Pro Pro Thr Thr
1 5

Arg Arg Arg Asp Asp
1 5
peptide
<400> SEQUENCE: 115
Arg Arg Arg Glu Glu 1 5

<210> SEQ ID NO 116
<211> LENGTH: 5
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide
<400> SEQUENCE: 116
Arg Arg Arg Gly Gly 1 5

<210> SEQ ID NO 117
<211> LENGTH: 5
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide
<400> SEQUENCE: 117
Arg Arg Arg Lys Lys 1 5

<210> SEQ ID NO 118
<211> LENGTH: 5
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide
<400> SEQUENCE: 118
Arg Arg Arg Pro Pro 1 5

<210> SEQ ID NO 119
<211> LENGTH: 5
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide
<400> SEQUENCE: 119
Arg Arg Arg Ser Ser 1 5

<210> SEQ ID NO 120
<211> LENGTH: 5
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide
<400> SEQUENCE: 120
Arg Arg Arg Thr Thr 1 5
<210> SEQ ID NO 121
<211> LENGTH: 5
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 121
Ser Ser Ser Asp Asp
1 5

<210> SEQ ID NO 122
<211> LENGTH: 5
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 122
Ser Ser Ser Glu Glu
1 5

<210> SEQ ID NO 123
<211> LENGTH: 5
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 123
Ser Ser Ser Gly Gly
1 5

<210> SEQ ID NO 124
<211> LENGTH: 5
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 124
Ser Ser Ser Lys Lys
1 5

<210> SEQ ID NO 125
<211> LENGTH: 5
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 125
Ser Ser Ser Pro Pro
1 5

<210> SEQ ID NO 126
<211> LENGTH: 5
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURES:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 126
Ser Ser Ser Arg Arg
1 5

<210> SEQ ID NO 127
<211> LENGTH: 5
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURES:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 127
Ser Ser Ser Thr Thr
1 5

<210> SEQ ID NO 128
<211> LENGTH: 5
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURES:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 128
Thr Thr Thr Asp Asp
1 5

<210> SEQ ID NO 129
<211> LENGTH: 5
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURES:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 129
Thr Thr Thr Glu Glu
1 5

<210> SEQ ID NO 130
<211> LENGTH: 5
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURES:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 130
Thr Thr Thr Gly Gly
1 5

<210> SEQ ID NO 131
<211> LENGTH: 5
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURES:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 131
Thr Thr Thr Lys Lys
1   5

<210> SEQ ID NO 132
<211> LENGTH: 5
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 132

Thr Thr Thr Pro Pro
1   5

<210> SEQ ID NO 133
<211> LENGTH: 5
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 133

Thr Thr Thr Arg Arg
1   5

<210> SEQ ID NO 134
<211> LENGTH: 5
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 134

Thr Thr Thr Ser Ser
1   5

<210> SEQ ID NO 135
<211> LENGTH: 7
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 135

Asp Asp Asp Asp Glu Glu Glu
1   5

<210> SEQ ID NO 136
<211> LENGTH: 7
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 136

Asp Asp Asp Gln Gly Gly
1   5

<210> SEQ ID NO 137
<211> LENGTH: 7
Asp Asp Asp Lys Lys Lys
1
5

Asp Asp Asp Pro Pro Pro
1
5

Asp Asp Asp Arg Arg Arg
1
5

Asp Asp Asp Ser Ser Ser
1
5

Asp Asp Asp Thr Thr Thr
1
5

Asp Asp Asp Thr Thr Thr
1
5
Glu Glu Glu Glu Asp Asp Asp
1 5

Glu Glu Glu Gly Gly Gly
1 5

Glu Glu Glu Lys Lys Lys
1 5

Glu Glu Glu Pro Pro Pro
1 5

Glu Glu Glu Arg Arg Arg
1 5

Glu Glu Glu Ser Ser Ser
1 5
Glu Glu Glu Glu Thr Thr Thr
1   5

Lys Lys Lys Lys Asp Asp Asp
1   5

Lys Lys Lys Glu Glu Glu
1   5

Lys Lys Lys Gly Gly Gly
1   5

Lys Lys Lys Pro Pro Pro
1   5
peptide

<400> SEQUENCE: 153

Lys Lys Lys Lys Arg Arg Arg
1 5

<210> SEQ ID NO 154
<211> LENGTH: 7
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 154

Lys Lys Lys Ser Ser Ser Ser
1 5

<210> SEQ ID NO 155
<211> LENGTH: 7
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 155

Lys Lys Lys Thr Thr Thr Thr
1 5

<210> SEQ ID NO 156
<211> LENGTH: 7
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 156

Pro Pro Pro Pro Asp Asp Asp
1 5

<210> SEQ ID NO 157
<211> LENGTH: 7
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 157

Pro Pro Pro Pro Glu Glu Glu
1 5

<210> SEQ ID NO 158
<211> LENGTH: 7
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 158

Pro Pro Pro Pro Gly Gly Gly
1 5
Pro Pro Pro Arg Arg Arg
1 5

Pro Pro Pro Ser Ser Ser
1 5

Pro Pro Pro Thr Thr Thr
1 5

Arg Arg Arg Asp Asp Asp
1 5
<220> FEATURES:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 164

Arg Arg Arg Arg Glu Glu Glu
1     5

<210> SEQ ID NO 165
<211> LENGTH: 7
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURES:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 165

Arg Arg Arg Gly Gly Gly
1     5

<210> SEQ ID NO 166
<211> LENGTH: 7
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURES:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 166

Arg Arg Arg Lys Lys Lys
1     5

<210> SEQ ID NO 167
<211> LENGTH: 7
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURES:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 167

Arg Arg Arg Pro Pro Pro
1     5

<210> SEQ ID NO 168
<211> LENGTH: 7
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURES:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 168

Arg Arg Arg Ser Ser Ser
1     5

<210> SEQ ID NO 169
<211> LENGTH: 7
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURES:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 169
Arg Arg Arg Thr Thr Thr
1  5

<210> SEQ ID NO 170
<211> LENGTH: 7
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 170
Ser Ser Ser Ser Asp Asp Asp
1  5

<210> SEQ ID NO 171
<211> LENGTH: 7
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 171
Ser Ser Ser Ser Glu Glu Glu
1  5

<210> SEQ ID NO 172
<211> LENGTH: 7
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 172
Ser Ser Ser Ser Gly Gly Gly
1  5

<210> SEQ ID NO 173
<211> LENGTH: 7
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 173
Ser Ser Ser Ser Lys Lys Lys
1  5

<210> SEQ ID NO 174
<211> LENGTH: 7
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 174
Ser Ser Ser Ser Pro Pro Pro
1  5

<210> SEQ ID NO 175
<211> LENGTH: 7
-continued

<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 175
Ser Ser Ser Ser Arg Arg Arg
1    5

<210> SEQ ID NO 176
<211> LENGTH: 7
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 176
Ser Ser Ser Ser Thr Thr Thr
1    5

<210> SEQ ID NO 177
<211> LENGTH: 7
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 177
Thr Thr Thr Thr Asp Asp Asp
1    5

<210> SEQ ID NO 178
<211> LENGTH: 7
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 178
Thr Thr Thr Thr Glu Glu Glu
1    5

<210> SEQ ID NO 179
<211> LENGTH: 7
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 179
Thr Thr Thr Thr Gly Gly Gly
1    5

<210> SEQ ID NO 180
<211> LENGTH: 7
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide
Thrrr Lys Lys
1 5

<210> SEQ ID NO 181
<211> LENGTH: 7
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 181
Thrrr Pro Pro Pro
1 5

<210> SEQ ID NO 182
<211> LENGTH: 7
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 182
Thrrr Arg Arg Arg
1 5

<210> SEQ ID NO 183
<211> LENGTH: 7
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 183
Thrrr Ser Ser Ser
1 5

<210> SEQ ID NO 184
<211> LENGTH: 4
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 184
Asp Asp Glu Glu
1

<210> SEQ ID NO 185
<211> LENGTH: 4
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 185
Asp Asp Gly Gly
1
Asp Asp Lys Lys
1

Asp Asp Pro Pro
1

Asp Asp Arg Arg
1

Asp Asp Ser Ser
1

Asp Asp Thr Thr
1
peptide

<400> SEQUENCE: 191

Glu Glu Asp Asp
1

<210> SEQ ID NO 192
<211> LENGTH: 4
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 192

Glu Glu Gly Gly
1

<210> SEQ ID NO 193
<211> LENGTH: 4
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 193

Glu Glu Lys Lys
1

<210> SEQ ID NO 194
<211> LENGTH: 4
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 194

Glu Glu Pro Pro
1

<210> SEQ ID NO 195
<211> LENGTH: 4
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 195

Glu Glu Arg Arg
1

<210> SEQ ID NO 196
<211> LENGTH: 4
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 196

Glu Glu Ser Ser
1
<210> SEQ ID NO 197
<211> LENGTH: 4
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide
<400> SEQUENCE: 197

Glu Glu Thr Thr

1

<210> SEQ ID NO 198
<211> LENGTH: 4
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide
<400> SEQUENCE: 198

Gly Gly Asp Asp

1

<210> SEQ ID NO 199
<211> LENGTH: 4
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide
<400> SEQUENCE: 199

Gly Gly Glu Glu

1

<210> SEQ ID NO 200
<211> LENGTH: 4
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide
<400> SEQUENCE: 200

Gly Gly Lys Lys

1

<210> SEQ ID NO 201
<211> LENGTH: 4
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide
<400> SEQUENCE: 201

Gly Gly Pro Pro

1

<210> SEQ ID NO 202
<211> LENGTH: 4
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURES:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 202
Gly Gly Arg Arg 1

<210> SEQ ID NO 203
<211> LENGTH: 4
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURES:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 203
Gly Gly Ser Ser 1

<210> SEQ ID NO 204
<211> LENGTH: 4
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURES:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 204
Gly Gly Thr Thr 1

<210> SEQ ID NO 205
<211> LENGTH: 4
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURES:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 205
Lys Lys Asp Asp 1

<210> SEQ ID NO 206
<211> LENGTH: 4
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURES:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 206
Lys Lys Glu Glu 1

<210> SEQ ID NO 207
<211> LENGTH: 4
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURES:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 207
Lys Lys Gly Gly

1

<210> SEQ ID NO 208
<211> LENGTH: 4
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 208

Lys Lys Pro Pro

1

<210> SEQ ID NO 209
<211> LENGTH: 4
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 209

Lys Lys Arg Arg

1

<210> SEQ ID NO 210
<211> LENGTH: 4
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 210

Lys Lys Ser Ser

1

<210> SEQ ID NO 211
<211> LENGTH: 4
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 211

Lys Lys Thr Thr

1

<210> SEQ ID NO 212
<211> LENGTH: 4
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 212

Pro Pro Asp Asp

1

<210> SEQ ID NO 213
<211> LENGTH: 4
<210> SEQ ID NO 214
<211> LENGTH: 4
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 214
Pro Pro Gly Gly
1

<210> SEQ ID NO 215
<211> LENGTH: 4
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 215
Pro Pro Lys Lys
1

<210> SEQ ID NO 216
<211> LENGTH: 4
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 216
Pro Pro Arg Arg
1

<210> SEQ ID NO 217
<211> LENGTH: 4
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 217
Pro Pro Ser Ser
1

<210> SEQ ID NO 218
<211> LENGTH: 4
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide
<400> SEQUENCE: 218

Pro Pro Thr Thr
1

<210> SEQ ID NO 219
<211> LENGTH: 4
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide
<400> SEQUENCE: 219

Arg Arg Asp Amp
1

<210> SEQ ID NO 220
<211> LENGTH: 4
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide
<400> SEQUENCE: 220

Arg Arg Glu Glu
1

<210> SEQ ID NO 221
<211> LENGTH: 4
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide
<400> SEQUENCE: 221

Arg Arg Gly Gly
1

<210> SEQ ID NO 222
<211> LENGTH: 4
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide
<400> SEQUENCE: 222

Arg Arg Lys Lys
1

<210> SEQ ID NO 223
<211> LENGTH: 4
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide
<400> SEQUENCE: 223

Arg Arg Pro Pro
1
Arg Arg Ser Ser
1

Arg Arg Thr Thr
1

Ser Ser Asp Asp
1

Ser Ser Glu Glu
1

Ser Ser Gly Gly
1
peptide

<400> SEQUENCE: 229

Ser Ser Lys Lys
1

<210> SEQ ID NO 230
<211> LENGTH: 4
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 230

Ser Ser Pro Pro
1

<210> SEQ ID NO 231
<211> LENGTH: 4
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 231

Ser Ser Arg Arg
1

<210> SEQ ID NO 232
<211> LENGTH: 4
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 232

Ser Ser Thr Thr
1

<210> SEQ ID NO 233
<211> LENGTH: 4
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 233

Thr Thr Asp Asp
1

<210> SEQ ID NO 234
<211> LENGTH: 4
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 234

Thr Thr Glu Glu
1
<210> SEQ ID NO 235
<211> LENGTH: 4
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 235
Thr Thr Gly Gly
1

<210> SEQ ID NO 236
<211> LENGTH: 4
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 236
Thr Thr Lys Lys
1

<210> SEQ ID NO 237
<211> LENGTH: 4
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 237
Thr Thr Pro Pro
1

<210> SEQ ID NO 238
<211> LENGTH: 4
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 238
Thr Thr Arg Arg
1

<210> SEQ ID NO 239
<211> LENGTH: 4
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 239
Thr Thr Ser Ser
1

<210> SEQ ID NO 240
<211> LENGTH: 5
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

SEQUENCE: 240
Gly Gly Gly Ser Ser
1
5

SEQ ID NO 241
LENGTH: 5
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

SEQUENCE: 241
Gly Gly Gly Thr Thr
1
5

SEQ ID NO 242
LENGTH: 5
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

SEQUENCE: 242
Asp Asp Glu Glu Glu
1
5

SEQ ID NO 243
LENGTH: 5
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

SEQUENCE: 243
Asp Asp Gly Gly Gly
1
5

SEQ ID NO 244
LENGTH: 5
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

SEQUENCE: 244
Asp Asp Lys Lys Lys
1
5

SEQ ID NO 245
LENGTH: 5
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

SEQUENCE: 245
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 251
Glu Glu Lys Lys Lys
1 5

<210> SEQ ID NO 252
<211> LENGTH: 5
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 252
Glu Glu Pro Pro Pro
1 5

<210> SEQ ID NO 253
<211> LENGTH: 5
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 253
Glu Glu Arg Arg Arg
1 5

<210> SEQ ID NO 254
<211> LENGTH: 5
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 254
Glu Glu Ser Ser Ser
1 5

<210> SEQ ID NO 255
<211> LENGTH: 5
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 255
Glu Glu Thr Thr Thr
1 5

<210> SEQ ID NO 256
<211> LENGTH: 5
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide
Gly Gly Asp Asp Asp
1  5

Gly Gly Glu Glu Glu
1  5

Gly Gly Lys Lys Lys
1  5

Gly Gly Pro Pro Pro
1  5

Gly Gly Arg Arg Arg
1  5

Gly Gly Ser Ser Ser
1  5
Gly Gly Thr Thr Thr
1 5

Lys Lys Asp Asp Asp
1 5

Lys Lys Glu Glu Glu
1 5

Lys Lys Gly Gly Gly
1 5

Lys Lys Pro Pro Pro
1 5
peptide

Lys Lys Arg Arg Arg
1 5

SEQ ID NO 268
LENGTH: 5
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE: OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

Lys Lys Ser Ser Ser
1 5

SEQ ID NO 269
LENGTH: 5
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE: OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

Lys Lys Thr Thr Thr
1 5

SEQ ID NO 270
LENGTH: 5
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE: OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

Pro Pro Asp Asp Asp
1 5

SEQ ID NO 271
LENGTH: 5
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE: OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

Pro Pro Glu Glu Glu
1 5

SEQ ID NO 272
LENGTH: 5
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE: OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

Pro Pro Gly Gly Gly
1 5
Pro Pro Lys Lys Lys
1 5

Pro Pro Arg Arg Arg
1 5

Pro Pro Ser Ser Ser Ser
1 5

Pro Pro Thr Thr Thr Thr
1 5

Arg Arg Asp Asp Asp
1 5
<220> FEATURES:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 278
Arg Arg Glu Glu Glu
1  5

<210> SEQ ID NO 279
<211> LENGTH: 5
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURES:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 279
Arg Arg Gly Gly Gly
1  5

<210> SEQ ID NO 280
<211> LENGTH: 5
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURES:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 280
Arg Arg Lys Lys Lys
1  5

<210> SEQ ID NO 281
<211> LENGTH: 5
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURES:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 281
Arg Arg Pro Pro Pro
1  5

<210> SEQ ID NO 282
<211> LENGTH: 5
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURES:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 282
Arg Arg Ser Ser Ser
1  5

<210> SEQ ID NO 283
<211> LENGTH: 5
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURES:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 283
<table>
<thead>
<tr>
<th>Arg</th>
<th>Arg</th>
<th>Thr</th>
<th>Thr</th>
<th>Thr</th>
<th>1</th>
<th>5</th>
</tr>
</thead>
</table>

```
<210> SEQ ID NO 284
<211> LENGTH: 5
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide
<400> SEQUENCE: 284
Ser Ser Asp Asp Asp
1  5
```

<table>
<thead>
<tr>
<th>Ser</th>
<th>Ser</th>
<th>Glu</th>
<th>Glu</th>
<th>Glu</th>
<th>1</th>
<th>5</th>
</tr>
</thead>
</table>

```
<210> SEQ ID NO 285
<211> LENGTH: 5
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide
<400> SEQUENCE: 285
Ser Ser Glu Glu Glu
1  5
```

<table>
<thead>
<tr>
<th>Ser</th>
<th>Ser</th>
<th>Gly</th>
<th>Gly</th>
<th>Gly</th>
<th>1</th>
<th>5</th>
</tr>
</thead>
</table>

```
<210> SEQ ID NO 286
<211> LENGTH: 5
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide
<400> SEQUENCE: 286
Ser Ser Gly Gly Gly
1  5
```

<table>
<thead>
<tr>
<th>Ser</th>
<th>Ser</th>
<th>Lys</th>
<th>Lys</th>
<th>Lys</th>
<th>1</th>
<th>5</th>
</tr>
</thead>
</table>

```
<210> SEQ ID NO 287
<211> LENGTH: 5
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide
<400> SEQUENCE: 287
Ser Ser Lys Lys Lys
1  5
```

