The invention provides peptides with high affinity for streptavidin. These peptides may be expressed as part of fusion proteins to facilitate the detection, quantitation, and purification of proteins of interest.


(End of abstract on next page.)

Primary Examiner—Karen Cochrane Carlson
Assistant Examiner—Sheridan K Sneedlen
(74) Attorney, Agent, or Firm—Clark & Elbing LLP

ABSTRACT

The invention provides peptides with high affinity for streptavidin. These peptides may be expressed as part of fusion proteins to facilitate the detection, quantitation, and purification of proteins of interest.

21 Claims, 22 Drawing Sheets
OTHER PUBLICATIONS


* cited by examiner
Fig. 1A
Fig. 1B
**Fig. 2A**

- **X-axis:** Selection Round
- **Y-axis:** Percent of counts binding to SA and eluting with biotin

**Fig. 2B**

- **X-axis:** Column volumes
- **Y-axis:** Percent of counts binding to SA and eluting with biotin

- **Legend:**
  - mRNA-pep
  - mRNA-pep/RNAse
  - mRNA-pep/RNAse/biotin

**Biotin Added**
GCTCAGCTTGTAGGTATCTCAGTGGTGTTAGTGCTTGGCTGCTCCAAGCTGGCTGTGCAAGAACCACCCCGGTATCCAGCC
GACCCGCTGCCTTTATCCGGTAACCGGTATCGCTTTGAGTCACCCGGTAGAAGACAGCTTTTCTGCAGTCTATGCCCTG
TGTAACAGGATTAGCAGAGCCAGGATAATGAGTTGCCGCTACAGAGGTTTCTGAGGTTGCTGTCATTACCGGCTACACTA
GAAGGACTAATTTTGAATCTCGCTCGCTCTGAAGCGATTTACCCTTCGGAAGAGATTGTGATTGAATCGTCCTGATCCGG
CAAAACCACCGGCTTGAGCGGAGGTTTTTTGTTTGCAAGCCAGCAGATTACCGCCAGAAAAAAGGATTCAAGAGATCT
CTTGATCTTTTCCTAGGGGTCTGACGTCTAGTTGAAGGAACAAACACTGTTAGTTTTTGGTCATGAAACTAATACGG
TCTGCCATATAACAGTAATACAAAGGGGTGTTATGAGCCATATATTCAACGGAAACAGCTCTGCTCTAGCCCGATTTA
TTCCAACATCGGATCGTAAATTATTTGGTAATAAATGCGTCGGCAGTAATGTCCGGCAATCGAGTTCGCAAACTATCGT
TGATGGAAGCCCGATGCCCCCAGGATTTGTTTCTGAAAACATGCGCAAGAGTTAGCTGTGCAATGTGATCGAGT
GTCAAGACTAAGACCGGATTTATGCTCTTCCGACACTCAGCATTTTATCCGTACTCCGTAGTGTAGCATGTT
ACCTACACGTGATCCCAGGGAAACTCGATTCGGATTTAGAGATTTACCGTATGGGAAATATTGGTGTGAC
CGCTGGCAGCTGTCCTGCGCCGGTGTCATTGCATTCTGTTTTGTAAATTTGCCATCAGCGAGGCAGTTAAATGGCTGCC
GCTAGAGCCTAGTACGAATGAATAACGTTTGGTTTGATGACAGGTGGTTTCTACTTCTGTACAGGAAAGCCGT
TCGTTTATGGGATCAAGTATCATTTTTTTCTAAGAATATTAATCTGACTGCGATCAATTTAGGA
TGATATTGAAAATACAAATAGGGGTCCCCGCAATCCTCCGGGAAAATGGCACCCTGGAATGGTAAACATTATAT
TTGTTAAATACCGGTATTTGTTAATAATCGCTGATTTTTTAAACCATAAGCCCAAAATCGCGCAAAAATCCCTATA
ATCAAAAAGATAAGACCGAGATTTGTTGTTTTCGACTGATTGAAACAGAGCTCACTTTAAAGAAGCTGGACT
AACGTCGAAAGCGGAAACACACATTAGCAGCCGATTTTGCGCTATGGGAAAGACCGGGAGACGTTG
CGAGAAAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAA
ACACCAGCCCGCTCTATGCGCGCTTAAGCGCGCTACAGGGCGGTCCATTGCCCA

Fig. 6A (continued)
Fig. 8A

RNA

Protein

TMV AUG FLAG ORF His6

Fig. 8B

AUG

Fig. 8C

AUG AUG

Fig. 8D

AUG deletion

Nickel purification + anti-FLAG purification

Fig. 8E

RT-PCR

Type II restriction enzymes

Fig. 8F
Fig. 9A
Fig. 9A (continued)
**Fig. 10B**

- RU
- 0 50 100 150 200 250 300 350 400
- TIME (SECONDS)
- b c d e f

Line graphs showing changes over time with labels a, b, c, d, e, and f.
Fig. 11
Fig. 13

- Maltose Elution
- Imidazole Elution
- Biotin Elution
- Soluble Fraction
- Induced
- Uninduced

<table>
<thead>
<tr>
<th>Molecular Weight (kDa)</th>
<th>83</th>
<th>62</th>
<th>47.5</th>
<th>32.5</th>
<th>25</th>
</tr>
</thead>
<tbody>
<tr>
<td>浓缩</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>浓缩</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
STREPTAVIDIN-BINDING PEPTIDES AND
USES THEREOF

CROSS-REFERENCE TO RELATED
APPLICATIONS

This application claims benefit from U.S. Provisional application Ser. No. 60/244,541, filed Oct. 31, 2000, hereby incorporated by reference.

STATEMENT AS TO FEDERALLY SPONSORED
RESEARCH

This invention was funded by grant number R01GM53936 from the National Institutes of Health and grant number NCC-2-1069 from NASA. The government may have certain rights in the invention.

BACKGROUND OF THE INVENTION

In general, the invention features novel compounds and methods for purifying or detecting proteins of interest.

Determining the enzymatic activity, binding specificity, or three-dimensional structure of a protein often requires the purification of the protein from a complex mixture of other components, such as compounds present in a cell lysate or in vitro translation extract. With the number of known proteins increasing dramatically as a result of whole genome sequencing projects, it has become crucial to find alternatives to traditional, time-consuming monoclonal antibody production for generating affinity reagents for the detection and purification of proteins. In addition, purifying a novel protein using traditional column chromatography methods often requires much trial and error to develop a purification protocol that results in the recovery of the protein in high yield and purity.

Thus, purification methods are needed that may be generally applied to proteins of interest, that utilize inexpensive reagents, and that result in highly purified protein without requiring multiple chromatography steps.

SUMMARY OF THE INVENTION

The purpose of the present invention is to provide improved reagents for the purification, detection, or quantitation of proteins of interest. In particular, the high affinity, streptavidin-binding peptides of the present invention may be used as affinity tags for the purification and detection of fusion proteins containing proteins of interest.