<table>
<thead>
<tr>
<th>Ser</th>
<th>Ser</th>
<th>Pro</th>
<th>Pro</th>
<th>Pro</th>
<th>1</th>
<th>5</th>
</tr>
</thead>
</table>

```
<210> SEQ ID NO 288
<211> LENGTH: 5
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide
<400> SEQUENCE: 288
Ser Ser Pro Pro Pro
1  5
```

```
<210> SEQ ID NO 289
<211> LENGTH: 5
```
<table>
<thead>
<tr>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ser Ser Arg Arg Arg</td>
</tr>
<tr>
<td>Ser Ser Thr Thr Thr</td>
</tr>
<tr>
<td>Thr Thr Asp Asp Asp</td>
</tr>
<tr>
<td>Thr Thr Glu Glu Glu</td>
</tr>
<tr>
<td>Thr Thr Gly Gly Gly</td>
</tr>
</tbody>
</table>
-continued

<400> SEQUENCE: 294
Thr Thr Lys Lys Lys
1  5

<210> SEQ ID NO 295
<211> LENGTH: 5
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 295
Thr Thr Pro Pro Pro
1  5

<210> SEQ ID NO 296
<211> LENGTH: 5
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 296
Thr Thr Arg Arg Arg
1  5

<210> SEQ ID NO 297
<211> LENGTH: 5
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 297
Thr Thr Ser Ser Ser
1  5

<210> SEQ ID NO 298
<211> LENGTH: 6
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 298
Asp Asp Asp Glu Glu Glu
1  5

<210> SEQ ID NO 299
<211> LENGTH: 6
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 299
Asp Asp Asp Gly Gly Gly
1  5
Asp Asp Asp Lys Lys Lys
1 5

Asp Asp Asp Pro Pro Pro
1 5

Asp Asp Asp Arg Arg Arg
1 5

Asp Asp Asp Ser Ser Ser
1 5

Asp Asp Asp Thr Thr Thr
1 5

Asp Asp Asp Thr Thr Thr
1 5
peptide

<400> SEQUENCE: 305

Glu Glu Glu Asp Asp Asp
1 5

<210> SEQ ID NO 306
<211> LENGTH: 6
<212> TYPE: PRF
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 306

Glu Glu Glu Gly Gly Gly
1 5

<210> SEQ ID NO 307
<211> LENGTH: 6
<212> TYPE: PRF
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 307

Glu Glu Glu Lys Lys Lys
1 5

<210> SEQ ID NO 308
<211> LENGTH: 6
<212> TYPE: PRF
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 308

Glu Glu Glu Pro Pro Pro
1 5

<210> SEQ ID NO 309
<211> LENGTH: 6
<212> TYPE: PRF
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 309

Glu Glu Glu Arg Arg Arg
1 5

<210> SEQ ID NO 310
<211> LENGTH: 6
<212> TYPE: PRF
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 310

Glu Glu Glu Ser Ser Ser
1 5
Glu Glu Glu Thr Thr Thr
1 5

Gly Gly Gly Asp Asp Asp
1 5

Gly Gly Gly Glu Glu Glu
1 5

Gly Gly Gly Lys Lys Lys
1 5

Gly Gly Gly Pro Pro Pro
1 5
<220> FEATURES:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 316
Gly Gly Gly Arg Arg Arg
1  5

<210> SEQ ID NO 317
<211> LENGTH: 6
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURES:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 317
Gly Gly Gly Ser Ser Ser
1  5

<210> SEQ ID NO 318
<211> LENGTH: 6
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURES:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 318
Gly Gly Gly Thr Thr Thr
1  5

<210> SEQ ID NO 319
<211> LENGTH: 6
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURES:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 319
Lys Lys Lys Asp Asp Asp
1  5

<210> SEQ ID NO 320
<211> LENGTH: 6
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURES:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 320
Lys Lys Lys Glu Glu Glu
1  5

<210> SEQ ID NO 321
<211> LENGTH: 6
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURES:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 321
Lys Lys Lys Gly Gly Gly
1 5

<210> SEQ ID NO 322
<211> LENGTH: 6
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 322

Lys Lys Lys Pro Pro Pro
1 5

<210> SEQ ID NO 323
<211> LENGTH: 6
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 323

Lys Lys Lys Arg Arg Arg
1 5

<210> SEQ ID NO 324
<211> LENGTH: 6
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 324

Lys Lys Lys Ser Ser Ser
1 5

<210> SEQ ID NO 325
<211> LENGTH: 6
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 325

Lys Lys Lys Thr Thr Thr
1 5

<210> SEQ ID NO 326
<211> LENGTH: 6
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 326

Pro Pro Pro Asp Asp Asp
1 5

<210> SEQ ID NO 327
<211> LENGTH: 6
Pro Pro Pro Glu Glu Glu
1  5

Pro Pro Pro Gly Gly Gly
1  5

Pro Pro Pro Lys Lys Lys
1  5

Pro Pro Pro Arg Arg Arg
1  5

Pro Pro Pro Ser Ser Ser
1  5
<400> SEQUENCE: 332
Pro Pro Pro Thr Thr Thr
1 5

<410> SEQ ID NO 333
<411> LENGTH: 6
<412> TYPE: PRT
<413> ORGANISM: Artificial Sequence
<420> FEATURE:
<423> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 333
Arg Arg Arg Asp Asp Asp
1 5

<410> SEQ ID NO 334
<411> LENGTH: 6
<412> TYPE: PRT
<413> ORGANISM: Artificial Sequence
<420> FEATURE:
<423> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 334
Arg Arg Arg Glu Glu Glu
1 5

<410> SEQ ID NO 335
<411> LENGTH: 6
<412> TYPE: PRT
<413> ORGANISM: Artificial Sequence
<420> FEATURE:
<423> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 335
Arg Arg Arg Gly Gly Gly
1 5

<410> SEQ ID NO 336
<411> LENGTH: 6
<412> TYPE: PRT
<413> ORGANISM: Artificial Sequence
<420> FEATURE:
<423> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 336
Arg Arg Arg Lys Lys Lys
1 5

<410> SEQ ID NO 337
<411> LENGTH: 6
<412> TYPE: PRT
<413> ORGANISM: Artificial Sequence
<420> FEATURE:
<423> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 337
Arg Arg Arg Pro Pro Pro
1 5
Arg Arg Arg Ser Ser Ser
1 5

Arg Arg Arg Thr Thr Thr
1 5

Ser Ser Ser Asp Asp Asp
1 5

Ser Ser Ser Glu Glu Glu
1 5

Ser Ser Ser Gly Gly Gly
1 5
peptide  

<400> SEQUENCE: 343  
Ser Ser Ser Lys Lys Lys  
1  5  

<210> SEQ ID NO 344  
<211> LENGTH: 6  
<212> TYPE: PRT  
<213> ORGANISM: Artificial Sequence  
<220> FEATURE:  
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide  

<400> SEQUENCE: 344  
Ser Ser Ser Pro Pro Pro  
1  5  

<210> SEQ ID NO 345  
<211> LENGTH: 6  
<212> TYPE: PRT  
<213> ORGANISM: Artificial Sequence  
<220> FEATURE:  
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide  

<400> SEQUENCE: 345  
Ser Ser Ser Arg Arg Arg  
1  5  

<210> SEQ ID NO 346  
<211> LENGTH: 6  
<212> TYPE: PRT  
<213> ORGANISM: Artificial Sequence  
<220> FEATURE:  
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide  

<400> SEQUENCE: 346  
Ser Ser Ser Thr Thr Thr  
1  5  

<210> SEQ ID NO 347  
<211> LENGTH: 6  
<212> TYPE: PRT  
<213> ORGANISM: Artificial Sequence  
<220> FEATURE:  
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide  

<400> SEQUENCE: 347  
Thr Thr Thr Asp Asp Asp  
1  5  

<210> SEQ ID NO 348  
<211> LENGTH: 6  
<212> TYPE: PRT  
<213> ORGANISM: Artificial Sequence  
<220> FEATURE:  
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide  

<400> SEQUENCE: 348  
Thr Thr Thr Glu Glu Glu  
1  5
Thr Thr Thr Gly Gly Gly
1 5

Thr Thr Thr Lys Lys Lys
1 5

Thr Thr Thr Pro Pro Pro
1 5

Thr Thr Thr Arg Arg Arg
1 5

Thr Thr Thr Ser Ser Ser
1 5

Thr Thr Thr Ser Ser Ser
1 5
<220> FEATUER:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 354
Gly Gly Gly Gly Asp Asp Asp
1  5

<210> SEQ ID NO 355
<211> LENGTH: 7
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 355
Gly Gly Gly Gly Glu Glu Glu
1  5

<210> SEQ ID NO 356
<211> LENGTH: 7
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 356
Gly Gly Gly Gly Lys Lys Lys
1  5

<210> SEQ ID NO 357
<211> LENGTH: 7
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 357
Gly Gly Gly Gly Pro Pro Pro
1  5

<210> SEQ ID NO 358
<211> LENGTH: 7
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 358
Gly Gly Gly Gly Arg Arg Arg
1  5

<210> SEQ ID NO 359
<211> LENGTH: 7
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 359
Gly Gly Gly Gly Ser Ser Ser
1  5

<210> SEQ ID NO 360
<211> LENGTH: 7
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 360
Gly Gly Gly Gly Thr Thr Thr
1  5

<210> SEQ ID NO 361
<211> LENGTH: 7
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 361
Pro Pro Pro Pro Thr Thr Thr
1  5

<210> SEQ ID NO 362
<211> LENGTH: 12
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 362
Ser Ser Ser Glu Ser Ser Glu Ser Ser Ser Glu
1  5  10

<210> SEQ ID NO 363
<211> LENGTH: 9
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 363
Gly Gly Glu Gly Gly Gly Gly Gly Glu
1  5

<210> SEQ ID NO 364
<211> LENGTH: 24
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 364
Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu Ser
1  5  10  15

Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu
20
<210> SEQ ID NO 365
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 365
Ser Ser Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu
1     5     10    15

<210> SEQ ID NO 366
<211> LENGTH: 18
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 366
Ser Ser Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Ser Ser
1     5     10    15

Ser Glu

<210> SEQ ID NO 367
<211> LENGTH: 14
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 367
Ser Ser Ser Ser Glu Ser Ser Ser Ser Glu
1     5     10

<210> SEQ ID NO 368
<211> LENGTH: 12
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 368
Ser Ser Ser Ser Glu Ser Ser Ser Glu
1     5     10

<210> SEQ ID NO 369
<211> LENGTH: 24
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 369
Ser Ser Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Ser Ser
1     5     10    15
Ser Ser Glu Ser Ser Ser Glu
20
<210> SEQ ID NO 370
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 370
1 5 10 15

<210> SEQ ID NO 371
<211> LENGTH: 17
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 371
1 5 10 15

Gly

<210> SEQ ID NO 372
<211> LENGTH: 12
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 372
1 5 10

<210> SEQ ID NO 373
<211> LENGTH: 37
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 373
1 5 10 15

20 25 30

Gly Pro Gly Gly Gly
35

<210> SEQ ID NO 374
<211> LENGTH: 33
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 374
1 5 10 15

20 25 30

Gly
<210> SEQ ID NO 375
<211> LENGTH: 32
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 375

1  5  10  15
20  25  30

<210> SEQ ID NO 376
<211> LENGTH: 27
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 376

1  5  10  15
Gly Gly Ser Gly Gly Gly Gly Gly Gly Gly
20  25

<210> SEQ ID NO 377
<211> LENGTH: 25
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 377

1  5  10  15
Gly Gly Ser Gly Gly Gly Gly Gly
20  25

<210> SEQ ID NO 378
<211> LENGTH: 23
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 378

1  5  10  15
Gly Gly Gly Arg Gly Gly Gly
20

<210> SEQ ID NO 379
<211> LENGTH: 21
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 379

1  5  10  15
Gly Gly Gly Gly
20

<210> SEQ ID NO 380
<211> LENGTH: 18
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 380
1    5   10
Ser Gly

<210> SEQ ID NO 381
<211> LENGTH: 12
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 381
1    5   10

<210> SEQ ID NO 382
<211> LENGTH: 46
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 382
1    5   10
20   25  30
Ser Ser Gly Gly Ser Gly Thr Ala Gly Gly His Ser Gly
35   40  45

<210> SEQ ID NO 383
<211> LENGTH: 35
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 383
1    5    10
Ser Gly Ser Gly Gly Gly Gly Ser Thr Gly Gly gly Gly Thr Ala
20   25   30
Gly Gly Gly
35

<210> SEQ ID NO 384
<211> LENGTH: 32
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 384
Gly His Pro Gly Ser Gly Ser Gly Ser Gly Gly Gly Gly Gly Gly Gly
1    5    10
20   25   30

<210> SEQ ID NO 385
<211> LENGTH: 31
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 385
1    5    10
20   25   30
<210> SEQ ID NO 386
<211> LENGTH: 31
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 386

Gly Gly Gly Ser Gly Ser Thr Gly Gly Gly Ser Gly Arg Ala Gly
20 25 30

<210> SEQ ID NO 387
<211> LENGTH: 30
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 387

1 5 10 15

20 25 30

<210> SEQ ID NO 388
<211> LENGTH: 29
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 388

1 5 10 15

20 25

<210> SEQ ID NO 389
<211> LENGTH: 28
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 389

Gly Gly Gly Gly Thr Ser Ser Gly Gly Ser Gly Ser Gly Gly Gly
1 5 10 15

Gly Ser Gly Gly Gly Gly Gly Ser Gly Gly Ser Gly
20 25

<210> SEQ ID NO 390
<211> LENGTH: 27
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 390

1 5 10 15

Gly Gly Ser Gly Gly Gly Gly Arg Gly Ala Gly Gly
20 25

<210> SEQ ID NO 391
<211> LENGTH: 27
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 391

Gly Gly Gly Ala Ala Gly Ala Gly Gly Gly Ser Gly Ala Gly Gly
1  5  10  15
Gly Ser Gly Gly Ser Gly Gly Arg Gly Thr Gly
20  25

<210> SEQ ID NO 392
<211> LENGTH: 27
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 392

1  5  10  15
Ser Gly Ala Glu Gly Gly Gly Gly Ala Gly Gly
20  25

<210> SEQ ID NO 393
<211> LENGTH: 26
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 393

1  5  10  15
Gly Gly Gly Gly Gly Gly Ala Gly
20  25

<210> SEQ ID NO 394
<211> LENGTH: 26
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 394

1  5  10  15
Pro Gly Gly Gly Gly Gly Ala Gly
20  25

<210> SEQ ID NO 395
<211> LENGTH: 25
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 395

1  5  10  15
Gly Gly Gly Pro Gly Gly Gly Gly Gly
20  25

<210> SEQ ID NO 396
<211> LENGTH: 25
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 396
1  5  10  15
20  25

<210> SEQ ID NO 397
<211> LENGTH: 25
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 397
1  5  10  15
20  25

<210> SEQ ID NO 398
<211> LENGTH: 25
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 398
1  5  10  15
20  25

<210> SEQ ID NO 399
<211> LENGTH: 25
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 399
1  5  10  15
20  25

<210> SEQ ID NO 400
<211> LENGTH: 25
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 400
1  5  10  15
20  25

<210> SEQ ID NO 401
<211> LENGTH: 24
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 401
1  5  10  15
20

Ala Gly Gly Ala Gly Gly Arg Gly Gly
<210> SEQ ID NO 402
<211> LENGTH: 24
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 402
Gly Gly Gly Gly Gly Asp Ala Gly Gly Asp Ala Gly Gly Ala
1    5  10   15
Gly Gly Arg Ala Gly Arg Ala Gly
20

<210> SEQ ID NO 403
<211> LENGTH: 23
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 403
1    5   10   15
Gly Gly Gly Ser Ser Gly Gly
20

<210> SEQ ID NO 404
<211> LENGTH: 23
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 404
Gly Ser Gly Pro Gly Thr Gly Gly Gly Ser Gly Ser Gly Gly Gly Gly
1    5   10   15
Gly Gly Gly Ser Gly Gly
20

<210> SEQ ID NO 405
<211> LENGTH: 23
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 405
1    5   10   15
Gly Gly Gly Gly Gly Pro Gly
20

<210> SEQ ID NO 406
<211> LENGTH: 23
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 406
1    5   10   15
Gly Gly Gly Gly Gly Asp Gly
20

<210> SEQ ID NO 407
<211> LENGTH: 23
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens
Gly Gly Thr Arg Gly Gly Thr Arg Gly Gly Thr Arg Gly Gly Asp Arg
1  5  10  15
Gly Arg Gly Arg Gly Ala Gly
20

<210> SEQ ID NO 408
<211> LENGTH: 23
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 408
1  5  10  15
Ala Gly Gly Gly Gly Gly Gly
20

<210> SEQ ID NO 409
<211> LENGTH: 22
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 409
1  5  10  15
Gly Arg Gly Gly Gly Gly
20

<210> SEQ ID NO 410
<211> LENGTH: 22
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 410
1  5  10  15
Gly Arg Gly Gly Ala Gly
20

<210> SEQ ID NO 411
<211> LENGTH: 22
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 411
1  5  10  15
Gly Gly Gly Gly Gly Gly
20

<210> SEQ ID NO 412
<211> LENGTH: 22
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 412
1  5  10  15
Ser Gly Gly Ser Gly Gly
-continued

<210> SEQ ID NO 413
<211> LENGTH: 22
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 413

1  5   10  15
Gly Ser Ala Gly Ser Gly
20

<210> SEQ ID NO 414
<211> LENGTH: 22
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 414

Gly Gly Pro Gly Thr Gly Ser Gly Gly Gly Gly Gly Ala Gly Thr Gly Gly
1  5   10  15
Gly Ala Gly Gly Pro Gly
20

<210> SEQ ID NO 415
<211> LENGTH: 22
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 415

1  5   10  15
Gly Ser Ala Gly Gly Gly
20

<210> SEQ ID NO 416
<211> LENGTH: 21
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 416

1  5   10  15
Gly Gly Gly Ala Gly
20

<210> SEQ ID NO 417
<211> LENGTH: 21
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 417

1  5   10  15
Gly Gly Gly Gly Gly
20

<210> SEQ ID NO 418
<211> LENGTH: 21
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 418

Gly Pro Gly Ala Gly Ala Gly Ser Gly Ala Gly Ser Ser Gly Gly
1  5  10  15
Gly Gly Gly Pro Gly
20

<210> SEQ ID NO 419
<211> LENGTH: 21
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 419

1  5  10  15
Gly Ser Ser Gly Gly
20

<210> SEQ ID NO 420
<211> LENGTH: 21
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 420