Accordingly, in a first aspect, the invention provides a peptide which binds streptavidin with a dissociation constant less than 10 \( \mu M \) (that is, binds streptavidin more tightly than a \( K_d \) of 10 \( \mu M \)) and which is not disulfide bonded or cyclized. Preferably, the dissociation constant is equal to or less than 5 \( \mu M \), 1 \( \mu M \), 100 nM, 50 nM, 25 nM, 10 nM, or even 5 nM.

In one preferred embodiment, the dissociation constant is less than 10 \( \mu M \), 5 \( \mu M \), 1 \( \mu M \), 100 nM, 50 nM, or 25 nM; and greater than 0.01 nM, 0.1 nM, 1 nM, 5 nM, or 10 nM.

In another preferred embodiment, the peptide is disulfide bonded or cyclized. In another preferred embodiment, the dissociation constant is less than 23 nM, 10 nM, or 5 nM; and greater than 0.01 nM, 0.1 nM, or 1 nM.

In another related aspect, the invention provides a peptide which binds streptavidin with a dissociation constant less than 23 nM, 10 nM, or 5 nM. In one preferred embodiment, the peptide is disulfide bonded or cyclized. In another preferred embodiment, the dissociation constant is less than 23 nM, 10 nM, or 5 nM; and greater than 0.01 nM, 0.1 nM, or 1 nM.

In another preferred embodiment, the value of the dissociation constant is contained in one of the following ranges: 5 nM to 10 nM, 10 nM to 5 nM, 5 nM to 1 nM, or 5 nM to 0.1 nM.

In another related aspect, the invention provides a peptide which binds streptavidin with a dissociation constant less than 10 \( \mu M \), 5 \( \mu M \), 1 \( \mu M \), 100 nM, 50 nM, or 25 nM; and greater than 0.01 nM, 0.1 nM, 1 nM, 5 nM, or 10 nM.

In another preferred embodiment, the value of the dissociation constant is contained in one of the following ranges: 5 nM to 10 nM, 10 nM to 5 nM, 5 nM to 1 nM, or 5 nM to 0.1 nM.

In another related aspect, the invention provides a peptide which binds streptavidin with a dissociation constant less than 10 \( \mu M \), 5 \( \mu M \), 1 \( \mu M \), 100 nM, 50 nM, or 25 nM; and greater than 0.01 nM, 0.1 nM, 1 nM, 5 nM, or 10 nM. In another preferred embodiment, the dissociation constant is less than 5 \( \mu M \), 1 \( \mu M \), 100 nM, 50 nM, or 25 nM; and greater than 0.01 nM, 0.1 nM, 1 nM, 5 nM, or 10 nM.

In another related aspect, the invention provides a peptide which binds streptavidin with a dissociation constant less than 10 \( \mu M \), 5 \( \mu M \), 1 \( \mu M \), 100 nM, 50 nM, or 25 nM; and greater than 0.01 nM, 0.1 nM, 1 nM, 5 nM, or 10 nM. In another preferred embodiment, the dissociation constant is less than 10 \( \mu M \), 5 \( \mu M \), 1 \( \mu M \), 100 nM, 50 nM, or 25 nM; and greater than 0.01 nM, 0.1 nM, 1 nM, 5 nM, or 10 nM.

In another related aspect, the invention provides a peptide which binds streptavidin with a dissociation constant less than 10 \( \mu M \), 5 \( \mu M \), 1 \( \mu M \), 100 nM, 50 nM, or 25 nM; and greater than 0.01 nM, 0.1 nM, 1 nM, 5 nM, or 10 nM. In another preferred embodiment, the dissociation constant is less than 10 \( \mu M \), 5 \( \mu M \), 1 \( \mu M \), 100 nM, 50 nM, or 25 nM; and greater than 0.01 nM, 0.1 nM, 1 nM, 5 nM, or 10 nM.

In another related aspect, the invention provides a peptide which binds streptavidin with a dissociation constant less than 10 \( \mu M \), 5 \( \mu M \), 1 \( \mu M \), 100 nM, 50 nM, or 25 nM; and greater than 0.01 nM, 0.1 nM, 1 nM, 5 nM, or 10 nM. In another preferred embodiment, the dissociation constant is less than 10 \( \mu M \), 5 \( \mu M \), 1 \( \mu M \), 100 nM, 50 nM, or 25 nM; and greater than 0.01 nM, 0.1 nM, 1 nM, 5 nM, or 10 nM.
of interest from the sample. In one preferred embodiment, the protein of interest is recovered from the fusion protein by cleaving the streptavidin-binding peptide from the fusion protein.

In yet another aspect, the invention provides a method of detecting the presence of a fusion protein of the present invention in a sample. This method includes (a) contacting the sample with streptavidin or a streptavidin-containing compound under conditions that allow complex formation between the fusion protein and either streptavidin or the streptavidin-containing compound; (b) isolating the complex, and (c) detecting the presence of streptavidin or the streptavidin-containing compound in the complex or following recovery from the complex. The presence of streptavidin or the streptavidin-containing compound indicates the presence of the fusion protein in the sample. Preferably, step (c) also involves measuring the amount of streptavidin or the streptavidin-containing compound in the complex or following recovery from the complex. The amount of fusion protein in the sample is correlated with, and may be calculated from, the measured amount of streptavidin. For example, for a fusion protein containing a peptide that binds one molecule of streptavidin per molecule of peptide, the amount of fusion protein in the sample is predicted to be approximately the same as the amount of streptavidin measured. In one preferred embodiment, the amount of streptavidin is determined using Western or ELISA analysis with an antibody that reacts with streptavidin or that reacts with a compound that is covalently linked to streptavidin. In another preferred embodiment, streptavidin is covalently linked to an enzyme, radiolabel, fluorescent label, or other detectable group, and the amount of streptavidin is determined using standard techniques based on a characteristic of the detectable group such as its enzyme activity, radioactivity, or fluorescence.

In another aspect, the invention features a method of determining the affinity of a compound of interest for a target molecule. The method includes incubating a solution having a compound of interest with a detectable group and a free target molecule under conditions that allow complex formation between the compound of interest and the free target molecule. The solution is contacted with a target molecule immobilized on a solid support under conditions that allow complex formation between the compound of interest and the immobilized target molecule, and the compound of interest bound to the immobilized target molecule is then separated from the compound of interest not bound to the immobilized target molecule. To determine the affinity, either the amount of the compound of interest that is bound to the immobilized target or the amount of the compound of interest not bound to the target molecule is then measured. In yet another preferred embodiment, the steps of this method are repeated one or more times with a different concentration of free target molecule. In other preferred embodiments, both the amount of the compound of interest bound to the immobilized target molecule and the amount of the compound of interest not bound to the immobilized target molecule are measured. In another embodiment, only the amount of the compound of interest that is not bound to the immobilized target molecule is measured. In another embodiment, the target is immobilized on a solid support such as a microtiter plate, a bead, or the matrix of a spin-filter column. In another embodiment, the compound of interest that is bound to the immobilized target molecule is separated from the compound of interest that is not bound to the immobilized target molecule by centrifugation, for example, by spin-filtration or ultracentrifugation. In still other preferred embodiments, the target is a small molecule such as streptavidin, a streptavidin fusion protein, streptavidin covalently bound to an enzyme (e.g., horseradish peroxidase or alkaline phosphatase), or a ligand. The compound of interest may be a peptide that binds streptavidin, or may be a fusion protein that includes a peptide that binds streptavidin and a protein of interest. In other embodiments, the compound of interest may be labeled with a detectable group, such as a radiolabel, a fluorescent molecule, or an enzyme.

In preferred embodiments of various aspects of the invention, the amino acid sequence of the peptide includes at least 10, 25, 50, 75, or 100 consecutive amino acids or consists of between 5 and 150, 10 and 100, 20 and 75, or 50 and 50 amino acids, inclusive, of any one of SEQ ID Nos. 1–29 or 35. Preferably, the amino acid sequence of the peptides includes an LPQ, QPQ, EPQ, HPA, HPD, or HPL motif. In other preferred embodiments, the amino acid sequence includes any one of SEQ ID Nos. 1–29 or 35. In still other preferred embodiments, the peptide has an amino acid sequence that is at least 20, 30, 40, 50, 60, 70, 80, 90, 95, or 100% identical to any one of SEQ ID Nos. 1–29 or 35.

It is also contemplated that the affinity of the peptides of the present invention for streptavidin may be increased by incorporating disulfide bonds into, or cyclizing, the peptides. By constraining the peptides, the amount of disorder inherent in the peptides (i.e., entropy) decreases, and thus binding of these peptides to streptavidin may require less energy. It is also contemplated that the three-dimensional structure of peptides of the invention bound to streptavidin may be experimentally determined or modeled based on the known crystal structure of streptavidin and used to determine possible modifications to the peptides that may further improve their affinity for streptavidin.

As used herein, by “nucleic acid” is meant a sequence of two or more covalently bonded naturally-occurring or modified deoxyribonucleotides or ribonucleotides. By “peptide” is meant a sequence of two or more covalently bonded naturally-occurring or modified amino acids. The terms “peptide” and “protein” are used interchangeably herein.

By “covalently linked” is meant covalently bonded or connected through a series of covalent bonds. A group is covalently linked to a protein may be attached to the amino-terminus, carboxy-terminus, between the amino- and carboxy-termini, or to a side chain of an amino acid in the protein.

By “streptavidin” is meant any streptavidin molecule or fragment thereof or any protein that has an amino acid sequence that is at least 80, 90, 95, or 100% identical to a streptavidin molecule or fragment thereof (see, for example, Haeuptle et al. J. Biol. Chem. 258: 305, 1983). A preferred fragment of streptavidin is “core” streptavidin, which is a proteolytic cleavage product of streptavidin (Bayer et al. Biochem. J. 259,369–376, 1989). Preferably, a streptavidin molecule or fragment thereof is capable of binding biotin or any other streptavidin-binding molecule. Streptavidin or a streptavidin fragment may be modified chemically or through gene fusion technology or protein synthesis so that it is covalently linked to an enzyme, radiolabel, fluorescent label, or other detectable group. These detectable groups may be used to determine the presence or location of a streptavidin-bound fusion protein in a cell or sample or to quantify the amount of a streptavidin-bound fusion protein, using standard methods.

By a “streptavidin-containing compound” is meant any compound that includes streptavidin covalently bound to...
protein. For example, the streptavidin-bound fusion protein show the result of successive Oligo-dT and Ni-NTA peptide, is a linear peptide that does not have any of the fusions in which thousands of purified proteins need to be generated from the output of the seventh round of selection for binding streptavidin. Preferably, either the fusion protein or the streptavidin that has been released from the protein for binding streptavidin more tightly than the strength of binding represented by a particular dissociation constant. FIG. 1A is a schematic illustration of an in vitro selection process according to the invention, showing the structure of the library and the selection scheme. Members of the DNA library have, from the 5' to 3' end, a T7 RNA polymerase promoter (T7), a tobacco mosaic virus translation enhancer (TMV), a start codon (ATG), 88 random amino acids, a hexahistidine tag (H6), and a 3' constant region (Const). FIG. 1B is a picture of an SDS-PAGE gel of samples from the library at different stages of preparation. The first lane shows the result of translating the mRNA display template with 35S-methionine. Most of the counts represent free peptide (free pep), but a significant amount of mRNA-peptide covalent fusions are also present (mRNA-pep). There is also another band that is independent of added template (NS, non-specific), and some counts remain in the gel well. The band corresponding to the mRNA-peptide can be shifted to a position slightly higher than that for the free peptide by the addition of RNase A. The remaining lanes show the result of successive oligo-dT and Ni-NTA purifications, and finally reverse transcription (RT).

FIG. 2A is a bar graph showing the fraction of 35S counts from the displayed peptides that bound to streptavidin and eluted with biotin, at each round of selection. FIG. 2B is a graph showing the elution profile for the peptide library generated from the output of the seventh round of selection in FIG. 2A. The first fraction represents the flow-through. Biotin was added at the point indicated. The plot compares the binding of the intact, reverse-transcribed, displayed peptides (mRNA-pep), the same sample treated with RNase A, and the RNase-treated sample applied to a streptavidin column pre-saturated with biotin (excess biotin was washed away prior to exposing the library to the matrix).

Using the methods described herein, purification of a fusion protein based on its affinity for streptavidin has allowed the isolation of the fusion protein in significantly higher purity than that obtained using a hexahistidine affinity tag or maltose-binding protein affinity tag. Moreover, streptavidin is an inexpensive reagent that may be purchased unmodified or covalently labeled with a detectable group (such as FITC-streptavidin or alkaline phosphatase-conjugated streptavidin) or with a chromatography matrix (such as streptavidin-agarose). The availability of these reagents simplifies methods for detecting and purifying the fusion proteins of the present invention.

In addition, the streptavidin binding peptide tag are particularly useful when intermediate amounts of protein (~10–500 μg) need to be produced and purified in a high throughput manner. For example, highly parallel purification protocols may be performed in 96-well plates using streptavidin-derivatized magnetic beads and streptavidin binding peptide -tagged proteins expressed in E. coli or in coupled in vitro transcription/translation reactions. This facilitates purification, for example, for proteomics applications in which thousands of purified proteins need to be generated and purified in parallel.

Other features and advantages of the invention will be apparent from the following detailed description and from the claims.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1A is a schematic illustration of an in vitro selection process according to the invention, showing the structure of the library and the selection scheme. Members of the DNA library have, from the 5' to 3' end, a T7 RNA polymerase promoter (T7), a tobacco mosaic virus translation enhancer (TMV), a start codon (ATG), 88 random amino acids, a hexahistidine tag (H6), and a 3' constant region (Const).