1  5  10  15
Gly His Gly Ser Gly
20

<210> SEQ ID NO 421
<211> LENGTH: 21
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 421

1  5  10  15
Gly Pro Gly Pro Gly
20

<210> SEQ ID NO 422
<211> LENGTH: 21
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 422

Gly Ala Gly Ser Gly Gly Gly Ala Ala Gly Ala Gly Ala Gly Ser
1  5  10  15
Ala Gly Gly Gly Gly
20

<210> SEQ ID NO 423
<211> LENGTH: 21
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 423

1  5  10  15
Gly Gly Gly Gly Gly
20

<210> SEQ ID NO 424
<211> LENGTH: 21
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 424

1 5 10 15
Gly Arg Gly Arg Gly
20

<210> SEQ ID NO 425
<211> LENGTH: 21
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 425

1 5 10 15
Ser Gly Gly Gly Gly
20

<210> SEQ ID NO 426
<211> LENGTH: 21
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 426

Gly Gly Glu Glu Gly Gly Ala Ser Gly Gly Pro Gly Ala Gly Ser Gly
1 5 10 15
Gly Ser Ala Gly Gly
20

<210> SEQ ID NO 427
<211> LENGTH: 21
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 427

1 5 10 15
Gly Arg Gly Arg Gly
20

<210> SEQ ID NO 428
<211> LENGTH: 20
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 428

1 5 10 15
Ala Gly Ala Gly
20

<210> SEQ ID NO 429
-continued

<211> LENGTH: 20
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 429

1   5  10  15
Ser Gly His Gly
20

<210> SEQ ID NO 430
<211> LENGTH: 20
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 430

1   5  10  15
Gly Gly Gly Gly
20

<210> SEQ ID NO 431
<211> LENGTH: 20
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 431

Gly Gly Thr Gly Gly Ser Ser Gly Ser Gly Ser Gly Ser Gly Ser Gly Gly Gly Gly
1   5  10  15
Gly Arg Arg Gly
20

<210> SEQ ID NO 432
<211> LENGTH: 20
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 432

Gly Ser Gly Thr Gly Thr Gly Thr Gly Ser Ser Gly Ala Gly Gly Gly Gly
1   5  10  15
Thr Pro Gly Gly
20

<210> SEQ ID NO 433
<211> LENGTH: 20
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 433

1   5  10  15
Ala Gly Ala Gly
20

<210> SEQ ID NO 434
<211> LENGTH: 20
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 434
<table>
<thead>
<tr>
<th>Gly</th>
<th>Ser</th>
<th>Ser</th>
<th>Gly</th>
<th>Gly</th>
<th>Gly</th>
<th>Gly</th>
<th>Gly</th>
<th>Gly</th>
<th>Gly</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td>10</td>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gly</td>
<td>Gly</td>
<td>Ala</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**SEQ ID NO: 435**
LENGTH: 20
TYPE: PRT
ORGANISM: Homo sapiens

**SEQUENCE: 435**
1   5   10  15
Gly Gly Gly Gly
20

**SEQ ID NO: 436**
LENGTH: 20
TYPE: PRT
ORGANISM: Homo sapiens

**SEQUENCE: 436**
1   5   10  15
Gly Ala Gly Gly
20

**SEQ ID NO: 437**
LENGTH: 20
TYPE: PRT
ORGANISM: Homo sapiens

**SEQUENCE: 437**
Gly Ser Ala Gly Gly Ser Ser Gly Ala Gly Ala Gly Ala Gly Gly Gly
1   5   10  15
Ala Gly Ala Gly
20

**SEQ ID NO: 438**
LENGTH: 600
TYPE: PRT
ORGANISM: Artificial Sequence

**FEATURES:**
OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide
OTHER INFORMATION: This sequence may encompass 4 to 200 "Ser Ser Asp" repeating units

**SEQUENCE: 438**
Ser Ser Asp Ser Ser Asp Ser Ser Asp Ser Ser Asp Ser Ser Asp Ser Ser Asp Ser
1   5   10  15
Ser Asp Ser Ser Asp Ser Ser Asp Ser Ser Asp Ser Ser Asp Ser Ser Asp Ser
20  25  30
Asp Ser Ser Asp Ser Ser Asp Ser Ser Asp Ser Ser Asp Ser Ser Asp Ser
35  40  45
Ser Ser Asp Ser Ser Asp Ser Ser Asp Ser Ser Asp Ser Ser Asp Ser
50  55  60
Ser Asp Ser Ser Asp Ser Ser Asp Ser Ser Asp Ser Ser Asp Ser
65  70  75  80
<table>
<thead>
<tr>
<th>Ser</th>
<th>Ser</th>
<th>Ser</th>
<th>Ser</th>
<th>Ser</th>
<th>Ser</th>
<th>Ser</th>
<th>Ser</th>
<th>Ser</th>
<th>Ser</th>
<th>Ser</th>
<th>Ser</th>
<th>Ser</th>
<th>Ser</th>
<th>Ser</th>
<th>Ser</th>
<th>Ser</th>
<th>Ser</th>
<th>Ser</th>
<th>Ser</th>
<th>Ser</th>
</tr>
</thead>
<tbody>
<tr>
<td>85</td>
<td>90</td>
<td>95</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>105</td>
<td>110</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>115</td>
<td>120</td>
<td>125</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>130</td>
<td>135</td>
<td>140</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>145</td>
<td>150</td>
<td>155</td>
<td>160</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>165</td>
<td>170</td>
<td>175</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>180</td>
<td>185</td>
<td>190</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>195</td>
<td>200</td>
<td>205</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>210</td>
<td>215</td>
<td>220</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>225</td>
<td>230</td>
<td>235</td>
<td>240</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>245</td>
<td>250</td>
<td>255</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>260</td>
<td>265</td>
<td>270</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>275</td>
<td>280</td>
<td>285</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>290</td>
<td>295</td>
<td>300</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>305</td>
<td>310</td>
<td>315</td>
<td>320</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>325</td>
<td>330</td>
<td>335</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>340</td>
<td>345</td>
<td>350</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>355</td>
<td>360</td>
<td>365</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>370</td>
<td>375</td>
<td>380</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>385</td>
<td>390</td>
<td>395</td>
<td>400</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>405</td>
<td>410</td>
<td>415</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>420</td>
<td>425</td>
<td>430</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>435</td>
<td>440</td>
<td>445</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>450</td>
<td>455</td>
<td>460</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>465</td>
<td>470</td>
<td>475</td>
<td>480</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
<210> SEQ ID NO 439
<211> LENGTH: 1200
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide
<220> FEATURE:
<223> OTHER INFORMATION: This sequence may encompass 4 to 200 "Ser Ser Asp Ser Ser Asn" repeating units

<400> SEQUENCE: 439
Ser Ser Asp Ser Ser Asp Ser Ser Asp Ser Ser Asp Ser Ser Asp Ser
1     5         10          15
Ser Asn Ser Ser Asp Ser Ser Asn Ser Ser Asp Ser Ser Asn Ser Ser
5     10         15          20
Asp Ser Ser Asp Ser Ser Asp Ser Ser Asp Ser Ser Asn Ser Ser Asn
20     25         30          35
Ser Asn Ser Ser Asp Ser Ser Asn Ser Ser Asp Ser Ser Asn Ser Ser
35     40         45          50
Ser Ser Asp Ser Ser Asp Ser Ser Asp Ser Ser Asp Ser
50     55          60
Ser Asn Ser Ser Asp Ser Ser Asp Ser Ser Asp Ser Ser Asn Ser Ser
65     70         75          80
Asp Ser Ser Asn Ser Ser Asp Ser Ser Asp Ser Ser Asp Ser Ser Asn
85     90          95
Ser Ser Asp Ser Ser Asp Ser Ser Asp Ser Ser Asp Ser
100     105         110
Ser Asn Ser Ser Asp Ser Ser Asn Ser Ser Asp Ser Ser Asp Ser Ser
110    115         120          125
Asp Ser Ser Asp Ser Ser Asp Ser Ser Asp Ser Ser Asn Ser Ser Asn
125    130         135          140
Ser Ser Asp Ser Ser Asp Ser Ser Asp Ser Ser Asp Ser Ser Asp Ser
135    140         145          150
Ser Asn Ser Ser Asp Ser Ser Asn Ser Ser Asp Ser Ser Asn Ser Ser
145    150         155          160
Ser Asn Ser Ser Asp Ser Ser Asn Ser Ser Asp Ser Ser Asn Ser Ser
155    160         165          170
Asp Ser Ser Asp Ser Ser Asp Ser Ser Asp Ser Ser Asn Ser Ser Asn
170    175         180          185
Ser Ser Asp Ser Ser Asp Ser Ser Asp Ser Ser Asp Ser
185    190         195          200
Ser Ser Asp Ser Ser Asp Ser Ser Asp Ser Ser Asp Ser
200    205
<210> SEQ ID NO 440
<211> LENGTH: 690
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE: OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide
<220> FEATURE: OTHER INFORMATION: This sequence may encompass 4 to 200 "Ser Ser Glu" repeating units
<400> SEQUENCE: 440

Ser Ser Glu Ser Ser Glu Ser Ser Glu Ser Ser Glu Ser Ser Glu Ser Ser Glu Ser Ser Glu Ser Ser Glu Ser
1  5 10  15
Ser Glu Ser Ser Glu Ser Ser Glu Ser Ser Glu Ser Ser Glu Ser Ser Glu Ser Ser Glu Ser Ser Glu Ser
20 25 30
Glu Ser Glu Ser Ser Glu Ser Ser Glu Ser Ser Glu Ser Ser Glu Ser Ser Glu Ser Ser Glu Ser Ser Glu Ser
35 40 45
Ser Glu Ser Ser Glu Ser Ser Glu Ser Ser Glu Ser Ser Glu Ser Ser Glu Ser Ser Glu Ser Ser Glu Ser
50 55 60
Ser Glu Ser Ser Glu Ser Ser Glu Ser Ser Glu Ser Ser Glu Ser Ser Glu Ser Ser Glu Ser Ser Glu Ser
65 70 75 80
Glu Ser Ser Glu Ser Ser Glu Ser Ser Glu Ser Ser Glu Ser Ser Glu Ser Ser Glu Ser Ser Glu Ser Glu
85 90 95
Ser Ser Glu Ser Ser Glu Ser Ser Glu Ser Ser Glu Ser Ser Glu Ser Ser Glu Ser Ser Glu Ser Ser Glu Ser
100 105 110
Ser Glu Ser Ser Glu Ser Ser Glu Ser Ser Glu Ser Ser Glu Ser Ser Glu Ser Ser Glu Ser Ser Glu Ser
115 120 125
Ser Glu Ser Glu Ser Glu Ser Glu Ser Glu Ser Glu
Ser Glu Ser Glu Ser Glu Ser Glu Ser Glu Ser Glu Ser Glu Ser Glu Ser Glu Ser Glu Ser Glu Ser Glu Ser Glu Ser Glu Ser Glu Ser Glu Ser Glu Ser Glu Ser Glu Ser Glu Ser Glu Ser Glu Ser Glu
Ser Glu Ser Glu Ser Glu Ser Glu Ser Glu Ser Glu Ser Glu Ser Glu Ser Glu Ser Glu Ser Glu Ser Glu Ser Glu Ser Glu Ser Glu Ser Glu Ser Glu Ser Glu Ser Glu Ser Glu Ser Glu Ser Glu Ser Glu Ser Glu Ser Glu Ser Glu Ser Glu Ser Glu Ser Glu Ser Glu

<210> SEQ ID NO 441
<211> LENGTH: 40
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic poly peptide
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (2) .. (2)
<223> OTHER INFORMATION: Ser, Asp, Glu, Thr or Pro
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (4) .. (4)
<223> OTHER INFORMATION: Ser, Asp, Glu, Thr or Pro
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (6) .. (6)
<223> OTHER INFORMATION: Ser, Asp, Glu, Thr or Pro
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (8) .. (8)
<223> OTHER INFORMATION: Ser, Asp, Glu, Thr or Pro
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (10) .. (10)
<223> OTHER INFORMATION: Ser, Asp, Glu, Thr or Pro
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (12) .. (12)
<223> OTHER INFORMATION: Ser, Asp, Glu, Thr or Pro
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (14) .. (14)
<223> OTHER INFORMATION: Ser, Asp, Glu, Thr or Pro
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (16) .. (16)
<223> OTHER INFORMATION: Ser, Asp, Glu, Thr or Pro
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (18) .. (18)
<223> OTHER INFORMATION: Ser, Asp, Glu, Thr or Pro
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (20) .. (20)
<223> OTHER INFORMATION: Ser, Asp, Glu, Thr or Pro
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (22) .. (22)
<223> OTHER INFORMATION: Ser, Asp, Glu, Thr or Pro
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (24) .. (24)
<223> OTHER INFORMATION: Ser, Asp, Glu, Thr or Pro
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (26) .. (26)
<223> OTHER INFORMATION: Ser, Asp, Glu, Thr or Pro
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (28) .. (28)
<223> OTHER INFORMATION: Ser, Asp, Glu, Thr or Pro
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (30) .. (30)
<223> OTHER INFORMATION: Ser, Asp, Glu, Thr or Pro
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (32) .. (32)
<223> OTHER INFORMATION: Ser, Asp, Glu, Thr or Pro
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (34) .. (34)
<223> OTHER INFORMATION: Ser, Asp, Glu, Thr or Pro
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (36) .. (36)
<223> OTHER INFORMATION: Ser, Asp, Glu, Thr or Pro
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (38) .. (38)
<223> OTHER INFORMATION: Ser, Asp, Glu, Thr or Pro
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (40) .. (40)
<223> OTHER INFORMATION: Ser, Asp, Glu, Thr or Pro

<400> SEQUENCE: 441
Gly Xaa Gly Xaa Gly Xaa Gly Xaa Gly Xaa Gly Xaa Gly Xaa Gly Xaa
  1  5  10  15
Gly Xaa Gly Xaa Gly Xaa Gly Xaa Gly Xaa Gly Xaa Gly Xaa Gly Xaa
  20  25  30
Gly Xaa Gly Xaa Gly Xaa
  35  40

<210> SEQ ID NO: 442
<211> LENGTH: 39
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (3) .. (3)
<223> OTHER INFORMATION: Ser, Asp, Glu, Thr or Pro
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (6) .. (6)
<223> OTHER INFORMATION: Ser, Asp, Glu, Thr or Pro
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (9) .. (9)
<223> OTHER INFORMATION: Ser, Asp, Glu, Thr or Pro
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (12) .. (12)
<223> OTHER INFORMATION: Ser, Asp, Glu, Thr or Pro
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (15) .. (15)
<223> OTHER INFORMATION: Ser, Asp, Glu, Thr or Pro
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (18) .. (18)
<223> OTHER INFORMATION: Ser, Asp, Glu, Thr or Pro
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (21) .. (21)
<223> OTHER INFORMATION: Ser, Asp, Glu, Thr or Pro
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<223> OTHER INFORMATION: Ser, Asp, Glu, Thr or Pro
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (27) (27)
<223> OTHER INFORMATION: Ser, Asp, Glu, Thr or Pro
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (30) (30)
<223> OTHER INFORMATION: Ser, Asp, Glu, Thr or Pro
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (33) (33)
<223> OTHER INFORMATION: Ser, Asp, Glu, Thr or Pro
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (36) (36)
<223> OTHER INFORMATION: Ser, Asp, Glu, Thr or Pro
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (39) (39)
<223> OTHER INFORMATION: Ser, Asp, Glu, Thr or Pro
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (42) (42)
<223> OTHER INFORMATION: Ser, Asp, Glu, Thr or Pro
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (45) (45)

1  5  10  15
Gly Xaa Gly Gly Xaa Gly Xaa Gly Gly Xaa Gly Xaa Gly Xaa Gly
20  25
Xaa Gly Xaa Gly Xaa Gly Xaa
30  35

<210> SEQ ID NO 443
<211> LENGTH: 40
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (4) (4)
<223> OTHER INFORMATION: Ser, Asp, Glu, Thr or Pro
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (8) (8)
<223> OTHER INFORMATION: Ser, Asp, Glu, Thr or Pro
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (12) (12)
<223> OTHER INFORMATION: Ser, Asp, Glu, Thr or Pro
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (16) (16)
<223> OTHER INFORMATION: Ser, Asp, Glu, Thr or Pro
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (20) (20)
<223> OTHER INFORMATION: Ser, Asp, Glu, Thr or Pro
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (24) (24)
<223> OTHER INFORMATION: Ser, Asp, Glu, Thr or Pro
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (28) (28)
<223> OTHER INFORMATION: Ser, Asp, Glu, Thr or Pro
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (32) (32)
<223> OTHER INFORMATION: Ser, Asp, Glu, Thr or Pro
<220> FEATURE:
<221> NAME/KEY: MOD_RES
1  5  10  15

20  25  30

Gly Gly Gly Xaa Gly Gly Gly Xaa
35  40
OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide

NAME/KEY: MOD_RES
LOCATION: (31) (31)

OTHER INFORMATION: Ser, Asp, Glu, Thr or Pro

FEATURE:

NAME/KEY: MOD_RES
LOCATION: (41) (41)

OTHER INFORMATION: This region may encompass 1 to 20 "Gly" repeating residues

FEATURE:

NAME/KEY: MOD_RES
LOCATION: (42) (42)

OTHER INFORMATION: Ser, Asp, Glu, Thr or Pro

FEATURE:

NAME/KEY: MOD_RES
LOCATION: (43) (43)

OTHER INFORMATION: This region may encompass 1 to 20 "Gly" repeating residues

FEATURE:

NAME/KEY: MOD_RES
LOCATION: (44) (44)

OTHER INFORMATION: Ser, Asp, Glu, Thr or Pro

FEATURE:

NAME/KEY: MOD_RES
LOCATION: (63) (63)

OTHER INFORMATION: This region may encompass 1 to 20 "Gly" repeating residues

FEATURE:

NAME/KEY: MOD_RES
LOCATION: (64) (64)

OTHER INFORMATION: Ser, Asp, Glu, Thr or Pro

FEATURE:

NAME/KEY: MOD_RES
LOCATION: (84) (84)

OTHER INFORMATION: This region may encompass 1 to 20 "Gly" repeating residues

FEATURE:

NAME/KEY: MOD_RES
LOCATION: (85) (85)

OTHER INFORMATION: Ser, Asp, Glu, Thr or Pro

FEATURE:

NAME/KEY: MOD_RES
LOCATION: (105) (105)

OTHER INFORMATION: This region may encompass 1 to 20 "Gly" repeating residues

FEATURE:

NAME/KEY: MOD_RES
LOCATION: (106) (106)

OTHER INFORMATION: Ser, Asp, Glu, Thr or Pro

FEATURE:

NAME/KEY: MOD_RES
LOCATION: (125) (125)

OTHER INFORMATION: This region may encompass 1 to 20 "Gly" repeating residues

FEATURE:

NAME/KEY: MOD_RES
LOCATION: (126) (126)

OTHER INFORMATION: Ser, Asp, Glu, Thr or Pro

FEATURE:

NAME/KEY: MOD_RES
LOCATION: (127) (127)

OTHER INFORMATION: This region may encompass 1 to 20 "Gly" repeating residues

FEATURE:

NAME/KEY: MOD_RES
LOCATION: (147) (147)

OTHER INFORMATION: Ser, Asp, Glu, Thr or Pro

FEATURE:

NAME/KEY: MOD_RES
LOCATION: (148) (148)

OTHER INFORMATION: This region may encompass 1 to 20 "Gly" repeating residues

FEATURE:

NAME/KEY: MOD_RES
LOCATION: (168) (168)

OTHER INFORMATION: Ser, Asp, Glu, Thr or Pro

FEATURE:

NAME/KEY: MOD_RES
LOCATION: (169) (169)

OTHER INFORMATION: This region may encompass 1 to 20 "Gly" repeating residues

FEATURE:

NAME/KEY: MOD_RES
LOCATION: (187) (187)

OTHER INFORMATION: This region may encompass 1 to 20 "Gly" repeating residues

FEATURE:
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (189) .. (189)
<223> OTHER INFORMATION: Ser, Asp, Glu, Thr or Pro

<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (190) .. (209)
<223> OTHER INFORMATION: This region may encompass 1 to 20 "Gly" repeating residues

<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (210) .. (210)
<223> OTHER INFORMATION: Ser, Asp, Glu, Thr or Pro

<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (211) .. (230)
<223> OTHER INFORMATION: This region may encompass 1 to 20 "Gly" repeating residues

<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (231) .. (251)
<223> OTHER INFORMATION: This region may encompass 1 to 20 "Gly" repeating residues

<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (252) .. (252)
<223> OTHER INFORMATION: Ser, Asp, Glu, Thr or Pro

<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (253) .. (272)
<223> OTHER INFORMATION: This region may encompass 1 to 20 "Gly" repeating residues

<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (273) .. (273)
<223> OTHER INFORMATION: Ser, Asp, Glu, Thr or Pro

<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (274) .. (293)
<223> OTHER INFORMATION: This region may encompass 1 to 20 "Gly" repeating residues

<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (294) .. (294)
<223> OTHER INFORMATION: Ser, Asp, Glu, Thr or Pro

<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (295) .. (314)
<223> OTHER INFORMATION: This region may encompass 1 to 20 "Gly" repeating residues

<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (315) .. (315)
<223> OTHER INFORMATION: Ser, Asp, Glu, Thr or Pro

<400> SEQUENCE: 445

1 5 10 15
20 25 30
35 40 45
50 55 60
65 70 75 80
85 90 95
100 105 110
115 120 125
130 135 140
145 150 155 160
165 170 175
180 185 190
195 200 205
210 215 220
225 230 235 240
245 250 255
260 265 270
275 280 285
290 295 300
305 310 315

<210> SEQ ID NO 446
<211> LENGTH: 4
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 446

Lys Cys Lys Lys
1

<210> SEQ ID NO 447
<211> LENGTH: 4
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 447

Asp Asp Asp Lys
1

<210> SEQ ID NO 448
<211> LENGTH: 4
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 448

Asp Asp Asp Lys
1
Ile Asp Gly Arg
1

<210> SEQ ID NO 449
<211> LENGTH: 6
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 449
Leu Val Pro Arg Gly Ser
1  5

<210> SEQ ID NO 450
<211> LENGTH: 8
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 450
Leu Glu Val Leu Phe Gln Gly Pro
1  5

<210> SEQ ID NO 451
<211> LENGTH: 7
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 451
Glu Glu Leu Tyr Phe Gln Gly
1  5

<210> SEQ ID NO 452
<211> LENGTH: 7
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 452
Glu Thr Leu Phe Gln Gly Pro
1  5

<210> SEQ ID NO 453
<211> LENGTH: 5
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 453
Leu Pro Glu Thr Gly
1  5

<210> SEQ ID NO 454
<211> LENGTH: 191
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 454
Phe Pro Thr Ile Pro Leu Ser Arg Leu Phe Asp Asn Ala Met Leu Arg
1  5  10  15

Ala His Arg Leu His Gln Leu Ala Phe Asp Thr Tyr Gln Glu Phe Glu
20  25

Glu Ala Tyr Ile Pro Lys Glu Gln Lys Tyr Ser Phe Leu Gln Asn Pro
35  40  45

Gln Thr Ser Leu Cys Phe Ser Glu Ser Ile Pro Thr Pro Ser Asn Arg
<table>
<thead>
<tr>
<th></th>
<th>50</th>
<th>55</th>
<th>60</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glu</td>
<td>Glu</td>
<td>Thr</td>
<td>Gln</td>
</tr>
<tr>
<td></td>
<td>65</td>
<td>70</td>
<td>75</td>
</tr>
<tr>
<td>Leu</td>
<td>Leu</td>
<td>Ile</td>
<td>Gln</td>
</tr>
<tr>
<td></td>
<td>95</td>
<td>90</td>
<td>95</td>
</tr>
<tr>
<td>Phe</td>
<td>Ala</td>
<td>Asn</td>
<td>Ser</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>105</td>
<td>110</td>
</tr>
<tr>
<td>Leu</td>
<td>Leu</td>
<td>Lys</td>
<td>Asp</td>
</tr>
<tr>
<td></td>
<td>115</td>
<td>120</td>
<td>125</td>
</tr>
<tr>
<td>Glu</td>
<td>Asp</td>
<td>Gly</td>
<td>Ser</td>
</tr>
<tr>
<td></td>
<td>130</td>
<td>135</td>
<td>140</td>
</tr>
<tr>
<td>Lys</td>
<td>Phe</td>
<td>Asp</td>
<td>Thr</td>
</tr>
<tr>
<td></td>
<td>145</td>
<td>150</td>
<td>155</td>
</tr>
<tr>
<td>Gly</td>
<td>Leu</td>
<td>Leu</td>
<td>Tyr</td>
</tr>
<tr>
<td></td>
<td>165</td>
<td>170</td>
<td>175</td>
</tr>
<tr>
<td>Leu</td>
<td>Arg</td>
<td>Ile</td>
<td>Val</td>
</tr>
<tr>
<td></td>
<td>180</td>
<td>185</td>
<td>190</td>
</tr>
</tbody>
</table>

<210> SEQ ID NO 455
<211> LENGTH: 288
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide

<400> SEQUENCE: 455

  1   2   3   4   5   6   7   8   9  10  11  12  13  14  15

  16  17  18  19  20  21  22  23  24  25  26  27  28  29  30

  31  32  33  34  35  36  37  38  39  40  41  42  43  44  45

  46  47  48  49  50  51  52  53  54  55  56  57  58  59  60

  61  62  63  64  65  66  67  68  69  70  71  72  73  74  75

  76  77  78  79  80  81  82  83  84  85  86  87  88  89  90

  91  92  93  94  95  96  97  98  99 100 101 102 103 104 105

 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120

 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135

 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150

 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165

 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180

 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195

 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210
-continued

210 235 220
225 230 235 240
245 250 255
260 265 270
275 280 285

<210> SEQ ID NO 456
<211> LENGTH: 36
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 456
aggtagtgw ggwgargwg wgtcyggwgg agaag 36

<210> SEQ ID NO 457
<211> LENGTH: 36
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 457
acctctct cwwcrgawc cwwwctcwc wccact 36

<210> SEQ ID NO 458
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 458
agttgctct tccactcgaggtgtc 24

<210> SEQ ID NO 459
<211> LENGTH: 16
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 459
cctcgagta gacga 16

<210> SEQ ID NO 460
<211> LENGTH: 289
<212> TYPE: PepT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide
<400> SEQUENCE: 461
Ser Ser Ser Glu Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu
  1   5   10   15
Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu
  20  25   30  
Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu
  35  40   45  
Ser Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu
  50  55   60  
Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu
  45  70  75   80  
Ser Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu
  85  90   95  
Ser Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu
 100 105  110  
Ser Ser Ser Glu Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu
 115 120  125  
Ser Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu
 130 135  140  
Ser Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu
 145 150  155  160  
Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu
 165 170  175  
Ser Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu
 180 185  190  
Ser Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu
 195 200  205  
Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu
 210 215  220  
Ser Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu
 225 230  235  240  
Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu
 245 250  255  
Ser Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu
 260 265  270  
Ser Ser Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu
 275 280  285  

<210> SEQ ID NO 461
<211> LENGTH: 36
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 461
Ser Ser Ser Glu Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu
  1   5   10   15
Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu
  20  25  30  
Ser Ser Ser Glu
<210> SEQ ID NO 462
<211> LENGTH: 12
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

Ser Ser Glu Ser Ser Glu Ser Ser Glu Ser
1 5 10

<210> SEQ ID NO 463
<211> LENGTH: 36
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 463

ttctagtgar tcysgycggt cyagtcyag ygaattc 36

<210> SEQ ID NO 464
<211> LENGTH: 36
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 464

agaagattc ctryarctrg ayrctcrga ytcact 36

<210> SEQ ID NO 465
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 465

ttctctgct tcactcagg gtac 24

<210> SEQ ID NO 466
<211> LENGTH: 288
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide

<400> SEQUENCE: 466

1 5 10 15
20 25 30
35 40 45
50 55 60
65    70    75    80
85    90    95
100   105   110
115   120   125
130   135   140
145   150   155   160
165   170   175
180   185   190
195   200   205
210   215   220
225   230   235   240
245   250   255
260   265   270
275   280   285

<210> SEQ ID NO 467
<211> LENGTH: 27
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 467
aggtgaagw garggwgwgg wgwaagg

<210> SEQ ID NO 468
<211> LENGTH: 27
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 468
acctccttc ccwccwccyt ccwcttc

<210> SEQ ID NO 469
<211> LENGTH: 7
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (7)...(7)
OTHER INFORMATION: Any amino acid

SEQUENCE: 469

Glu Asn Leu Tyr Phe Gln Xaa
1 5

SEQ ID NO: 470
LENGTH: 165
TYPE: PRT
ORGANISM: Homo sapiens

SEQUENCE: 470
Cys Asp Leu Pro Gln Thr His Ser Leu Gly Ser Arg Arg Thr Leu Met
1 5 10 15
Leu Leu Ala Gln Met Arg Lys Ile Ser Leu Phe Ser Cys Leu Lys Asp
20 25 30
Arg His Asp Phe Gly Phe Pro Gln Glu Gly Phe Gly Asn Gln Phe Gln
35 40 45
Lys Ala Glu Thr Ile Pro Val Leu His Glu Met Ile Gln Gln Ile Phe
50 55 60
Ann Leu Phe Ser Thr Lys Asp Ser Ser Ala Ala Trp Asp Glu Thr Leu
65 70 75 80
Leu Asp Lys Phe Tyr Thr Glu Leu Tyr Gln Gln Leu Ann Asp Leu Glu
85 90 95
Ala Cys Val Ile Gln Gly Val Gly Val Thr Glu Thr Pro Leu Met Lys
100 105 110
Glu Asp Ser Ile Leu Ala Val Arg Lys Tyr Phe Gly Arg Ile Thr Leu
115 120 125
Tyr Leu Lys Glu Lys Tyr Ser Pro Cys Ala Trp Glu Val Arg
130 135 140
Ala Glu Ile Met Arg Ser Phe Ser Leu Ser Thr Asn Leu Gln Glu Ser
145 150 155 160
Leu Arg Ser Lys Glu
165

SEQ ID NO: 471
LENGTH: 174
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURES:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide

SEQUENCE: 471
Thr Pro Leu Gly Pro Ala Ser Ser Leu Pro Gln Ser Phe Leu Leu Lys
1 5 10 15
Cys Leu Glu Gln Val Arg Lys Ile Gln Gly Asp Gly Ala Ala Leu Gln
20 25 30
Glu Lys Leu Cys Ala Thr Tyr Lys Leu Cys His Pro Glu Glu Leu Val
35 40 45
Leu Leu Gly His Ser Leu Gly Ile Pro Trp Ala Pro Leu Ser Ser Cys
50 55 60
Pro Ser Gln Ala Leu Gln Leu Ala Gly Cys Leu Ser Gln Leu His Ser
65 70 75 80
Gly Leu Phe Leu Tyr Gln Gly Leu Leu Gln Ala Leu Glu Gly Ile Ser
85 90 95
-continued

Pro Glu Leu Gly Pro Thr Leu Asp Thr Leu Gln Leu Asp Val Ala Asp 100 105 110
Fhe Ala Thr Thr Ile Trp Gln Gln Met Glu Leu Gly Met Ala Pro 115 120 125
Ala Leu Gln Pro Thr Glu Ala Met Pro Ala Phe Ala Ser Ala Phe 130 135 140
Gln Arg Arg Ala Gly Gly Val Leu Val Ala Ser His Leu Gln Ser Phe 145 150 155 160
Leu Glu Val Ser Tyr Arg Val Leu Arg His Leu Ala Gln Pro 165 170

<210> SEQ ID NO: 472
<211> LENGTH: 288
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<222> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide

<400> SEQUENCE: 472
275 280 285

<210> SEQ ID NO 473
<211> LENGTH: 288
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<222> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide

<400> SEQUENCE: 473

Gly Glu Gly Glu Gly Gly Glu Gly Gly Gly Glu Gly Glu Gly Glu Gly
1   5  10 15
Glu Gly Glu Gly Glu Gly Gly Glu Gly Gly Gly Glu Gly Glu Gly
20  25  30
35  40  45
Glu Gly Glu Gly Glu Gly Gly Glu Gly Gly Gly Glu Gly Glu Gly
50  55  60
Glu Gly Glu Gly Glu Gly Gly Glu Gly Gly Gly Glu Gly Glu Gly Glu
65  70  75  80
Gly Glu Gly Glu Gly Glu Gly Gly Glu Gly Glu Gly Glu Gly Glu
85  90  95
Gly Glu Gly Glu Gly Glu Gly Gly Glu Gly Glu Gly Glu Gly Glu
100 105 110
Gly Glu Gly Glu Gly Glu Gly Gly Glu Gly Glu Gly Glu Gly Glu
115 120 125
Gly Glu Gly Glu Gly Glu Gly Gly Glu Gly Glu Gly Glu Gly Glu
130 135 140
Gly Glu Gly Glu Gly Glu Gly Gly Glu Gly Glu Gly Glu Gly Glu
145 150 155 160
Gly Glu Gly Glu Gly Glu Gly Gly Glu Gly Glu Gly Glu Gly Glu
165 170 175
Gly Glu Gly Glu Gly Glu Gly Gly Glu Gly Glu Gly Glu Gly Glu
180 185 190
Glu Gly Glu Gly Glu Gly Glu Gly Gly Glu Gly Glu Gly Glu Gly Glu
195 200 205
Glu Gly Glu Gly Glu Gly Glu Gly Gly Glu Gly Glu Gly Glu Gly Glu
210 215 220
Gly Glu Gly Glu Gly Glu Gly Gly Glu Gly Glu Gly Glu Gly Glu
225 230 235 240
Gly Glu Gly Glu Gly Glu Gly Gly Glu Gly Glu Gly Glu Gly Glu
245 250 255
Gly Glu Gly Glu Gly Glu Gly Gly Glu Gly Glu Gly Glu Gly Glu
260 265 270
Gly Glu Gly Glu Gly Glu Gly Gly Glu Gly Glu Gly Glu Gly Glu
275 280 285
<table>
<thead>
<tr>
<th></th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Ser Ser Glu Ser Ser Glu Ser Ser Ser Glu Ser Ser Glu Ser</td>
</tr>
<tr>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>20</td>
<td>25</td>
</tr>
<tr>
<td>35</td>
<td>40</td>
</tr>
<tr>
<td>50</td>
<td>60</td>
</tr>
<tr>
<td>65</td>
<td>70</td>
</tr>
<tr>
<td>90</td>
<td>95</td>
</tr>
<tr>
<td>105</td>
<td>110</td>
</tr>
<tr>
<td>115</td>
<td>120</td>
</tr>
<tr>
<td>130</td>
<td>135</td>
</tr>
<tr>
<td>145</td>
<td>150</td>
</tr>
<tr>
<td>165</td>
<td>170</td>
</tr>
<tr>
<td>185</td>
<td>190</td>
</tr>
<tr>
<td>205</td>
<td></td>
</tr>
<tr>
<td>210</td>
<td></td>
</tr>
<tr>
<td>225</td>
<td></td>
</tr>
<tr>
<td>245</td>
<td></td>
</tr>
<tr>
<td>260</td>
<td></td>
</tr>
<tr>
<td>275</td>
<td></td>
</tr>
</tbody>
</table>

**SEQ ID NO:** 475
**LENGTH:** 336
**TYPE:** PRT
**ORGANISM:** Artificial Sequence
**FEATURE:**
**OTHER INFORMATION:** Description of Artificial Sequence: Synthetic polypeptide
-continued

Ser Glu Ser Ser Ser Ser Ser Glu Ser Ser Ser Ser Ser Glu 35 40 45
Ser Ser Ser Ser Ser Glu Ser Ser Ser Ser Ser Glu Ser 50 55 60
Ser Ser Ser Ser Glu Ser Ser Ser Ser Ser Glu Ser Ser 65 70 75 80
Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Ser Ser Ser 85 90 95
Glu Ser Ser Ser Ser Glu Ser Ser Ser Ser Ser Ser Glu Ser 100 105 110
Ser Ser Ser Ser Glu Ser Ser Ser Ser Ser Glu Ser Ser 115 120 125
Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Ser Ser Ser 130 135 140
Ser Glu Ser Ser Ser Ser Glu Ser Ser Ser Ser Ser 145 150 155 160
Ser Ser Ser Ser Glu Ser Ser Ser Ser Ser Glu Ser Ser 165 170 175
Ser Ser Ser Ser Glu Ser Ser Ser Ser Ser Glu Ser Ser 180 185 190
Ser Ser Glu Ser Ser Ser Ser Glu Ser Ser Ser Ser Ser 195 200 205
Glu Ser Ser Ser Ser Ser Glu Ser Ser Ser Ser Ser 210 215 220
Ser Ser Ser Ser Glu Ser Ser Ser Ser Ser Glu Ser Ser 225 230 235 240
Ser Ser Glu Ser Ser Ser Ser Glu Ser Ser Ser Ser Ser 245 250 255
Ser Glu Ser Ser Ser Ser Glu Ser Ser Ser Ser Ser 260 265 270
Ser Ser Ser Ser Ser Glu Ser Ser Ser Ser Ser Glu Ser 275 280 285
Ser Ser Ser Ser Glu Ser Ser Ser Ser Ser Glu Ser Ser 290 295 300
Ser Ser Glu Ser Ser Ser Ser Glu Ser Ser Ser Ser Ser 305 310 315 320
Glu Ser Ser Ser Ser Glu Ser Ser Ser Ser Ser Glu Ser 325 330 335