FIG. 1B is a picture of an SDS-PAGE gel of samples from the library at different stages of preparation. The first lane shows the result of translating the mRNA display template with 35S-methionine. Most of the counts represent free peptide (free pep), but a significant amount of mRNA-peptide covalent fusions are also present (mRNA-pep). There is also another band that is independent of added template (NS, non-specific), and some counts remain in the gel well. The band corresponding to the mRNA-peptide can be shifted to a position slightly higher than that for the free peptide by the addition of RNase A. The remaining lanes show the result of successive oligo-dT and Ni-NTA purifications, and finally reverse transcription (RT).

FIG. 2A is a bar graph showing the fraction of 35S counts from the displayed peptides that bound to streptavidin and eluted with biotin, at each round of selection. FIG. 2B is a graph showing the elution profile for the peptide library generated from the output of the seventh round of selection in FIG. 2A. The first fraction represents the flow-through. Biotin was added at the point indicated. The plot compares the binding of the intact, reverse-transcribed, displayed peptides (mRNA-pep), the same sample treated with RNase A, and the RNase-treated sample applied to a streptavidin column pre-saturated with biotin (excess biotin was washed away prior to exposing the library to the matrix).

FIG. 3 is a list of the sequences of 20 clones from the seventh round of selection (SEQ ID Nos.: 1–20). The "#" column indicates the number of times each sequence was observed. The HPQ sequence is in bold type. Non-random sequences at the termini are underlined. The six C-terminal-most residues are not shown.
The present methods stem from the discovery of peptides encoding a fusion protein containing maltose-binding protein, a streptavidin-binding peptide (SEQ ID No.: 35) and a hexahistidine tag. Fig. 8E illustrates the reverse transcription of the mRNA display template that has initiated internally and displays the observed signal is for the streptavidin binding peptide-streptavidin interaction using the Spin-filter Binding Inhibition Assay (SBIA). The labeled streptavidin binding peptide-tagged peptide was incubated with a range of streptavidin concentrations and the amount not complexed was then determined after a short incubation with immobilized streptavidin. This analysis gave a $K_{D}$ of 2.5 nM for the interaction of the streptavidin binding peptide sequence with streptavidin.

Fig. 13 is a purity assay for a single-step purification of streptavidin binding peptide-tagged (multiply-tagged) protein from lysed cells. Purity is compared to samples processed upon the basis of the His-tag or the maltose-binding protein sequences contained in the same protein. Lanes 1 and 2 show lysed E. coli prior to and after IPTG-induction, respectively. Lane 3 shows the soluble fraction of E. coli lysate in streptavidin-binding buffer. Lanes 4 through 6 show approximately equal amounts of purified proteins from the elution fractions of the streptavidin binding peptide-tag purification. The His-tag purification, and the maltose-binding protein purification, respectively.

Fig. 14 illustrates a detection of a streptavidin binding peptide-tagged protein with streptavidin-derivatized horseradish peroxidase. The left-hand panel shows the Coomassie Brilliant Blue staining of an SDS-Tricine PAGE gel of the purified streptavidin binding peptide-tagged (multiply-tagged) protein and an E. coli extract. These same samples were also run on a different portion of the same gel and then transferred to a nitrocellulose membrane. The right-hand panel shows the result of the probing of this membrane with streptavidin-derivatized horseradish peroxidase. The only observed signal is for the streptavidin binding peptide-tagged protein, with no staining of other proteins in the extract.

Fig. 15 is a graph showing a $K_{D}$ determination of a streptavidin binding peptide-streptavidin interaction. The y axis shows the fraction of surface-bound peptide that is competed by the free streptavidin. This analysis gave a $K_{D}$ of 2.4 nM for the interaction of the streptavidin binding peptide sequence with streptavidin.

**DETAILED DESCRIPTION**

The present methods stem from the discovery of peptides that have unusually high affinities for streptavidin ($K_{D}$ of less than 10 nM). These peptides were selected from a library of randomized, non-constrained peptides using the mRNA display method. The high affinity of the selected
peptides was particularly surprising, given the fact that non-constrained linear peptide libraries generally do not yield high affinity ligands to proteins, except in cases where the protein normally functions in peptide recognition (Clackton et al., Trends Biotechnol. 12:173–184 (1994); Katz, Annu. Rev. Biophys. Biomol. Struct. 26:27–45, 1997). Many other peptides with high affinity for streptavidin may be isolated using the mRNA display method or any other selection method, such as ribosome display (Roberts, Curr. Opin. Chem. Biol. 3(3):268–73, 1999), or phage display (U.S. Pat. No. 5,821,047).

The binding characteristics of exemplary selected streptavidin-binding peptides are described in Table 1, and the sequences of these peptides are listed in FIG. 3. The first column of Table 1 lists the peptide name (SB1–SB20). For comparison, a non-selected sequence with two HPQ motifs spaced by 19 residues (called “non-selected”) is listed in row one. SB19-C4 is a truncation mutant of peptide SB19, described below. The peptides are grouped according to the number of HPQ and similar tripeptide motifs they possess. The second column shows the number of tripeptide motifs in each peptide, and the number of amino acid residues separating them. The third column represents the percentage of peptide binding and specifically eluting from a streptavidin column. This percentage ranged from 8.3% to as high as 88% for the selected peptides, compared to only 0.16% for the control, non-selected peptide with two HPQ motifs.

The fourth column shows the $K_{d}$, when known, for the interaction between streptavidin and the peptides, as measured in the EMSA assay described herein. The standard deviation in the $K_{d}$ is shown in the fifth column, based on the number of independent measurements (n, shown in parentheses). The dissociation constant ranged from 110 nM for peptide SB5 to 4.8 nM for peptides SB2.

### TABLE 1

<table>
<thead>
<tr>
<th>Peptide</th>
<th>Structure</th>
<th>binding and eluting %</th>
<th>$K_{d}$ (nM)</th>
<th>deviation (n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-selected</td>
<td>HPQ 19 HPQ</td>
<td>0.16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SB1</td>
<td>HPQ 19 HPQ</td>
<td>86</td>
<td>50</td>
<td>5.7 (4)</td>
</tr>
<tr>
<td>SB2</td>
<td>HPQ 19 HPQ</td>
<td>48</td>
<td>48</td>
<td>0.91 (8)</td>
</tr>
<tr>
<td>SB3</td>
<td>HPQ 23 HPQ</td>
<td>20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SB4</td>
<td>HPQ 43 HPQ</td>
<td>49</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SB5</td>
<td>HPQ 52 HPQ</td>
<td>72</td>
<td>110</td>
<td>22 (6)</td>
</tr>
</tbody>
</table>

Two HPQ motifs

<table>
<thead>
<tr>
<th>Peptide</th>
<th>Structure</th>
<th>binding and eluting %</th>
<th>$K_{d}$ (nM)</th>
<th>deviation (n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SB6</td>
<td>HPQ 4 HPQ</td>
<td>49</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SB7</td>
<td>HPQ 7 HPQ</td>
<td>28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SB8</td>
<td>HPQ 12 HP</td>
<td>27</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SB9</td>
<td>HPQ 12 HP</td>
<td>64</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SB10</td>
<td>HPQ 21 HPQ</td>
<td>15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SB11</td>
<td>HPQ 26 HPQ</td>
<td>88</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SB12</td>
<td>HPQ 30 HPQ</td>
<td>73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SB13</td>
<td>HPQ 32 HPQ</td>
<td>64</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SB14</td>
<td>HPQ 43 HPQ</td>
<td>11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SB15</td>
<td>HPQ 50 HPQ</td>
<td>44</td>
<td>92</td>
<td>16 (4)</td>
</tr>
<tr>
<td>SB16</td>
<td>HPQ 74 HPQ</td>
<td>50</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

One HPQ motif

<table>
<thead>
<tr>
<th>Peptide</th>
<th>Structure</th>
<th>binding and eluting %</th>
<th>$K_{d}$ (nM)</th>
<th>deviation (n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SB17</td>
<td>HPQ 88</td>
<td>83</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SB18</td>
<td>HPQ 88</td>
<td>88</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SB19</td>
<td>HPQ 88</td>
<td>85</td>
<td>10</td>
<td>1.8 (10)</td>
</tr>
<tr>
<td>SB19-C4</td>
<td>HPQ 88</td>
<td>9.8</td>
<td>4.9</td>
<td>0.88 (10)</td>
</tr>
</tbody>
</table>

No HPQ motif

To further characterize the binding of the selected peptides to streptavidin, truncation mutants for peptide SB19 were constructed to determine which regions were necessary for high affinity streptavidin-binding (FIG. 5). Deletion of up to 56 residues had no observable effect on the binding strength. For example, peptide SB19-C4 retained only the first 38 residues from the selected construct (plus the C-terminal sequence MMSGGCCKLG, SEQ ID No.: 36) and had a dissociation constant of 4.9 nM for streptavidin (Table 1). In contrast, N-terminal truncation mutations (N1–N3) resulted in a lower percentage of the encoded peptide specifically eluting from the streptavidin column (0.058 to 69% for the truncation mutants compared to 85% for full length SB19). These results suggested that the determinants for binding streptavidin were spread throughout the N-terminal 38 residues of the SB19 peptide.

High affinity streptavidin-binding peptides, such as those shown in Table 1, have a number of uses. For example, these peptides may be used for protein purification by expressing a protein of interest as a fusion protein joined to one or more of the streptavidin-binding peptides of the invention. In one such purification method, a sample containing the fusion protein is incubated with immobilized streptavidin. Proteins with no or weak affinity for streptavidin are washed away, and the fusion protein is then selectively eluted from the streptavidin matrix by addition of biotin, a biotin analog, another streptavidin-binding peptide, or any compound that competes with the fusion protein for binding to the matrix. Alternatively, the fusion protein may be eluted from the matrix by increasing or decreasing the pH of the buffer applied to the matrix.

As described in detail below, this general protocol was used in a one-step purification of a fusion protein containing a streptavidin-binding peptide from an E. coli extract, resulting in a high yield of very pure protein. This fusion protein contained the first 38 amino acids of the SB19-C4 peptide, which due to its small size was not expected to affect the three-dimensional structure or activity of the covalently-linked protein of interest. Purification of fusion proteins containing other streptavidin-binding peptides of the present invention may be performed similarly.

In addition, various modifications of the above purification protocol would be apparent to one skilled in the art (see, for example, Ausubel et al., supra), and such modifications are included in the invention. In particular, use of the streptavidin-binding peptides as affinity tags is desirable for high throughput protein production and purification. For example, purification of fusion proteins in a multi-well format may be conducted using magnetic streptavidin beads that are washed and eluted robotically. The methods of the present invention may also be adapted to purify fusion proteins from in vitro translation mixtures or from other extracts, such as those from prokaryotic, yeast, insect, or mammalian cells, using standard techniques. If necessary, avidin may be added to the extract to bind any free biotin in the extract before contacting a sample from the extract with streptavidin. Allowing any free biotin to bind avidin may prevent biotin from competing with the streptavidin-binding peptides for binding to streptavidin.

If desired, the presence of a fusion protein of the invention in a sample may be detected by incubating the fusion protein with streptavidin (i.e., unlabeled streptavidin or streptavidin that is labeled with a detectable group) under conditions that allow streptavidin to bind the fusion protein. Preferably, the unbound streptavidin is separated from the streptavidin-bound fusion protein. Then, the streptavidin that is bound to the fusion protein is detected. Alternatively, the streptavidin bound to the fusion protein is physically separated from the fusion protein and then detected, using standard methods.
For example, to detect streptavidin that is bound to the fusion protein or that has been separated from the fusion protein, Western or ELISA analysis may be performed using an antibody that reacts with streptavidin or that reacts with a compound that is covalently linked to streptavidin. If streptavidin is covalently linked to an enzyme, radiolabel, fluorescent label, or other detectable group, the amount of streptavidin may be determined using standard techniques based on a characteristic of the detectable group such as its enzyme activity, radioactivity, or fluorescence (see, for example, Ausubel et al., supra). Alternatively, streptavidin may be contacted with a streptavidin-binding compound that is covalently linked to an enzyme, radiolabel, fluorescent label, or other detectable group, and the detectable group may be assayed as described herein.

We have also developed an improved method to generate synthetic DNA libraries encoding full-length proteins, which may be used in a variety of selection methods to isolate proteins with desired binding affinities or activities. The generation of libraries of proteins containing a desired number of amino acids is often limited by the number of internal initiation events that result in truncated proteins and the number of frameshifts that result in premature stop codons or the removal of desired stop codons. For example, during solid phase DNA synthesis, insertions and deletions which cause frameshifts may occur due to imperfect coupling and capping efficiencies. In addition, the random regions in DNA templates may encode stop codons, resulting in premature truncation of the encoded protein. To address these problems, we have developed a method in which small DNA cassettes are synthesized, and an in vitro selection using the mRNA display technology is performed to enrich the library of DNA cassettes for sequences encoding two protein affinity tags. These DNA cassettes lack frameshifts and premature stop codons. The selected DNA cassettes are then cleaved with restriction enzymes and ligated to generate the full-length DNA library (FIGS. 8A–8F) (Cho et al., J. Mol. Biol. 297:309–319, 2000).