<210> SEQ ID NO 476
<211> LENGTH: 320
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide

<400> SEQUENCE: 476
Ser Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser 1 5 10 15
Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser 20 25 30
Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Ser 35 40 45
Glu Ser Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu Ser 50 55 60
Ser Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu Ser 65 70 75 80
Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu Ser 95 100 90 95
Ser Ser Glu Ser Ser Glu Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser 100 105 110
Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Ser Ser Ser 115 120 125
Glu Ser Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Ser Ser Glu Ser 130 135 140
Ser Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Ser Ser Ser Glu Ser 145 150 155 160
Ser Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Ser Ser Ser Glu Ser 165 170 175
Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Ser Ser Ser Glu Ser Ser Ser 180 185 190
Ser Glu Ser Ser Ser Glu Ser Ser Ser Ser Ser Ser Glu Ser Ser Ser 195 200 205
Glu Ser Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Ser Ser Glu Ser 210 215 220
Ser Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Ser Ser Ser Glu Ser 225 230 235 240
Ser Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Ser Ser Ser Glu Ser 245 250 255
Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Ser Ser Ser Glu Ser Ser Ser 260 265 270
Ser Glu Ser Ser Ser Glu Ser Ser Ser Ser Ser Ser Glu Ser Ser Ser 275 280 285
Glu Ser Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Ser Ser Glu Ser 290 295 300
Ser Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Ser Ser Glu Ser Ser Ser 305 310 315 320

<210> SEQ ID NO 477
<211> LENGTH: 280
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<222> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide
<400> SEQUENCE: 477

Gly Glu Gly Ser Gly Glu Gly Ser Gly Glu Gly Ser Gly 1 5 10 15
Gly Glu Gly Glu Gly Ser Gly Glu Gly Glu Gly Glu Gly Glu Gly Glu 20 25 30
Glu Gly Ser Glu Gly Gly Ser Gly Glu Gly Gly Glu Gly Gly 35 40 45
Gly Ser Gly Ser Gly Glu Gly Ser Gly Glu Gly Ser Gly Glu Gly 50 55 60
Glu Gly Gly Ser Gly Ser Gly Glu Gly Gly Ser Gly Glu Gly Ser
-continued

65    70    75    80
Glu Gly Glu Gly Ser Glu Gly Glu Gly Ser Glu Gly Glu Gly Ser Glu
85           90          95
Gly Ser Glu Gly Glu Gly Ser Glu Gly Ser Glu Gly Glu Gly Gly
100         105         110
Glu Gly Glu Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Glu
115        120        125
Gly Ser Gly Glu Gly Ser Glu Gly Ser Glu Gly Ser Gly Ser Gly Ser Gly Glu
130       135       140
Gly Glu Gly Ser Gly Ser Gly Glu Gly Ser Glu Gly Ser Gly Ser Gly Ser
145       150       155       160
Gly Glu Gly Glu Gly Ser Glu Gly Ser Glu Gly Ser Gly Gly Ser
165       170       175
Glu Gly Ser Glu Gly Glu Gly Ser Glu Gly Ser Glu Gly Ser Gly Glu Gly
180       185       190
Gly Glu Gly Ser Gly Glu Gly Glu Gly Ser Glu Gly Ser Glu Gly Ser Glu
195       200       205
Glu Gly Glu Gly Glu Gly Ser Glu Gly Ser Glu Gly Glu Gly Ser Glu
210       215       220
Glu Gly Glu Gly Ser Glu Gly Ser Glu Gly Glu Gly Ser Glu
225       230       235       240
Gly Glu Gly Glu Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Glu Gly
245       250       255
Glu Gly Gly Ser Glu Gly Glu Gly Ser Glu Gly Ser Glu Gly Ser Glu
260       265       270
Gly Ser Glu Gly Ser Glu Gly Glu Gly Ser Glu Gly Ser Glu Gly Ser Glu
275       280       285

<210> SEQ ID NO 478
<211> LENGTH: 288
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide

<400> SEQUENCE: 478
Gly Glu Gly Ser Gly Glu Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Ser
1      5      10      15
Gly Glu Gly Glu Gly Ser Glu Gly Glu Gly Ser Glu Gly Ser Glu
20     25     30
Gly Ser Glu Gly Glu Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Glu Gly
35     40     45
Gly Glu Gly Ser Gly Glu Gly Ser Gly Glu Gly Ser Glu Gly Ser Glu Gly
50     55     60
Glu Gly Glu Gly Ser Gly Glu Gly Ser Gly Glu Gly Ser Glu Gly Ser
65     70     75     80
Glu Gly Glu Gly Ser Gly Glu Gly Ser Gly Glu Gly Ser Glu Gly Glu Gly
85     90     95
Ser Gly Glu Gly Ser Gly Ser Gly Ser Gly Ser Gly Gly Glu Gly Ser
100    105    110
Gly Glu Ser Gly Glu Gly Ser Glu Gly Ser Gly Ser Gly Glu Gly
115    120    125
Gly Ser Glu Gly Ser Gly Glu Gly Glu Gly Ser Gly Ser Glu Gly Ser Glu
130 135 140
Gly Glu Gly Ser Gly Glu Gly Ser Gly Glu Gly Ser Gly Ser Glu
145 150 155 160
Gly Glu Gly Glu Gly Ser Gly Glu Gly Ser Gly Gly Ser Gly Ser Glu
165 170 175
Glu Gly Ser Glu Gly Gly Ser Gly Ser Gly Ser Glu Gly Gly Glu
180 185 190
Gly Ser Glu Gly Ser Gly Glu Gly Ser Gly Ser Gly Ser Glu Gly
195 200 205
Glu Gly Ser Glu Gly Gly Ser Gly Ser Gly Ser Gly Ser Glu Gly
210 215 220
Glu Gly Glu Gly Ser Gly Gly Ser Glu Gly Ser Gly Ser Gly Gly Ser
225 230 235 240
Gly Ser Glu Gly Gly Ser Gly Ser Gly Ser Glu Gly Gly Glu Gly
245 250 255
Glu Gly Ser Glu Gly Gly Ser Gly Ser Gly Ser Glu Gly Ser Glu
260 265 270
Gly Ser Glu Gly Ser Glu Gly Ser Gly Ser Gly Ser Glu Gly Ser Glu
275 280 285

<210> SEQ ID NO 479
<211> LENGTH: 272
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide
<400> SEQUENCE: 479
Ser Glu Gly Glu Ser Glu Ser Glu Ser Glu Ser Glu Ser Gly Glu Ser Ser 1  5 10 15
Ser Gly Gly Ser Glu Ser Gly Ser Glu Gly Ser Gly Gly Ser Gly Glu 20 25 30
Ser Glu Gly Glu Gly Ser Gly Ser Gly Ser Gly Ser Gly Ser Gly Glu 35 40 45
Ser Gly Glu Gly Ser Glu Gly Ser Gly Ser Glu Gly Ser Glu Ser Glu 50 55 60
Ser Glu Gly Glu Ser Glu Ser Glu Ser Glu Ser Gly Glu Ser Ser Glu 65 70 75 80
Ser Gly Gly Ser Glu Ser Glu Ser Gly Ser Gly Ser Gly Ser Gly Glu 95 100 95
Ser Gly Ser Gly Ser Gly Ser Gly Ser Gly Ser Glu Gly Ser Glu 105 110
Ser Glu Gly Glu Ser Glu Ser Gly Ser Glu Gly Ser Glu Ser Glu 115 120 125
Ser Glu Gly Glu Ser Glu Ser Glu Ser Gly Ser Glu Ser Glu 130 135 140
Ser Gly Gly Ser Glu Ser Glu Ser Glu Gly Ser Gly Ser Gly Glu 145 150 155 160
Ser Glu Gly Ser Gly Glu Gly Ser Gly Ser Gly Ser Glu Gly Ser Glu 165 170 175
Ser Gly Gly Ser Glu Ser Glu Ser Gly Ser Glu Gly Ser Glu Gly 180 185 190
Ser Gly Ser Glu Ser Glu Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Ser 210 215 220
Ser Gly Gly Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Ser 225 230 235 240
Ser Glu Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Ser 245 250 255

<210> SEQ ID NO: 480
<211> LENGTH: 264
<212> TYPE: FRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide

<400> SEQUENCE: 480
Ser Glu Gly Glu Ser Glu Ser Glu Ser Glu Ser Gly Gly Gly Ser Gly Ser 1 5 10 15
Ser Gly Gly Ser Glu Ser Glu Gly Ser Glu Gly Ser Gly Gly Gly 20 25 30
Ser Glu Gly Ser Glu Ser Glu Ser Glu Ser Gly Gly Ser Gly Gly Ser 35 40 45
Ser Gly Gly Ser Glu Ser Glu Ser Glu Ser Glu Ser Glu Gly Ser Gly Gly 50 55 60
Ser Gly Ser Gly Ser Gly Ser Gly Ser Gly Ser Glu Gly Gly Gly 65 70 75 80
Ser Glu Gly Glu Ser Glu Ser Glu Ser Glu Ser Glu Ser Glu Ser Gly Gly 85 90 95
Ser Glu Ser Gly Gly Gly Ser Glu Ser Glu Ser Gly Gly Ser Gly Gly 100 105 110
Ser Gly Gly Gly Gly Ser Gly Ser Gly Ser Gly Ser Glu Gly Ser Ser 115 120 125
Ser Gly Ser Gly Ser Gly Ser Gly Ser Gly Ser Glu Gly Gly Gly 130 135 140
Ser Glu Gly Glu Ser Glu Ser Glu Ser Glu Ser Gly Gly Gly Ser Gly Gly 145 150 155 160
Ser Gly Glu Ser Gly Ser Gly Ser Gly Ser Gly Ser Glu Ser Gly Gly 165 170 175
Ser Glu Gly Glu Ser Glu Ser Glu Ser Gly Ser Gly Glu Gly Ser Gly Gly 180 185 190
Ser Gly Ser Gly Ser Gly Ser Gly Ser Gly Ser Gly Ser Gly Gly Ser Gly 195 200 205
Ser Glu Gly Ser Glu Ser Glu Gly Ser Glu Gly Ser Glu Gly Ser Gly Gly 210 215 220
Ser Ser Glu Gly Ser Gly Ser Glu Ser Gly Ser Glu Gly Ser Gly Gly Ser 225 230 235 240
Ser Gly Gly Ser Gly Ser Glu Ser Gly Ser Gly Ser Gly Ser Gly Gly Gly 245 250 255
Ser Gly Gly Ser Glu Gly Ser Gly Ser Gly
<210> SEQ ID NO 481
<211> LENGTH: 264
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide

<400> SEQUENCE: 481
Ser Glu Gly Glu Ser Glu Ser Ser Glu Gly Glu Ser Ser Ser Ser
1      5      10     15
Ser Gly Gly Ser Glu Ser Ser Glu Gly Ser Gly Gly Gly Gly
20     25     30
Ser Glu Gly Gly Glu Gly Gly Ser Ser Gly Ser Gly Gly Gly Gly
35     40     45
Ser Gly Gly Ser Glu Gly Ser Ser Glu Gly Ser Ser Glu Gly Ser Ser
50     55     60
Ser Glu Gly Ser Ser Ser Ser Ser Glu Gly Ser Ser Gly Ser Ser Ser
65     70     75     80
Ser Gly Gly Ser Glu Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser
85     90     95
Ser Gly Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser
100    105    110
Ser Glu Gly Ser Glu Ser Ser Gly Gly Gly Gly Ser Ser Glu Gly
115    120    125
Ser Gly Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser
130    135    140
Ser Glu Gly Ser Glu Ser Ser Gly Gly Gly Gly Ser Ser Glu Gly
145    150    155    160
Ser Ser Gly Gly Gly GLu Ser Ser Ser Ser Ser Ser Ser Ser Ser
165    170    175
Ser Gly Gly Gly Gly Gly Ser Ser Ser Ser Ser Ser Ser Ser Ser
180    185    190
Ser Glu Gly Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser
195    200    205
Ser Gly Gly Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser
210    215    220
Ser Ser Gly Gly Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser
225    230    235    240
Ser Gly Gly Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser
245    250    255
Ser Gly Gly Ser Glu Gly Ser Ser Ser Ser Ser Ser Ser Ser Ser
260
<210> SEQ ID NO 483
<211> LENGTH: 28
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide
<400> SEQUENCE: 483

Ser Gly Gly Ser Gly Glu Ser Gly Glu Glu Gly Gly Ser Gly Gly Gly
1 5 10 15
Ser Gly Glu Ser Gly Ser Gly Ser Gly Ser Gly Ser Gly Glu Gly Gly
20 25 30
Ser Gly Glu Ser Gly Ser Gly Ser Gly Ser Gly Ser Gly Ser Gly Glu
35 40 45
Ser Gly Glu Ser Gly Glu Ser Gly Glu Ser Gly Gly Ser Gly Gly Ser
50 55 60
Ser Gly Glu Ser Gly Ser Gly Ser Gly Ser Gly Ser Gly Gly Glu Ser
65 70 75 80
Ser Gly Glu Gly Ser Gly Ser Gly Ser Gly Ser Gly Gly Ser Gly Glu
85 90 95
Ser Gly Glu Gly Ser Gly Ser Gly Ser Gly Ser Gly Ser Gly Ser Gly
100 105 110
Ser Gly Glu Gly Ser Gly Ser Gly Ser Gly Ser Gly Gly Glu Gly Glu
115 120 125
Ser Gly Glu Ser Gly Ser Gly Ser Gly Ser Gly Ser Gly Ser Gly Glu
130 135 140
Ser Gly Glu Ser Gly Ser Gly Ser Gly Ser Gly Ser Gly Ser Gly Ser
145 150 155 160
Ser Gly Ser Gly Ser Gly Ser Gly Ser Gly Ser Gly Ser Gly Ser Gly
165 170 175
Ser Gly Ser Gly Ser Gly Ser Gly Ser Gly Ser Gly Ser Gly Ser Gly
180 185 190
Ser Gly Ser Gly Ser Gly Ser Gly Ser Gly Ser Gly Ser Gly Ser Gly
195 200 205
Ser Gly Ser Gly Ser Gly Ser Gly Ser Gly Ser Gly Ser Gly Ser Gly
210 215 220
Ser Gly Glu Gly Ser Gly Gly Glu Gly Glu Gly Gly Ser Gly Gly Ser
225 230 235 240
Ser Gly Glu Gly Ser Gly Gly Glu Gly Glu Gly Gly Glu Gly Ser Ser
245 250 255
Ser Gly Glu Gly Ser Gly Ser Gly
260

<210> SEQ ID NO 484
<211> LENGTH: 664
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide
<220> FEATURE:
tct agt gag tcc agt gaa tcc agc gaa tct agt gag gaa tcc
Ser Ser Glu Ser Ser Glu Ser Ser Glu Ser Ser Glu Ser
1 5 10 15

agc gag tct agc tct gag gaa tct tct agt gag tcc agt gag tcc agt
Ser Glu Ser Ser Ser Glu Ser Ser Glu Ser Ser Glu Ser Glu Ser
20 25 30

tcc agt gaa tct tct agt gag tcc agt gaa tct tct gag tcc agt gaa tct
Ser Ser Glu Ser Ser Glu Ser Ser Glu Ser Ser Glu Ser Ser Glu Ser
35 40 45

tct agt gag tct agc gag taa tcc agc gaa tct tct ago tcc gag gag tcc tct gag tcc agt gag gaa tcc
Ser Ser Glu Ser Ser Glu Ser Ser Glu Ser Ser Glu Ser Ser Glu Ser
50 55 60

agc gaa tcc agc tct ago tcc gag taa tcc agc gag tcc agc gag tcc agc
Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Glu Ser
65 70 75 80

tcc agt gaa tct tct agt gag tcc agt gag tcc agt tct gag tcc agt gag tcc agt
Ser Ser Glu Ser Ser Glu Ser Ser Glu Ser Ser Glu Ser Ser Glu Ser
85 90 95

tct agt gag tcc tct gag tcc agc gag gag tcc tct gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gag gaz
<table>
<thead>
<tr>
<th>Ser</th>
<th>Ser</th>
<th>Glu</th>
<th>Ser</th>
<th>Ser</th>
<th>Glu</th>
<th>Ser</th>
<th>Ser</th>
<th>Glu</th>
<th>Ser</th>
<th>Ser</th>
<th>Glu</th>
<th>Ser</th>
<th>Ser</th>
<th>Glu</th>
<th>Ser</th>
<th>Ser</th>
<th>Glu</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td>10</td>
<td>15</td>
<td>20</td>
<td>25</td>
<td>30</td>
<td>35</td>
<td>40</td>
<td>45</td>
<td>50</td>
<td>55</td>
<td>60</td>
<td>65</td>
<td>70</td>
<td>75</td>
<td>80</td>
<td>85</td>
</tr>
<tr>
<td>Gly</td>
<td>Glu</td>
<td>Gly</td>
<td>Glu</td>
<td>Gly</td>
<td>Glu</td>
<td>Gly</td>
<td>Glu</td>
<td>Gly</td>
<td>Glu</td>
<td>Gly</td>
<td>Glu</td>
<td>Gly</td>
<td>Glu</td>
<td>Gly</td>
<td>Glu</td>
<td>Gly</td>
<td>Glu</td>
</tr>
<tr>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>25</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**SEQ ID NO 487**
**LENGTH:** 26
**TYPE:** PRT
**ORGANISM:** Artificial Sequence
**FEATURE:**
**OTHER INFORMATION:** Description of Artificial Sequence: Synthetic peptide

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>25</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**SEQ ID NO 488**
**LENGTH:** 27
**TYPE:** PRT
**ORGANISM:** Artificial Sequence
**FEATURE:**
**OTHER INFORMATION:** Description of Artificial Sequence: Synthetic peptide