In one preferred embodiment of this method, mRNA display templates that contain a translation enhancer sequence operably-linked to an open reading frame that terminate in puromycin are generated as described previously (Cho et al., supra). The open reading frame encodes two different protein affinity tags, such as a FLAG tag and a hexahistidine tag. Preferably, one of the tags is located at the amino-terminus of the encoded peptide, and the other tag is located at the carboxy-terminus. The mRNA display templates are in vitro translated to generate mRNA displayed peptides (Cho et al., supra). mRNA displayed peptides encoded by templates that do not contain frameshifts or premature stop codons should contain both affinity tags. In contrast, templates that contain frameshifts or premature stop codons encode peptides without the C-terminal affinity tag (FIG. 8D). Additionally, mRNA display templates that initiate internally produce peptides without the N-terminal affinity tag (FIG. 8C). The library of mRNA displayed peptides is enriched for peptides containing both affinity tags by purification of the mRNA displayed peptides based on the presence of these tags (see, for example, Ausubel et al., supra). For example, the mRNA displayed peptides may be applied to a matrix designed to bind peptides containing one of the affinity tags, and the mRNA display peptides without the affinity tag are washed away. The mRNA display peptides containing the affinity tag are then eluted and applied to a second matrix designed to bind the other affinity tag. The mRNA display peptides recovered from this purification step are enriched for members containing both affinity tags and thus for full-length peptides. These mRNA displayed peptides are reverse transcribed to generate double-stranded DNA. The amplified DNA is then cleaved by restriction enzymes. Preferably, this restriction digestion removes the sequences encoding the affinity tags from the DNA cassettes. The cleaved DNA cassettes are then ligated to generate the full-length DNA templates.

The experiments described above were carried out as follows:

Generation of a Streptavidin-Binding Peptide Library

The mRNA display method for selecting peptides or proteins of interest takes advantage of the translation-terminating antibiotic puromycin, which functions by entering the A-site of ribosomes and forming a covalent bond with the nascent peptide. By covalently attaching puromycin to the 3′ end of an mRNA, a covalent link between a polypeptide and its encoding message can be achieved in situ during in vitro translation (Roberts et al., Curr. Opin. Struct. Biol. 9:521–529, 1999; Liu et al., Methods Enzymol. 318:268–293, 2000). These mRNA-peptide fusions can then be purified and subjected to in vitro selection, yielding the isolation of novel peptide ligands. A DNA library encoding polypeptides of 108 amino acids was synthesized as described (Cho et al., supra). The library consisted of short cassettes concatamersized together. Each cassette encoded a random peptide with a pattern of polar versus non-polar amino acid side chains compatible with forming an amphipathic α-helix or β-strand (Cho et al., supra). The random region was 88 amino acids long, followed by a C-terminal invariant region containing a hexahistidine tag (FIG. 1A).

The library had a complexity of 2.4×10^14 at the DNA level. It was transcribed using T7 RNA polymerase (FIG. 1A), after which a “linker” oligonucleotide was added to the 3′ end using T4 DNA ligase as described (Liu et al., supra; Cho et al., supra). The linker consisted of a 21 nucleotide long dA stretch, followed by a polyethylene glycol linker, followed by the sequence dA-dC-dC-puromycin (Liu et al., supra).

This puromycin-terminated mRNA was translated in vitro, using the Ambion (Austin, Tex.) in vitro translation kit under standard conditions for capped mRNA. The 10 mL reaction mixture was supplemented with 2 mCi 35S-methionine and a total methionine concentration of 10 μM. The reaction mixture also included 300 nM of the library of puromycin-linked mRNA molecules. After 1 hour at 30°C, MgCl2, and KCl were added to 20 and 710 mM, respectively, and the reaction mixture was further incubated at room temperature for five minutes to increase the yield of displayed peptides. This in vitro translation produced 1.2×10^15 polyepitides linked via the puromycin moiety to their encoding mRNAs.

These mRNA displayed peptides were then purified on oligo-dT cellulose (which binds to the oligo-dA sequence in the linker) to remove polyepitides not fused to mRNA. For this purification procedure, the reaction mixture was diluted 10-fold into oligo-dT-binding buffer (1M NaCl, 50 mM HEPES, 10 mM EDTA, 0.25% Triton X-100, and 5 mM 2-mercaptoethanol at pH 7.9) and 80 μg oligo-dT cellulose (type 7, Amersham-Pharmacia, Piscataway, N.J.) and incubated with agitation at 4°C for 30 minutes. The mixture was applied to a column (Poly-Prep chromatography column, BioRad, Hercules, Calif.), drained, washed with 10 mL oligo-dT-binding buffer, washed with 10 mL oligo-dT-wash buffer (300 mM NaCl, 20 mM HEPES, 1 mM EDTA, 0.25% Triton X-100, and 5 mM 2-mercaptoethanol at pH 7.9), and washed with 1 mL of 0.5xoligo-dT-wash buffer. The mRNA-
displayed peptides were eluted with 4.5 mL water plus 5 mM 2-mercaptoethanol into tubes containing Triton X-100 and bovine serum albumin (BSA, New England Biolabs, Beverly, Mass.) at final concentrations of 0.15% and 5 mg/mL, respectively.

The mRNA-displayed peptides that eluted from the oligo-dT cellulose column were further purified on Ni-NTA agarose, which binds to the hexahistidine tags on the polypeptides, to remove any mRNA not fused to polypeptides. The eluted fractions from the oligo-dT cellulose purification were exposed to 0.5 mL Ni-NTA-agarose, 4 μg/mL RNA (Boehringer-Mannheim, Indianapolis, Ind.), and 5 μg/mL BSA at pH 8.0) and incubated for 30 minutes at room temperature. The matrix was then drained, washed with 12 column volumes Ni-binding buffer, and eluted with the same buffer plus 100 mM imidazole. Eluted fractions were combined, de-salted using two successive NAP columns (Amersham-Pharmacia, Piscataway, N.J.) equilibrated in 1 mM Tris(hydroxymethyl)aminomethane, 0.1% Triton X-100, 5 mM 2-mercaptoethanol, 4 μg/mL tRNA (Boehringer-Mannheim, Indianapolis, Ind.), and 5 μg/mL BSA at pH 7.6.

The mRNA portion was then reverse transcribed using Superscript II (Gibco BRL, Rockville, Md.) according to the manufacturers instructions, except that the mRNA concentration was about 5 nM and the enzyme concentration was 1 U/μL. To ensure a high yield in the reaction, a mixture of two primers were used: 1 μM of "splitl" from the splinted ligation (Cho et al., supra), and 1 μM of the 3' PCR primer. After 30 minutes at 42°C, the temperature of the reaction mixture was raised to 50°C for 2 minutes, and then cooled over 5 minutes to room temperature to allow gradual peptide folding. Finally, the contents were de-salted using NAP columns and subjected to scintillation counting. By comparing the 35S counts of the purified, reverse transcribed mRNA-peptide fusions to the 35S-methionine stock and taking into consideration the total methionine concentration in the translation reaction (10 μM), the number of displayed peptides in this sample was determined to be 6.7×10^-12. This number also represents the complexity of the library, since it contained virtually no redundancy (the complexity of the library constructs, were responsible for the interaction with streptavidin. Treatment of the library with RNase A did not reduce the extent of binding/elution from the matrix (FIG. 2B). Also, biotin-saturated streptavidin showed no binding to the peptide library (FIG. 2B). These results demonstrated that the interaction of the selected peptides with the streptavidin matrix was specific for the unligated protein, rather than for any other component of the matrix.

Sequence Analysis of Selected Peptides

Thirty-three randomly chosen clones from the PCR DNA from round seven were chosen for sequencing. Twenty different sequences were observed (FIG. 3). Surprisingly, all 20 sequences were frame-shifted from the intended frame (frame 1) to frame 3 by deletion of two nucleotides or addition of one nucleotide. The designed pattern of polar and non-polar residues was therefore discarded, leaving an unpatterned, essentially random sequence. Since the selection, about half of the library members were in frame 1 throughout their entire open reading frames (Cho et al., supra). Frame 3 appears to have been enriched over frame 1 due to the increased frequency of the sequence HPQ. Frame 1 has a low incidence (1:45,000 library members) of the sequence HPQ due to the designed polar/non-polar pattern. By contrast, frame 3 had a much higher expected incidence of the HPQ sequence (1:64), similar in frequency to that of a library of the same length and with equal mixtures of all four nucleotides at each position (1:193). Also, frame 3 was rich in histidine, thus allowing retention on the Ni—NTA column. The Ni—NTA purification protocol was intended to eliminate library mRNA molecules not displaying peptide, but was not performed under sufficiently stringent conditions so as to elimate peptides with small numbers of histidines. Frame 2 had a high incidence of stop codons. Nineteen of the 20 clones had at least one HPQ motif, and five clones contained two such motifs (Table 1). The clones were organized according to the number of times the HPQ and related tripeptide motifs occur (Table 1). The number of amino acids between the two motifs, when present, ranged from four to 74.

Binding Affinities of Peptides

To rapidly assay each of the 20 selected peptides to determine their affinity for streptavidin, a new method for...
preparing, tagging and purifying the peptides was employed. For generation of the DNA-tagged peptides, plasmids containing single inserts were used as templates for PCR-amplification using the same 5' PCR primer as described for the library construction (Cho et al., supra), and a new 3' primer 5'ATAGCCGGTGCCCAAGTGGTACCCGGCCCGAAGC GA GT-3'; SEQ ID No. 30), which altered the 3' RNA sequence to ACUGGUGCCGGCUGCAACGCUUGGCA CGGCUAAU (SEQ ID No. 31). This sequence was designed to anneal to the photo-crosslinking linker, which has the sequence 5'-psoralen-TAGCCGGTGA-A17-cc-puromycin-3', in which the underlined bases are 3'-methoxy nucleotides and the remaining bases are deoxynucleotides (the oligonucleotide was synthesized using reagents from Glen Research, Sterling, Va.). This new primer changed the constant C-terminal peptide sequence from WSGGCHHHHHHSSA (SEQ ID No. 32) to WSGGCGLGTGY (SEQ ID No. 33), of which the last three amino acids may not be translated because they are annealed to the linker. Each DNA template was transcribed and gel purified as described (Cho et al., supra), and then incubated with the psoralen linker under the following conditions: 2 pM mRNA, 4 pM linker, 50 mM Tris(hydroxymethyl) aminomethane, 200 mM KCl, and 10 mM spermidine at pH 7.4 and 70°C for 2 minutes, and then cooled to 4°C over 5 minutes. Samples were then placed in the cold room in a 96 well plate (50 µL/well), one inch above which was suspended a UV lamp (366 nm, Ultraviolet Products, Inc., San Gabriel, Calif., model number UVL-21) for 15 minutes. Then, the reactions mixtures were de-salted using a G-50 Sephadex spin column (Boehringer Mannheim, Indianapolis, Ind.). The translation display reactions and oligo-T purification were carried out as above. Finally, RNase A (200 ng/mL, 10 minutes, room temperature) was added to degrade the mRNA, leaving peptides fused to a short DNA oligonucleotide. Complete degradation was confirmed by SDS-PAGE analysis.

The resulting purified DNA-tagged peptides (DTP) were analyzed in a streptavidin column-binding assay, in which ~500 pM 35S-labeled DTP were mixed with 50 µL of the streptavidin matrix in streptavidin-binding buffer, in a total volume of 300 µL, and incubated for 10 minutes at room temperature with agitation. Then, the contents were loaded onto a chromatography column. The column was drained and washed with 80 column volumes of streptavidin-binding buffer, and then eluted with three consecutive aliquots (3 column volumes each) of streptavidin-binding buffer plus 2 mM biotin over a 15 minute period. All fractions (flow-through, washes, elutions, and irreversibly bound counts) were analyzed by scintillation counting to determine the fraction of DTP that bound streptavidin and eluted with biotin (Table 1). The non-selected clone in which two HPQ motifs (encoded by 19 amino acids) were introduced contained the sequence MDEAHPOAGPVDQADARLIVQOGA LOHIPQGDTRMSGSGCKLGTGY (SEQ ID No. 34), in which the underlined portions are identical the HPQ regions of clone SB2.

The results of this analysis are shown in Table 1. For comparison, two HPQ motifs, separated by 19 residues, were introduced into a control, unselected member of the library. The low percentage of this control peptide that specifically eluted from the streptavidin column (0.16%) indicated that the presence of two HPQ motifs was not sufficient for high affinity binding. In contrast, a greater percentage of the selected peptides (8.3 to 88%) was retained on the column during the washing step and then specifically eluted with biotin.

The dissociation constants of the selected peptides for streptavidin were measuring using an electrophoretic mobility shift assay (EMSA). In this assay, DTP's were incubated with varying amounts of pure streptavidin (Pierce ImmunoPure Streptavidin, Rockford, Ill.) in streptavidin-binding buffer plus 5% glycerol to increase the density of the solution so that it could collect at the bottom of the gel well. After incubating at room temperature for 20 minutes, the reactions mixtures were moved into the cold room, where they remained for 10 minutes before being carefully loaded onto a 10% polyacrylamide:bisacrylamide (37.5:1, National Diagnostics, Atlanta, Ga.) gel (thickness 0.7 mm, height 16 cm, width 18 cm) containing 2xTBE, 0.1% Triton X-100 and 5% glycerol. The gel, which had been pre-run for 30 minutes at 13 watts, and the running buffer were pre-cooled to 4°C. Then, the gel was run in the cold room at 13 watts, which increased the temperature of the gel to about 20°C. The gel was run for 45 to 120 minutes, depending on the mobility of the particular DTP. Then, the gel was fixed in 10% acetic acid and 10% methanol for 15 minutes, transferred to electrophoresis paper (Ahlstrom, Mt. Holly Springs, Pa.), dried, and analyzed using a PhosphorImager (Molecular Dynamics, Sunnyvale, Calif.).

The short DNA oligonucleotide tag on the DTP's allowed them to migrate in a native gel, and the addition of unlabeled ligand (i.e., streptavidin) caused a mobility shift for several of the clones. The concentration of DTP's was less than 1 nM in each titration, and thus the dissociation constant (K_d) can be approximated by the concentration of streptavidin that results in half of the DTP being mobility-shifted. To determine the K_d, several different measurements were taken in the range of 25–75% of DTP bound (values outside of this range were unreliable due to background and close proximity of the bound and unbound bands in the gel). The K_d was determined using the equation K_d = [streptavidin] * R when R is the ratio of unbound to bound DTP (ratio of unshifted to shifted band). Independent measurements on gels prepared at different times were used for each clone (the number of different measurements, n, is shown in Table 1). Streptavidin concentrations were measured by UV_260 using the molar extinction coefficient of 57,000 per monomer.