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>25</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**SEQ ID NO 489**
**LENGTH:** 29
**TYPE:** PRT
**ORGANISM:** Artificial Sequence
**FEATURE:**
**OTHER INFORMATION:** Description of Artificial Sequence: Synthetic peptide

<table>
<thead>
<tr>
<th>Ser</th>
<th>Glu</th>
<th>Ser</th>
<th>Ser</th>
<th>Gly</th>
<th>Ser</th>
<th>Glu</th>
<th>Ser</th>
<th>Ser</th>
<th>Ser</th>
<th>Gly</th>
<th>Ser</th>
<th>Ser</th>
<th>Ser</th>
<th>Glu</th>
<th>Ser</th>
<th>Ser</th>
<th>Ser</th>
<th>Glu</th>
<th>Ser</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>28</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**SEQ ID NO 490**
**LENGTH:** 13
**TYPE:** PRT
**ORGANISM:** Artificial Sequence
**FEATURE:**
**OTHER INFORMATION:** Description of Artificial Sequence: Synthetic peptide

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**SEQ ID NO 491**
**LENGTH:** 24
**TYPE:** PRT
<210> SEQ ID NO 491
<211> LENGTH: 24
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

1  5  10  15
Gly Ser Gly Gly Ser Gly Gly Glu
20

<210> SEQ ID NO 492
<211> LENGTH: 24
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

1  5  10  15
Gly Gly Gly Ser Gly Gly Gly Glu
20

<210> SEQ ID NO 493
<211> LENGTH: 24
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

1  5  10  15
Gly Gly Gly Gly Gly Gly Glu
20

<210> SEQ ID NO 494
<211> LENGTH: 18
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

1  5  10  15
Gly Glu

<210> SEQ ID NO 495
<211> LENGTH: 13
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

Glu Glu Glu Glu Glu Glu Glu Glu Glu Glu Glu Glu Glu Glu Glu Glu Glu
1  5  10
<210> SEQ ID NO: 496
<211> LENGTH: 22
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 496

Ser Ser Ser Ser Ser Glu Ser Ser Ser Ser Ser Glu Ser Ser Ser Ser 1 5 10 15
Ser Glu Ser Ser Ser Ser Ser
20

<210> SEQ ID NO: 497
<211> LENGTH: 24
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 497

Ser Ser Ser Glu Ser Ser Ser Glu Ser Ser Ser Ser Glu Ser Ser Glu Ser Ser Glu Ser Ser Glu 1 5 10 15
Ser Ser Glu Ser Ser Ser Glu
20

<210> SEQ ID NO: 498
<211> LENGTH: 18
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 498

Ser Ser Ser Ser Ser Ser Ser Glu Ser Ser Ser Ser Ser Glu Ser Ser Ser Ser 1 5 10 15
Ser Glu

<210> SEQ ID NO: 499
<211> LENGTH: 864
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide
<221> NAME/KEY: CDS
<222> LOCATION: (1) .. (864)

<400> SEQUENCE: 499


48
96
144
<210> SEQ ID NO 500
<211> LENGTH: 280
<212> TYPE: PAT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide

<400> SEQUENCE: 500
1  5  10  15

35    40    45
50    55    60
65    70    75    80
85    90    95
100   105   110
115   120   125
130   135   140
145   150   155   160
165   170   175
180   185   190
195   200   205
210   215   220
225   230   235   240
245   250   255
260   265   270
275   280   285

<210> SEQ ID NO 501
<211> LENGTH: 126
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
polynucleotide
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (1) . . (39)
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (58) . . (126)

<400> SEQUENCE: 501
atg gat tat aaa gag gat gac gat aaa ggg tct cca ggt tagttacctct
49
Met Asp Tyr Lys Asp Asp Asp Asp Lys Gly Ser Pro Gly
1     5     10

ggtgata gga ggt tcg tct tca ctc gag ggt acc cat cac cat cac cat
99
Gly Gly Ser Ser Ser Ser Leu Gly Thr His His His His
15    20    25

cac gag ctc gta cgc gta gaa aaa atg
126
<210> SEQ ID NO 502
<211> LENGTH: 13
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 502

Met Asp Tyr Lys Asp Asp Asp Lys Gly Ser Pro Gly
1 5 10

<210> SEQ ID NO 503
<211> LENGTH: 23
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 503

Gly Gly Ser Ser Ser Leu Glu Gly Thr His His His His His Glu
1 5 10 15
Leu Val Pro Val Glu Lys Met
20

<210> SEQ ID NO 504
<211> LENGTH: 6
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 504

Lys Lys Lys Lys Lys Lys
1 5

<210> SEQ ID NO 505
<211> LENGTH: 856
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide

<400> SEQUENCE: 505

Met Glu Gly Asp Ile Gln Met Thr Gln Ser Pro Ser Ser Leu Ser Ser Ala
1 5 10 15
Ser Val Gly Asp Arg Val Thr Ile Thr Cys Arg Ala Ser Gln Asp Val
20 25
Asn Thr Ala Val Ala Trp Tyr Gln Gln Lys Pro Gly Lys Ala Pro Lys
30 35 40
Leu Leu Ile Tyr Ser Ala Ser Phe Leu Tyr Ser Gly Val Pro Ser Arg
45 50 55 60
Phe Ser Gly Ser Arg Ser Gly Thr Asp Phe Thr Leu Thr Ile Ser Ser
65 70 75 80
Leu Gln Pro Glu Asp Phe Ala Thr Tyr Cys Gln Gln His Tyr Thr
95 90 95
Thr Pro Pro Thr Phe Gly Glu Gly Thr Lys Val Glu Ile Lys Thr Gly
100 105 110
Ser Gly Glu Gly Ser Glu Gly Gly Gly Gly Glu Gly Ser Glu Gly
115 120 125
Glu Gly Ser Gly Glu Gly Gly Glu Gly Ser Gly Thr Glu Val
130 135 140
Gln Leu Val Glu Ser Gly Gly Gly Leu Val Gin Pro Gly Gly Ser Leu
145 150 155 160
Arg Leu Ser Cys Ala Ala Ser Gly Phe Asn Ile Lys Asp Thr Tyr Ile
165 170 175
His Trp Val Arg Gin Ala Pro Gly Lys Gly Leu Glu Trp Val Ala Arg
180 185 190
Ile Tyr Pro Thr Asn Gly Tyr Thr Arg Tyr Ala Asp Ser Val Lys Gly
195 200 205
Arg Phe Thr Ile Ser Ala Asp Thr Ser Lys Asn Thr Ala Tyr Leu Gln
210 215 220
Met Asn Ser Leu Arg Ala Glu Asp Thr Ala Val Tyr Cys Ser Arg
225 230 235 240
Trp Gly Gly Asp Gly Phe Tyr Ala Met Asp Tyr Trp Gly Gin Gly Thr
245 250 255
Leu Val Thr Val Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Glu Gly
260 265 270
Glu Gly Ser Glu Gly Ser Gly Gly Gly Gly Ser Glu Gly Ser Glu Ser Gly
275 280 285
Glu Gly Gly Gly Ser Glu Gly Ser Glu Gly Gly Ser Glu Gly Ser Glu Gly
290 295 300
Ser Gly Glu Gly Glu Gly Gly Ser Glu Gly Gly Ser Glu Gly Gly Ser Glu
305 310 315 320
Gly Glu Gly Ser Glu Gly Glu Gly Gly Gly Glu Gly Ser Glu Gly Glu
325 330 335
Gly Ser Gly Glu Gly Glu Gly Gly Glu Gly Ser Glu Gly Ser Glu Ser Glu
340 345 350
Gly Glu Gly Ser Glu Gly Gly Glu Gly Ser Glu Gly Ser Glu Gly Ser Glu
355 360 365
Gly Glu Gly Ser Glu Gly Gly Ser Glu Gly Glu Gly Ser Glu Gly Ser Glu
370 375 380
Gly Ser Glu Gly Ser Gly Gly Ser Gly Gly Glu Gly Gly Gly
385 390 395 400
Ser Glu Gly Ser Gly Glu Gly Gly Ser Gly Ser Glu Gly Ser Glu Gly
405 410 415
Glu Gly Ser Glu Gly Ser Gly Gly Glu Gly Ser Glu Gly Ser Glu Gly Ser Glu
420 425 430
Glu Gly Glu Gly Ser Glu Gly Ser Glu Gly Glu Gly Ser Gly Glu
435 440 445
Gly Ser Glu Gly Gly Ser Gly Ser Gly Ser Gly Ser Gly Glu Gly
450 455 460
Gly Glu Gly Ser Glu Gly Gly Ser Gly Ser Gly Ser Glu Gly
465 470 475 480
Gly Ser Glu Gly Ser Glu Gly Ser Gly Gly Ser Glu Gly Ser Glu Gly Ser Glu
485 490 495
polypeptide

<400> SEQUENCE: 506

Met Glu Asp Ile Leu Leu Thr Gln Ser Pro Val Ile Leu Ser Val Ser
1 5 10 15
Pro Gly Glu Arg Val Ser Phe Ser Cys Arg Ala Ser Gln Ser Ile Gly
20 25 30
Thr Asn Ile His Trp Tyr Gln Gln Arg Thr Asn Gly Ser Pro Arg Leu
35 40 45
Leu Ile Lys Tyr Ala Ser Glu Ser Ile Ser Gly Ile Pro Ser Arg Phe
50 55 60
Ser Gly Ser Gly Ser Gly Thr Asp Phe Thr Leu Ser Ile Asn Ser Val
65 70 75 80
Glu Ser Glu Asp Ile Ala Asp Tyr Tyr Cys Gln Gln Asn Asn Thr
85 90 95
Pro Thr Thr Phe Gly Ala Gly Thr Lys Leu Glu Leu Lys Thr Gly Ser
100 105 110
Gly Glu Gly Ser Glu Gly Gly Gly Gly Gly Glu Gly Ser Glu Gly Glu
115 120 125
Gly Ser Gly Glu Gly Gly Gly Gly Gly Gly Gly Gly Thr Gln Val Gln
130 135 140
Leu Lys Gln Ser Gly Pro Gly Leu Val Gln Pro Ser Gln Ser Leu Ser
145 150 155 160
Ile Thr Cys Thr Val Ser Gly Phe Ser Leu Thr Asn Tyr Gly Val His
165 170 175
Trp Val Arg Gln Ser Pro Gly Lys Gly Leu Glu Trp Leu Gly Val Ile
180 185 190
Trp Ser Gly Gly Asn Thr Asp Tyr Asn Thr Phe Thr Ser Arg Leu
195 200 205
Ser Ile Asn Lys Asp Asn Ser Lys Ser Gln Val Phe Phe Lys Met Asn
210 215 220
Ser Leu Gln Ser Asp Thr Ala Ile Tyr Tyr Cys Ala Arg Ala Leu
225 230 235 240
Thr Tyr Tyr Asp Tyr Glu Phe Ala Tyr Trp Gly Gln Gly Thr Leu Val
245 250 255
Thr Val Ser Gly Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Glu
260 265 270
Ser Glu Gly Ser Gly Glu Gly Ser Gly Ser Gly Ser Gly Glu Gly Glu
275 280 285
Glu Gly Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Ser Gly Gly Glu
290 295 300
Glu Gly Glu Gly Gly Gly Ser Gly Gly Glu Gly Gly Ser Gly Gly Ser
305 310 315 320
Gly Ser Gly Glu Gly Gly Gly Gly Ser Gly Ser Gly Ser Gly Gly Ser
325 330 335
Gly Glu Gly Gly Glu Gly Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly
340 345 350
Gly Ser Glu Gly Gly Gly Ser Gly Gly Ser Gly Ser Gly Gly Ser
355 360 365
Gly Gly Gly Ser Gly Ser Glu Gly Gly Ser Gly Gly Ser Gly Gly
370 375 380
<table>
<thead>
<tr>
<th>Ser Glu Gly Glu Gly Ser Glu Gly Ser Gly Glu Gly Glu Gly Ser Glu</th>
<th>385</th>
<th>390</th>
<th>395</th>
<th>400</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gly Ser Gly Glu Gly Glu Gly Ser Glu Gly Ser Gly Glu Gly Glu</td>
<td>405</td>
<td>410</td>
<td>415</td>
<td></td>
</tr>
<tr>
<td>Gly Ser Gly Glu Gly Glu Gly Ser Glu Gly Ser Gly Glu Gly Glu</td>
<td>420</td>
<td>425</td>
<td>430</td>
<td></td>
</tr>
<tr>
<td>Gly Ser Gly Glu Gly Ser Gly Glu Gly Ser Gly Glu Gly Ser Gly</td>
<td>435</td>
<td>440</td>
<td>445</td>
<td></td>
</tr>
<tr>
<td>Gly Ser Gly Glu Gly Ser Gly Glu Gly Ser Gly Glu Gly Ser Gly</td>
<td>450</td>
<td>455</td>
<td>460</td>
<td></td>
</tr>
<tr>
<td>Gly Ser Gly Glu Gly Ser Gly Glu Gly Ser Gly Glu Gly Ser Gly</td>
<td>465</td>
<td>470</td>
<td>475</td>
<td>480</td>
</tr>
<tr>
<td>Gly Ser Gly Glu Gly Ser Gly Glu Gly Ser Gly Glu Gly Ser Gly</td>
<td>485</td>
<td>490</td>
<td>495</td>
<td></td>
</tr>
<tr>
<td>Gly Gly Glu Gly Ser Gly Glu Gly Ser Gly Glu Gly Ser Gly Glu</td>
<td>500</td>
<td>505</td>
<td>510</td>
<td></td>
</tr>
<tr>
<td>Gly Ser Gly Glu Gly Ser Gly Glu Gly Ser Gly Glu Gly Ser Gly</td>
<td>515</td>
<td>520</td>
<td>525</td>
<td></td>
</tr>
<tr>
<td>Gly Glu Gly Ser Gly Glu Gly Ser Gly Glu Gly Ser Gly Glu Gly</td>
<td>530</td>
<td>535</td>
<td>540</td>
<td></td>
</tr>
<tr>
<td>Gly Glu Gly Ser Gly Glu Gly Ser Gly Glu Gly Ser Gly Glu Gly</td>
<td>545</td>
<td>550</td>
<td>555</td>
<td>560</td>
</tr>
<tr>
<td>Ser Gly Glu Gly Ser Gly Glu Gly Ser Gly Glu Gly Ser Gly Glu</td>
<td>565</td>
<td>570</td>
<td>575</td>
<td></td>
</tr>
<tr>
<td>Ser Gly Glu Gly Ser Gly Glu Gly Ser Gly Glu Gly Ser Gly Glu</td>
<td>580</td>
<td>585</td>
<td>590</td>
<td></td>
</tr>
<tr>
<td>Gly Gly Glu Gly Ser Gly Glu Gly Ser Gly Glu Gly Ser Gly Glu</td>
<td>595</td>
<td>600</td>
<td>605</td>
<td></td>
</tr>
<tr>
<td>Gly Gly Glu Gly Ser Gly Glu Gly Ser Gly Glu Gly Ser Gly Glu</td>
<td>610</td>
<td>615</td>
<td>620</td>
<td></td>
</tr>
<tr>
<td>Gly Glu Gly Ser Gly Glu Gly Ser Gly Glu Gly Ser Gly Glu Gly</td>
<td>625</td>
<td>630</td>
<td>635</td>
<td>640</td>
</tr>
<tr>
<td>Gly Ser Gly Glu Gly Ser Gly Glu Gly Ser Gly Glu Gly Ser Gly</td>
<td>645</td>
<td>650</td>
<td>655</td>
<td></td>
</tr>
<tr>
<td>Gly Gly Glu Gly Ser Gly Glu Gly Ser Gly Glu Gly Ser Gly Glu</td>
<td>660</td>
<td>665</td>
<td>670</td>
<td></td>
</tr>
<tr>
<td>Gly Gly Glu Gly Ser Gly Glu Gly Ser Gly Glu Gly Ser Gly Glu</td>
<td>675</td>
<td>680</td>
<td>685</td>
<td></td>
</tr>
<tr>
<td>Gly Gly Glu Gly Ser Gly Glu Gly Ser Gly Glu Gly Ser Gly Glu</td>
<td>690</td>
<td>695</td>
<td>700</td>
<td></td>
</tr>
<tr>
<td>Ser Gly Glu Gly Ser Gly Glu Gly Ser Gly Glu Gly Ser Gly Glu</td>
<td>705</td>
<td>710</td>
<td>715</td>
<td>720</td>
</tr>
<tr>
<td>Gly Gly Ser Gly Glu Gly Ser Gly Glu Gly Ser Gly Glu Gly Ser</td>
<td>725</td>
<td>730</td>
<td>735</td>
<td></td>
</tr>
<tr>
<td>Gly Gly Glu Gly Ser Gly Glu Gly Ser Gly Glu Gly Ser Gly Glu</td>
<td>740</td>
<td>745</td>
<td>750</td>
<td></td>
</tr>
<tr>
<td>Ser Gly Glu Gly Ser Gly Glu Gly Ser Gly Glu Gly Ser Gly Glu</td>
<td>755</td>
<td>760</td>
<td>765</td>
<td></td>
</tr>
<tr>
<td>Gly Gly Glu Gly Ser Gly Glu Gly Ser Gly Glu Gly Ser Gly Glu</td>
<td>770</td>
<td>775</td>
<td>780</td>
<td></td>
</tr>
<tr>
<td>Gly Ser Gly Glu Gly Ser Gly Glu Gly Ser Gly Glu Gly Ser Gly</td>
<td>785</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Gly Glu Gly Gly Glu Gly Ser Gly Gly Glu Gly Gly Glu Gly Glu
805
Ser Glu Gly Gly Ser Gly Ser Gly Glu Gly Gly Ser Gly
820
Glu Gly Ser Gly Asp Tyr Lys Asp Asp Asp Lys Gly Gly Ser
835
His His His His His His
850
<210> SEQ ID NO: 507
<211> LENGTH: 852
<212> TYPE: PRT
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide
<400> SEQUENCE: 507
Met Glu Gly Asp Ile His Met Glu Asp Ile Gln Met Thr Gln Ser Pro 1 5 10 15
Ser Ser Leu Ser Ala Ser Val Gly Asp Arg Val Thr Ile Thr Cys Arg 20 25 30
Ala Ser Gln Amp Val Asn Thr Ala Ala Trp Tyr Gln Gln Lys Pro 35 40 45
Gly Lys Ala Pro Lys Leu Leu Ile Tyr Ser Ala Ser Phe Leu Tyr Ser 50 55 60
Gly Val Pro Ser Arg Phe Ser Gly Ser Arg Ser Gly Thr Asp Phe Thr 65 70 75 80
Leu Thr Ile Ser Ser Leu Gln Pro Glu Asp Phe Ala Thr Tyr Cys 85 90 95
Gln Gln His Tyr Thr Thr Pro Pro Thr Phe Gly Gln Gln Gly Thr Lys Val 100 105 110
Glu Ile Lys Ser Gly Glu Val Gln Leu Val Glu Ser Gly Gly Gly 115 120 125
Leu Val Gln Pro Gly Gly Ser Leu Arg Leu Ser Cys Ala Ala Ser Gly 130 135 140
Phe Asn Ile Lys Asp Thr Tyr Ile His Trp Val Arg Gln Ala Pro Gly 145 150 155 160
Lys Gly Leu Glu Trp Val Ala Arg Ile Tyr Pro Thr Asn Gly Tyr Thr 165 170 175
Arg Tyr Ala Asp Ser Val Lys Gly Arg Phe Thr Ile Ser Ala Asp Thr 180 185 190
Ser Lys Asn Thr Ala Tyr Leu Gln Met Asn Ser Leu Arg Ala Glu Asp 195 200 205
Thr Ala Val Tyr Tyr Cys Ser Arg Trp Gly Gly Asp Gly Phe Tyr Ala 210 215 220
Met Asp Tyr Trp Gly Gln Gly Thr Leu Val Thr Val Ser Gly Gly Glu 225 230 235 240
Gly Ser Gly Glu Gly Ser Gly Glu Gly Ser Gly Ser Gly Ser Gly Glu 245 250 255
<210> SEQ ID NO 508
<211> LENGTH: 684
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEE: CDS
<222> LOCATION: (1)...(684)

<400> SEQUENCE: 508

atg gat aac act cat act tgc cct cct tgt cca gcg ccc gaa ctt ctc
Met Asp Lys Thr His Thr Cys Pro Pro Cys Pro Ala Pro Glu Leu Leu
1    5    10   15

96

ggt ggc ccg ttc gtt ttc ctc gtc cca ccc ccc aaa gac acc ctc
Gly Gly Pro Ser Val Phe Leu Phe Pro Pro Lys Pro Lys Asp Thr Leu
20   25   30

144

atg att tcc cgt act ctt gac gta acc tgg gta gtt gta gtc ctt
tct Met Ile Ser Arg Thr Pro Glu Val Thr Cys Val Val Val Asp Val Ser
35   40   45

192

cac gaa gat ccc gaa gtt aaa ttc aac tgg tac gta gat ggt gtt gac
His Glu Asp Pro Glu Val Lys Phe Asn Trp Tyr Val Asp Gly Val Glu
50   55   60

240

gtg cat aac gtt aac ccc cgc ggc ggg gag caa tat aat tcc acc
Val His Asn Ala Lys Thr Lys Pro Arg Glu Glu Gln Tyr Asn Ser Thr
65   70   75   80

288

tac cgt gtt gtc ttt gtt ctc acc cgc ctt cac gat gtt gct tgc
tgc Tyr Arg Val Val Ser Val Leu Thr Val Leu His Glu Asp Trp Leu Asn
85   90   95

336

ggc aac gaa taa aac tgg tgg aac aac ggc cct cca ggc ccc
Gly Lys Glu Tyr Lys Cys Lys Val Ser Asn Lys Ala Leu Pro Ala Pro
100 105 110

384

atc gaa aac att tct aag ggc aac ggc cag ccc gaa cca caa
Ile Glu Lys Thr Ile Ser Lys Ala Lys Gly Glu Pro Arg Glu Pro Gln
115 120 125

432

gta tat acc ctt cgg ccc tcc cgt gat gaa ctt acc aag aac gaa gtt

-continued
Val Tyr Thr Leu Pro Pro Ser Arg Asp Glu Leu Thr Lys Asn Gln Val
130 135 140

tcc ctg acc tgc ctg gtt aag ggt ttc tcc cca tct gat atc gcc gtc
400
Ser Leu Thr Cys Leu Val Lys Gly Phe Tyr Pro Ser Asp Ile Ala Val
145 150 155 160

gag tgg gaa tcc aac ggt cag ccg gag aac ast tat aaa act atc cca
528
Glu Trp Glu Ser Arg Gly Gin Pro Glu Asn Arg Tyr Lys Thr Ile Pro
165 170 175

ccg gtt ctg gac tcc ggc tcc ttc ctg tat tcc aag ctg acc
576
Pro Val Leu Asp Ser Arg Gly Ser Phe Leu Tyr Ser Lys Leu Thr
180 185 190

gtt gat aag gaa cct tgt cag ggc acc gtt ttc tct tgc tct gta
624
Val Asp Lys Ser Arg Trp Gin Gin Gly Asn Val Phe Ser Cys Ser Val
195 200 205

atg cag gaa gca ctg cac aac cat acc cag aaa agg ctc ctc ctg
672
Met His Glu Ala Leu His Asn His Tyr Thr Gin Lys Ser Leu Ser Leu
210 215 220

tcg ccg gtt aag
684
Ser Pro Gly Lys
225

<210> SEQ ID NO 509
<211> LENGTH: 228
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide

<400> SEQUENCE: 509

Met Asp Lys Thr His Thr Cys Pro Pro Cys Pro Ala Pro Glu Leu Leu
1 5 10 15
Gly Gly Pro Ser Val Phe Leu Phe Pro Pro Lys Pro Lys Asp Thr Leu
20 25 30
Met Ile Ser Arg Thr Pro Glu Val Thr Cys Val Val Val Asp Val Ser
35 40 45
His Glu Asp Pro Glu Val Lys Phe Asn Trp Tyr Val Asp Gly Val Glu
50 55 60
Val His Asn Ala Lys Thr Lys Pro Arg Glu Gin Tyr Gin Asn Ser Thr
65 70 75 80
Tyr Arg Val Val Ser Val Leu Thr Val Leu His Gin Asp Trp Leu Asn
95 100 105 110
Gly Lys Glu Tyr Lys Cys Lys Val Ser Asn Lys Ala Leu Pro Ala Pro
125
Ile Glu Lys Thr Ile Ser Asn Ala Lys Gly Gin Pro Arg Glu Pro Gin
145 150 155 160
Val Tyr Thr Leu Pro Pro Ser Arg Asp Glu Leu Thr Lys Asn Gln Val
180 185 190 195
Ser Leu Thr Cys Leu Val Lys Gly Phe Tyr Pro Ser Asp Ile Ala Val
200 205 210
Glu Trp Glu Ser Arg Gly Gin Pro Glu Asn Gin Tyr Thr Ile Pro
225
Pro Val Leu Asp Ser Arg Gin Ser Phe Leu Tyr Ser Lys Leu Thr
240 245
Val Asp Lys Ser Arg Trp Gin Gin Gly Asn Val Phe Ser Cys Ser Val
255
Met His Glu Ala Leu His Asn His Tyr Thr Gln Lys Ser Leu Ser Leu Ser Pro Gly Lys
210 215 220 225

| SEQ ID NO: 510 |
| LENGTH: 510 |
| TYPE: PRT |
| ORGANISM: Artificial Sequence |
| FEATURE: OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide |

| SEQUENCE: 510 |
| Met Ser Lys Gly Glu Glu Leu Phe Thr Gln Val Val Pro Ile Leu Val |
| Glu Leu Asp Gly Asp Val Asn Gly His Lys Phe Ser Val Ser Gly Glu |
| Gly Glu Gly Asp Ala Thr Tyr Gly Lys Leu Thr Leu Lys Phe Ile Cys |
| Thr Thr Gly Lys Leu Pro Val Pro Trp Pro Thr Leu Val Thr Thr Phe |
| Ser Tyr Gly Val Gln Cys Phe Ser Arg Tyr Pro Asp His Met Lys Arg |
| His Asp Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr Val Gln Glu Arg |
| Thr Ile Ser Phe Lys Asp Gly Asn Tyr Lys Thr Arg Ala Glu Val |
| Lys Phe Glu Gly Asp Thr Leu Val Asn Arg Ile Glu Leu Lys Gly Ile |
| Asp Phe Lys Gly Asp Gly Asn Ile Leu Gly His Lys Leu Gly Tyr Asn |
| Tyr Asn Ser His Asn Val Tyr Ile Thr Ala Asp Lys Glu Asn Gly |
| Ile Lys Ala Asn Phe Lys Ile Arg His Asn Ile Glu Asp Gly Ser Val |
| Glu Leu Ala Asp His Tyr Gln Glu Asn Thr Pro Ile Gly Asp Gly Pro |
| Val Leu Leu Pro Asp Asn His Tyr Leu Ser Thr Gln Ser Ala Leu Ser |
| Lys Asp Pro Asn Glu Lys Arg Asp His Met Val Leu Leu Glu Phe Val |
| Thr Ala Ala Gly Ile Gly Glu Gly Ser Gly Glu Gly Ser Gly Glu |
| Gly Ser Gly Ser Gly Glu Gly Glu Gly Ser Gly Ser Gly Glu Ser Gly |
| Gly Glu Gly Ser Glu Ser Gly Glu Gly Ser Gly Glu Gly Ser Gly Ser Gly |
| Gly Glu Gly Gly Gly Gly Gly Gly Ser Gly Ser 265 270 |
| Gly Gly Ser Gly Gly Gly Gly Ser Gly Ser 275 280 285 |
| Gly Ser Gly Gly Gly Gly Ser Gly Ser 290 295 300 |
| Ser Gly Glu Gly Gly Gly Gly Ser Gly Gly Ser Gly Ser Glu Gly 320 |
| Gly Gly Ser Gly Gly Gly Gly Ser Gly Ser Gly Ser Gly Glu Gly |
305 Glu Gly Gly Ser Glu Gly Gly Glu Gly Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly
310 315 320
Glu Gly Gly Ser Glu Gly Gly Ser Glu Gly Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly 325
330 335
Gly Ser Glu Gly Gly Ser Glu Gly Gly Ser Glu Gly Gly Ser Glu Gly Ser Glu Gly Ser
340 345 350
Gly Ser Glu Gly Gly Ser Glu Gly Gly Ser Glu Gly Gly Ser Glu Gly Gly Ser Glu Gly 355
360 365
Glu Gly Ser Gly Glu Gly Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Ser
370 375 380
Gly Ser Glu Gly Ser Gly Glu Gly Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Ser Glu 385
390 395 400
Gly Gly Ser Gly Glu Gly Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Ser
405 410 415
Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Ser
420 425 430
Glu Gly Ser Glu Gly Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Ser Glu
435 440 445
Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Ser
450 455 460
Glu Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Ser
465 470 475 480
Glu Gly Gly Ser Gly Ser Glu Gly Glu Gly Ser Gly Ser Glu Gly Ser Glu Gly Ser Glu
485 490 495
Ser Gly Glu Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Ser Glu
500 505 510
Ser Glu Gly Ser Gly Glu Gly Ser Gly Glu Gly Ser Gly Glu Gly Ser Glu Gly Ser Glu
515 520 525
Gly Ser Glu Gly Ser Gly Glu Gly Ser Gly Glu Gly Ser Gly Glu Gly Ser Glu Gly Ser Glu
530 535 540
Gly Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Ser Glu
545 550 555 560
Ser Gly Glu Gly Ser Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly
565 570 575
Gly Ser Gly Glu Gly Ser Gly Ser Glu Gly Ser Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly
580 585 590
Ser Gly Glu Gly Ser Gly Glu Gly Ser Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Ser Glu
595 600 605
Glu Gly Ser Gly Ser Glu Gly Ser Gly Glu Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Ser
610 615 620
Glu Gly Ser Gly Ser Glu Gly Ser Gly Ser Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Glu
625 630 635 640
Ser Gly Glu Gly Ser Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Ser
645 650 655
Gly Glu Gly Ser Glu Gly Ser Gly Ser Glu Gly Ser Gly Ser Glu Gly Ser Glu Gly Ser Glu
660 665 670
Gly Ser Glu Gly Ser Gly Ser Glu Gly Ser Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Ser
675 680 695 700
Gly Gly Ser Glu Gly Ser Gly Ser Glu Gly Ser Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Ser
690 695 700
Ser Gly Glu Gly Ser Gly Ser Glu Gly Ser Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Ser
705 710 715 720
-continued

Gly Ser Gly Glu Gly Gly Ser Gly Ser Gly Glu Gly Gly Ser Gly Gly
725 730 735
Ser Gly Glu Gly Ser Glu Gly Gly Ser Glu Gly Gly Glu Gly Glu Gly
740 745 750
Glu Gly Ser Glu Gly Ser Gly Glu Gly Ser Gly Glu Gly Ser Gly Ser
755 760 765
Glu Gly Glu Gly Gly Ser Gly Glu Gly Gly Glu Gly Gly Gly Glu
770 775 780
Gly Ser Gly Glu Gly Ser Glu Gly Gly Ser Gly Glu Gly Gly Ser
785 790 795 800
Gly Glu Gly Ser Glu Gly
805

<210> SEQ ID NO 511
<211> LENGTH: 797
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide

<400> SEQUENCE: 511

Met Ala Asn Thr Pro Val Ser Gly Asn Leu Lys Val Glu Phe Tyr Asn
1     5   10  15
Ser Asn Pro Ser Asp Thr Thr Asn Ser Ile Asn Pro Gln Phe Lys Val
20   25   30
Thr Asn Thr Gly Ser Ser Ala Ile Asp Leu Ser Lys Leu Thr Leu Arg
35   40   45
Tyr Tyr Thr Thr Val Asp Gly Gln Lys Asp Gln Thr Phe Trp Ala Asp
50   55   60
His Ala Ala Ile Ile Gly Ser Asn Gly Ser Tyr Asn Gly Ile Thr Ser
65   70   75   80
Asn Val Lys Gly Thr Phe Val Val Met Ser Ser Ser Thr Asn Asn Ala
85   90   95
Asp Thr Tyr Leu Glu Ile Ser Phe Thr Gly Gly Thr Leu Glu Pro Gly
100  105  110
Ala His Val Glu Ile Gln Gly Arg Phe Ala Lys Asn Asp Trp Ser Asn
115  120  125
Tyr Thr Gln Ser Asn Asp Tyr Ser Phe Lys Ser Ala Ser Gln Phe Val
130  135  140
Glu Trp Asp Gln Val Thr Ala Tyr Leu Asn Gly Val Leu Val Trp Gly
145  150  155  160
Lys Glu Pro Gly Gly Ser Val Gly Ser Gly Ser Gly Ser Glu Asn
165  170  175
Leu Tyr Phe Gln His Gly Glu Gly Thr Phe Thr Ser Asp Leu Ser Lys
180  185  190
Gln Met Glu Glu Glu Ala Val Arg Leu Phe Ile Glu Trp Leu Lys Asn
195  200  205
Gly Gly Pro Ser Ser Gly Ala Pro Pro Ser Gly Gly Glu Gly Ser
210  215  220
Gly Glu Gly Ser Gly Gly Glu Gly Ser Gly Ser Gly Gly Glu Gly Glu
225  230  235  240
Gly Ser Gly Ser Gly Gly Gly Gly Ser Gly Ser Gly Ser Gly
245  250  255
Gly Glu Gly Ser Gly Ser Gly Gly Glu Gly Glu Gly Gly Gly Ser
260  265  270
Gly Glu Gly Ser Gly Gly Ser Gly Ser Gly Gly Glu Gly Gly Gly
275  280  285
Glu Gly Ser Gly Gly Ser Gly Ser Gly Gly Glu Gly Glu Gly Glu
290  295  300
Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Ser Gly Glu Gly
305  310  315  320
Glu Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Ser Gly
325  330  335
Glu Gly Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Ser Gly Ser
340  345  350
Ser Gly Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Glu Gly Gly
355  360  365
Glu Gly Ser Gly Gly Glu Gly Ser Gly Ser Gly Gly Ser Gly Glu
370  375  380
Gly Ser Gly Gly Ser Gly Gly Ser Gly Ser Gly Ser Gly Ser Glu
385  390  395  400
Gly Glu Gly Ser Gly Gly Glu Gly Gly Gly Ser Gly Ser Gly Glu
405  410  415
Ser Gly Gly Gly Gly Gly Ser Gly Ser Gly Gly Ser Gly Ser Gly
420  425  430
Glu Gly Ser Gly Gly Gly Ser Gly Ser Gly Ser Gly Ser Gly Glu
435  440  445
Gly Ser Gly Gly Ser Gly Gly Gly Ser Gly Ser Gly Ser Gly Glu
450  455  460
Glu Gly Ser Gly Gly Ser Gly Gly Gly Ser Gly Ser Gly Gly Ser
465  470  475  480
Glu Gly Ser Gly Gly Ser Gly Gly Ser Gly Ser Gly Ser Gly Glu
485  490  495
Ser Gly Gly Gly Ser Gly Ser Gly Ser Gly Ser Gly Ser Gly Ser
500  505  510
Gly Ser Gly Ser Gly Gly Ser Gly Ser Gly Ser Gly Ser Gly Glu
515  520  525
Gly Ser Gly Ser Gly Gly Ser Gly Ser Gly Ser Gly Ser Gly Glu
530  535  540
Gly Glu Gly Ser Gly Gly Ser Gly Ser Gly Gly Ser Gly Ser Gly
545  550  555  560
Ser Gly Glu Gly Gly Ser Gly Ser Gly Glu Gly Ser Gly Ser Glu
565  570  575
Gly Ser Gly Ser Gly Gly Ser Gly Ser Gly Ser Gly Ser Gly Glu
580  585  590
Ser Gly Ser Gly Ser Gly Ser Gly Ser Gly Ser Gly Ser Gly Glu
595  600  605
Glu Gly Ser Gly Ser Gly Ser Gly Ser Gly Ser Gly Ser Gly Glu
610  615  620
Glu Gly Gly Ser Gly Ser Gly Ser Gly Gly Ser Gly Glu Gly Ser
625  630  635  640
Gly Ser Gly Gly Ser Gly Ser Gly Ser Glu Gly Ser Glu Glu
Glu Gly Ser Gly Glu Gly Glu Gly Ser Gly Ser Gly Ser Gly Ser Gly Ser Gly 660 665 670
Ser Gly Glu Gly Ser Gly Ser Gly Ser Gly Gly Glu Gly Ser Ser Gly 705 710 715 720
Gly Ser Gly Glu Gly Ser Gly Gly Ser Gly Ser Gly Ser Gly Gly Glu Gly 725 730 735
Gly Ser Glu Gly Glu Gly Ser Gly Gly Ser Gly Gly Ser Gly Ser Glu 740 745 750
Glu Gly Ser Ser Gly Gly Ser Gly Glu Gly Gly Ser Gly Ser Gly 765 765
Glu Gly Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly 770 775 780
Ser Gly Glu Gly Ser Gly Gly Ser Gly Ser Gly Ser Gly 785 790 795

<210> SEQ ID NO 512
<211> LENGTH: 144
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide

<400> SEQUENCE: 512
Gly Glu Gly Ser Gly Glu Gly Ser Gly Gly Ser Gly Ser Gly Ser Gly 1 5 10 15
Gly Ser Gly Ser Gly Ser Gly Gly Ser Gly Ser Gly Ser Gly Gly Ser Gly 35 40 45
Gly Ser Glu Gly Glu Gly Ser Gly Gly Ser Gly Ser Gly Ser Gly Ser Gly 50 55 60
Glu Gly Ser Ser Gly Ser Gly Gly Ser Gly Ser Gly Ser Gly Ser Ser 65 70 75 80
Glu Gly Glu Gly Ser Gly Gly Ser Gly Ser Gly Ser Gly Ser Gly Ser 85 90 95
Gly Ser Ser Gly Gly Gly Ser Gly Ser Gly Ser Gly Ser Gly Ser Gly 100 105 110
Gly Ser Gly Ser Gly Gly Ser Gly Ser Gly Ser Gly Ser Gly Ser Gly 115 120 125
Gly Ser Gly Ser Gly Ser Gly Ser Gly Ser Gly Ser Gly Ser Gly Ser 130 135 140

<210> SEQ ID NO 513
<211> LENGTH: 144
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide

<400> SEQUENCE: 513
Gly Glu Gly Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Ser 1 5 10 15
Gly Glu Gly Glu Gly Ser Glu Gly Ser Glu Gly Glu Gly Ser Glu Gly Ser Glu Gly Ser 20 25 30
Glu Gly Ser Glu Gly Glu Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Glu Gly Ser 35 40 45
Gly Ser Glu Gly Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Ser Glu 50 55 60
Glu Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Ser 65 70 75 80
Glu Gly Glu Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Ser 85 90 95
Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Ser Glu 100 105 110
Gly Glu Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Ser 115 120 125
Gly Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Ser Glu 130 135 140

<210> SEQ ID NO 514
<211> LENGTH: 144
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide

<400> SEQUENCE: 514
Gly Glu Gly Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Ser 1 5 10 15
Gly Glu Gly Glu Gly Ser Glu Gly Ser Glu Gly Glu Gly Ser Glu Gly Ser Glu 20 25 30
Glu Gly Ser Glu Gly Glu Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Glu Gly 35 40 45
Gly Ser Glu Gly Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Ser 50 55 60
Glu Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Ser 65 70 75 80
Glu Gly Glu Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Ser 85 90 95
Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Ser Glu 100 105 110
Gly Glu Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Ser Glu 115 120 125
Gly Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Ser Glu 130 135 140

<210> SEQ ID NO 515
<211> LENGTH: 144
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide

<400> SEQUENCE: 515
Gly Glu Gly Ser Gly Glu Gly Ser Gly Glu Gly Ser Gly Glu Gly Ser Gly  
1     5     10     15
Gly Glu Gly Ser Gly Glu Gly Ser Gly Glu Gly Ser Gly Glu Gly Ser Gly  
20    25    30
Gly Gly Ser Glu Gly Glu Gly Ser Gly Glu Gly Ser Gly Glu Gly Glu Gly  
35    40    45
Gly Ser Glu Gly Glu Gly Ser Gly Glu Gly Ser Gly Glu Gly Ser Gly Glu Gly  
50    55    60
Gly Gly Ser Glu Gly Ser Gly Glu Gly Ser Gly Glu Gly Ser Gly Ser Gly Ser  
65    70    75    80
Gly Gly Glu Gly Ser Gly Glu Gly Ser Gly Glu Gly Ser Gly Glu Gly Ser Gly  
85    90    95
Gly Ser Glu Gly Glu Gly Ser Gly Glu Gly Ser Gly Glu Gly Ser Gly Ser Gly Ser  
100   105   110
Gly Gly Ser Glu Gly Ser Gly Glu Gly Ser Gly Glu Gly Ser Gly Ser Gly Ser  
115   120   125
Gly Gly Ser Glu Gly Ser Gly Glu Gly Ser Gly Glu Gly Ser Gly Ser Gly  
130   135   140

<210> SEQ ID NO 516
<211> LENGTH: 144
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide

<400> SEQUENCE: 516
Gly Glu Gly Gly Ser Gly Glu Gly Ser Gly Glu Gly Ser Gly Glu Gly  
1     5     10     15
Gly Glu Gly Gly Ser Gly Glu Gly Ser Gly Glu Gly Ser Gly Glu Gly Ser Gly  
20    25    30
Gly Gly Ser Glu Gly Glu Gly Ser Gly Glu Gly Ser Gly Glu Gly Glu Gly  
35    40    45
Gly Ser Glu Gly Glu Gly Ser Gly Glu Gly Ser Gly Glu Gly Ser Gly Glu Gly  
50    55    60
Gly Gly Ser Glu Gly Ser Gly Glu Gly Ser Gly Glu Gly Ser Gly Glu Gly  
65    70    75    80
Ser Gly Ser Glu Gly Glu Gly Ser Gly Glu Gly Ser Gly Glu Gly Glu Gly  
85    90    95
Gly Gly Ser Glu Gly Glu Gly Ser Gly Glu Gly Ser Gly Glu Gly Ser Gly  
100   105   110
Gly Gly Ser Glu Gly Glu Gly Ser Gly Glu Gly Ser Gly Glu Gly Ser Gly  
115   120   125
Gly Gly Ser Glu Gly Ser Gly Glu Gly Ser Gly Glu Gly Ser Gly Ser Gly  
130   135   140

<210> SEQ ID NO 517
<211> LENGTH: 144
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide

<400> SEQUENCE: 517
<table>
<thead>
<tr>
<th>Gly Glu Gly Ser Gly Glu Gly Ser Gly Glu Gly Ser Gly Glu</th>
</tr>
</thead>
<tbody>
<tr>
<td>1  5  10  15</td>
</tr>
<tr>
<td>20 25 30</td>
</tr>
<tr>
<td>35 40 45</td>
</tr>
<tr>
<td>50 55 60</td>
</tr>
<tr>
<td>65 70 75 80</td>
</tr>
<tr>
<td>95 90 95</td>
</tr>
<tr>
<td>100 105 110</td>
</tr>
<tr>
<td>115 120 125</td>
</tr>
<tr>
<td>130 135 140</td>
</tr>
</tbody>
</table>

<210> SEQ ID NO 518
<211> LENGTH: 144
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide

<400> SEQUENCE: 518

Gly Glu Gly Gly Ser Gly Glu Gly Ser Gly Glu Gly Ser Gly Glu 1  5  10  15  
Gly Glu Gly Gly Ser Gly Glu Gly Ser Gly Glu Gly Ser Gly Glu 20 25 30
Glu Gly Ser Glu Gly Glu Gly Ser Gly Glu Gly Ser Gly Glu Gly 35 40 45
Gly Ser Glu Gly Glu Gly Ser Gly Glu Gly Ser Gly Glu Gly Ser 50 55 60
Glu Gly Ser Glu Gly Glu Gly Ser Gly Glu Gly Ser Gly Glu Gly 65 70 75 80
Glu Gly Glu Gly Ser Glu Gly Glu Gly Ser Gly Glu Gly Ser Gly 95 90 95
Gly Ser Glu Gly Glu Gly Ser Gly Glu Gly Ser Gly Glu Gly Ser 100 105 110
Ser Gly Glu Gly Ser Gly Glu Gly Ser Gly Glu Gly Glu 115 120 125
Gly Gly Glu Gly Ser Gly Glu Gly Ser Gly Glu Gly Glu 130 135 140

<210> SEQ ID NO 519
<211> LENGTH: 144
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide

<400> SEQUENCE: 519
Gly Glu Gly Ser Gly Glu Gly Ser Gly Glu Gly Ser Gly Glu Gly 1 5 10 15
Glu Ser Gly Glu Gly Glu Gly Ser Gly Glu Gly Ser Gly 20 25 30
Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly 35 40 45
Gly Ser Glu Gly Glu Gly Ser Gly Ser Glu Gly Ser Gly Ser Glu Gly 50 55 60
Glu Gly Ser Glu Gly Ser Glu Gly Ser Gly Glu Ser Gly Ser Glu Ser 65 70 75 80
Glu Gly Glu Gly Ser Glu Gly Ser Gly Glu Gly Ser Glu Gly Ser Glu Gly 95 99 95
Gly Ser Glu Gly Glu Gly Ser Gly Ser Glu Gly Ser Glu Ser Glu Gly 100 105 110
Ser Gly Glu Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Glu Gly Glu 115 120 125
Gly Gly Glu Gly Ser Glu Gly Ser Glu Gly Ser Gly Glu Gly Glu 130 135 140

<210> SEQ ID NO 520
<211> LENGTH: 144
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide

<400> SEQUENCE: 520

Gly Glu Gly Gly Ser Gly Glu Gly Ser Gly Glu Gly Ser Gly Glu Gly 1 5 10 15
Gly Glu Gly Glu Gly Ser Gly Glu Gly Ser Gly Glu Gly Ser Glu Ser 20 25 30
Glu Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Glu Gly 35 40 45
Gly Ser Glu Gly Glu Gly Ser Gly Ser Glu Gly Ser Gly Ser Glu Gly 50 55 60
Glu Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Ser Glu Ser 65 70 75 80
Glu Gly Glu Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Glu Ser Glu 85 90 95
Gly Ser Glu Gly Glu Gly Ser Gly Ser Glu Gly Ser Glu Ser Glu Gly 100 105 110
Gly Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Ser Glu Gly Glu Gly 115 120 125
Gly Gly Ser Glu Gly Ser Gly Gly Ser Glu Gly Ser Gly Glu Gly Ser Glu 130 135 140

<210> SEQ ID NO 521
<211> LENGTH: 144
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide

<400> SEQUENCE: 521
-continued

<table>
<thead>
<tr>
<th>195</th>
<th>200</th>
<th>205</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glu Gly Ser Glu Gly Gly Ser Glu Gly Ser Gly Gly Ser Gly</td>
<td>225</td>
<td>230</td>
</tr>
<tr>
<td>Ser Gly Glu Gly Gly Gly Ser Glu Gly Ser Gly Gly Ser Gly</td>
<td>290</td>
<td>295</td>
</tr>
<tr>
<td>Gly Glu Gly Ser Glu Gly Gly Gly Ser Gly Gly Ser Gly Gly Glu Glu</td>
<td>305</td>
<td>310</td>
</tr>
<tr>
<td>Gly Glu Gly Ser Glu Gly Ser Glu Gly Gly Gly Gly Ser Gly Gly Ser Gly</td>
<td>360</td>
<td>365</td>
</tr>
<tr>
<td>Ser Glu Gly Glu Gly Gly Gly Gly Ser Gly Gly Ser Gly</td>
<td>375</td>
<td>380</td>
</tr>
<tr>
<td>Gly Ser Gly Gly Gly Ser Gly Ser Gly Ser Gly Gly Ser Gly Gly Glu Glu</td>
<td>395</td>
<td>400</td>
</tr>
<tr>
<td>Glu Gly Ser Glu Gly Ser Glu Gly Gly Ser Gly Gly Ser Gly</td>
<td>405</td>
<td>410</td>
</tr>
<tr>
<td>Glu Gly Glu Gly Ser Glu Gly Gly Ser Glu Gly Ser Gly Ser Gly Ser Gly</td>
<td>420</td>
<td>425</td>
</tr>
<tr>
<td>Gly Ser Ser Gly Gly Gly Ser Gly Gly Ser Gly Ser Gly Ser Gly Ser Gly</td>
<td>450</td>
<td>455</td>
</tr>
<tr>
<td>Gly Ser Glu Gly Gly Ser Glu Gly Ser Gly Ser Gly Ser Gly Ser Gly Ser Gly</td>
<td>465</td>
<td>470</td>
</tr>
<tr>
<td>Gly Glu Gly Ser Glu Gly Ser Gly Gly Ser Gly Gly Ser Gly</td>
<td>485</td>
<td>490</td>
</tr>
<tr>
<td>Ser Glu Gly Gly Ser Glu Gly Ser Gly Gly Glu Gly Gly Ser Gly</td>
<td>500</td>
<td>505</td>
</tr>
<tr>
<td>Gly Ser Glu Gly Gly Gly Gly Ser Gly Ser Gly Ser Gly Ser Gly Gly Glu</td>
<td>515</td>
<td>520</td>
</tr>
<tr>
<td>Gly Ser Glu Gly Gly Gly Gly Ser Glu Gly Ser Gly Ser Gly Glu</td>
<td>530</td>
<td>535</td>
</tr>
<tr>
<td>Glu Gly Ser Glu Gly Gly Ser Glu Gly Gly Gly Ser Gly Ser Gly</td>
<td>545</td>
<td>550</td>
</tr>
<tr>
<td>Gly Gly Glu Gly Gly Ser Glu Gly Gly Ser Gly Gly Ser Gly</td>
<td>565</td>
<td>570</td>
</tr>
<tr>
<td>Ser Gly Gly Ser Gly Ser Gly Gly Ser Gly Ser Lys Gly</td>
<td>580</td>
<td>585</td>
</tr>
<tr>
<td>Glu Glu Leu Phe Thr Gly Val Val Pro Ile Leu Val Glu Leu Asp Gly</td>
<td>595</td>
<td>600</td>
</tr>
</tbody>
</table>
Asp Val Asn Gly His Lys Phe Ser Val Ser Gly Glu Gly Gly Asp  
   610  615  620
Ala Thr Tyr Gly Lys Leu Thr Leu Lys Phe Ile Cys Thr Thr Gly Lys  
   625  630  635  640
Leu Pro Val Pro Trp Pro Thr Leu Val Thr Thr Phe Ser Tyr Gly Val  
   645  650  655
Gln Cys Phe Ser Arg Tyr Pro Asp His Met Lys Arg His Asp Phe Phe  
   660  665  670
Lys Ser Ala Met Pro Glu Gly Tyr Val Gln Glu Arg Thr Ile Ser Phe  
   675  680  685
Lys Asp Asp Asp Tyr Lys Thr Arg Ala Glu Val Lys Phe Glu Gly  
   690  695  700
Asp Thr Leu Val Asn Arg Ile Glu Leu Lys Gly Ile Asp Phe Lys Glu  
   705  710  715  720
Asp Gly Asn Ile Leu Gly His Leu Glu Gly Tyr Asn Tyr Asn Ser His  
   725  730  735
Asn Val Tyr Ile Thr Ala Asp Lys Gln Lys Asn Gly Ile Lys Ala Asn  
   740  745  750
Phe Lys Ile Arg His Asn Ile Glu Asp Gly Ser Val Gln Leu Ala Asp  
   755  760  765
His Tyr Glu Gln Asn Thr Pro Ile Gly Asp Gly Pro Val Leu Leu Pro  
   770  775  780
Asp Asn His Tyr Leu Ser Thr Gln Ser Ala Leu Ser Lys Asp Pro Asn  
   785  790  795  800
Glu Lys Arg Asp His Met Val Leu Leu Glu Phe Val Thr Ala Ala Gly  
   805  810  815
Ile Thr His Gly Met Asp Glu Leu Tyr Lys  
   820  825

<210> SEQ ID NO 524
<211> LENGTH: 6
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<222> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide
<400> SEQUENCE: 524
Ser Lys Val Ile Leu Phe  
    1      5

<210> SEQ ID NO 524
<211> LENGTH: 7
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 524
Glu Asn Leu Tyr Phe Gln Gly  
    1      5
A biologically active protein comprising at least two domains wherein (a) a first domain of said at least two domains comprises an amino acid sequence having and/or mediating said biological activity; and (b) a second domain of said at least two domains comprises an amino acid sequence consisting preferably of at least about 100 amino acid residues forming random coil conformation whereby said random coil conformation mediates an increased in vivo and/or in vitro stability of said biologically active protein.

The biologically active protein according to claim 149, wherein said second domain forming random coil conformation consists of alanine, serine and proline residues.

The biologically active protein according to claim 149 or 150, wherein said second domain forming random coil conformation comprises a plurality of amino acid repeats, wherein said repeat consist of Ala, Ser, and Pro residues and wherein no more than 6 consecutive amino acid residues are identical.

The biologically active protein according to any one of claims 149 to 151, wherein said proline residues constitute more than 4% and less than 40% of the amino acids of said second domain forming random coil conformation.

The biologically active protein according to any one of claims 149 to 152, wherein said second domain of said at least two domains comprises an amino acid sequence consisting of about 100 to 3000 amino acid residues forming random coil conformation.

The biologically active protein according to any one of claims 149 to 153, wherein said polypeptide with biological activity is selected from the group consisting of binding molecules, antibody fragments, cytokines, growth factors, hormones or enzymes.

The biologically active protein according to claim 154 wherein said binding molecule is selected from the group consisting of antibodies, Fab fragments, F(ab\'2)2 fragments, CDR derived peptidomimetics, single chain variable fragments (scFv), domain antibodies and lipocalins.

The biologically active protein according to any one of claims 149 to 154, wherein said polypeptide with biological activity is selected from the group consisting of granulocyte colony stimulating factor, human growth hormone, alpha-interferon, beta-interferon, gamma-interferon, tumor necrosis factor, erythropoietin, coagulation factor VIII, gp120/gp160, soluble tumor necrosis factor I and II receptor, interleukin 2 and neutrophil gelatinase associated lipocalin.

The biologically active protein according to any one of claims 149 to 156, wherein said increased in vivo stability of said biologically active protein is a prolonged plasma half life of said biologically active protein comprising said random coil forming second domain when compared to said biologically active protein lacking said random coil forming second domain.

A composition comprising the biologically active protein according to any one of claims 149 to 157.

The composition according to claim 158, which is a pharmaceutical composition, optionally further comprising a pharmaceutical acceptable carrier.

A nucleic acid molecule encoding the biologically active protein of any one of claims 149 to 157.

A vector comprising the nucleic acid of claim 159.

A cell comprising the nucleic acid according to claim 160 or the vector according to claim 161.

A method for the preparing a biologically active protein comprising culturing the cell according to claim 162 and isolating said biologically active protein from the culture.

A method of treating a disease condition selected from the group consisting of hormone deficiency related disorders, autoimmune disease, cancer, anemia, neovascular diseases, infectious/inflammatory diseases, thrombosis, myocardial infarction, diabetes, reperfusion injury, and a kidney disease, comprising administering to a subject in need thereof a composition according to claim 159.

* * * * *