Examples of these mobility shifts in the presence of streptavidin are shown in FIG. 4A. Some clones showed either no shift or poorly defined bands, suggesting that the lifetime of these complexes was too short for detection using this method. We chose five of the most well behaved clones and quantitatively examined their mobility shifts in response to a range of streptavidin concentrations. An example of a streptavidin titration experiment for peptide SB19 is shown in FIG. 4B, and the data is graphed in FIG. 4C. The dissociation constants for the clones ranged from 110 nM to less than 5 nM (Table 1). These surprisingly high affinities were comparable to those for monoclonal antibody-antigen interactions, demonstrating that even random, non-constrained peptide libraries can be a source of avid ligands to proteins that do not normally function in peptide binding.

Dissection of Clone SB19

Clone SB19 possessed only one HPQ motif, bound to 85% in the column binding assay, and had a K_d for streptavidin of 10 nM (Table 1). A series of C-terminal truncation constructs (C1–C4) were constructed and assayed in the streptavidin column-binding assay (FIG. 5). C-terminal truncation analysis of clone SB19 was performed using standard methods by amplifying the clone with the original 5' primer and a series of 3' primers that truncated the
The majority (87%) however, had an apparent KD of halogous primers were used for the N-terminal truncation the ability to bind the streptavidin column. Thirty percent of proper folding of the domains. same buffer containing 20 mM imidazole and then eluted with methionine codons to increase the 35S-incorporation. retained the streptavidin-binding peptide and thus retained peptide was inactive even at streptavidin concentrations >1 Pierce, Rockford Ill.) and then incubated at 4°C for 30 minutes of a buffer appropriate for the subsequent affinity Triton X-100, pH 7.4). The sample was applied directly to the immobilized streptavidin matrix (e.g., column volume 100 μl; UltraLink Immobilized Streptavidin Plus, Pierce, Rockford Ill.) and then incubated at 4°C for 30 minutes. The matrix was then washed with 40 column volumes of SBB and then eluted with 3 successive 2 column volume aliquots of SBB containing 2 mM biotin for 10 minutes each. Samples of each of the lysed uninduced and lysed induced cells, the soluble fraction, and the elution fraction were then analyzed on an 8% SDS-Tricine PAGE gel, and then stained with Coomassie Brilliant Blue. In a typical experiment, 1 ml of the soluble fraction of lysed induced cells was loaded onto 0.1 ml of the affinity matrix.

For the expression of the fusion protein, BL21 (DE3) cells were transformed with a plasmid containing a Malose Binding Protein—Streptavidin-binding Peptide—His,-Protein of Interest insert (FIG. II) which encodes a fusion protein containing, from the amino- to carboxy-terminus, maltose-binding protein, the first 38 residues of the SB19-C4 sequence, a hexahistidine tag, and another peptide called 2r18-19N (FIGS. 6A-6C). This insert was constructed using standard molecular biology techniques (see, for example, Ausubel et al., supra). Each of these domains of the fusion protein is separated by a few amino acids to allow proper folding of the domains.

A kanamycin-resistant colony was selected and grown overnight in 10 ml LB media with 50 mg/liter kanamycin at 37°C. This starter culture was diluted 100-fold into 1000 ml LB with 50 mg/liter kanamycin, and the culture was grown at 37°C to OD600 of 1.6–1.8 at 37°C. Expression of the fusion protein was induced by addition of 1 mM IPTG, and the culture was grown for another two hours. The cells were pelleted by centrifugation at 3000–5000 x g for 20 minutes. The pellet was resuspended in 5% of the original volume of a buffer appropriate for the subsequent affinity purification method. For purification on an amylose column the cells were resuspended in 1 mM EDTA and MBP buffer (10 mM HEPES.HCl, 10 mM HEPES.Na+, 200 mM KCl, 0.25% w/w Triton X-100, and 10 mM BME at pH 7.4) and frozen slowly at −20°C overnight. The sample was thawed in the morning and sonicated on ice. The cell lysate was obtained by collection of the supernatant after centrifugation at 14,000 x g for 20 minutes at 4°C.

To purify the fusion protein, the cell lysate was applied to a column containing immobilized streptavidin, with a capacity of about 1 mg/ml, that had been washed with an eight column volumes of MBP buffer. Then, the column was washed with 12 column volumes of MBP buffer. The fusion protein was eluted with MBP buffer containing 2 mM biotin. Samples of the cell lysate and eluted protein were analyzed by SDS-PAGE on an 8% gel (FIG. 7B). The lane containing the purified protein had a band of the expected size. No other bands were observed, except for a faint band of slightly higher mobility (FIG. 7B). This band was probably a deg-
This gel is shown in FIG. 13. A band of the correct size was purified using each of the three columns. Purification on immobilized streptavidin gave the highest purity sample. No other bands were visible by Coomassie Brilliant Blue-stained SDS-Tricine PAGE gel analysis. At least one impurity was detected in the Ni-NTA-purified sample, and several contaminants were apparent in the amylose-purified sample. These results suggest that the streptavidin binding peptide-tag provides superior purity after a single purification step from lysed induced cells.

Detection of Fusion Proteins Containing Streptavidin Binding Peptides Using Streptavidin-Derived Reagents

Recombinant proteins containing streptavidin affinity tags may also be detected using reagents that bind to these affinity tags. A wide range of streptavidin-derivatized reagents is commercially available, and as a consequence the streptavidin binding peptide-tag provides a versatile detection tool. To demonstrate the utility of this interaction, an experiment was performed in which a recombinant protein was probed with streptavidin-derivatized horseradish peroxidase.

In this experiment the multiply-tagged protein was purified on the streptavidin column. The purified protein and an E. coli cell lysate were electrophoresed in adjacent lanes of a SDS-Tricine PAGE gel as described above. The protein on the gel was transferred to nitrocellulose (Trans-blot transfer medium, 0.2 μm, Biorad catalog number 162-0112) at 10 V for 30 minutes using the manufacturer’s instructions (Trans-blot semi-dry transfer cell, Biorad catalog number 170-3940). Efficient protein transfer was confirmed by the presence of the pre-stained molecular weight markers on the nitrocellulose.

After transfer, the nitrocellulose was incubated in TBS (25 mM Tris(hydroxymethyl)aminomethane, 138 mM NaCl, 2.68 mM KC1, pH 7.4) plus 0.05% Polyoxyethylene-sorbitan monolaurate (TWEEN-20) and 3% BSA for 1 hour at room temperature. The blot was then briefly rinsed with the same buffer without BSA, and then a streptavidin-derivatized horseradish peroxidase conjugate (Amersham-Pharmacia, product number RPNI231) was added at a 1,000-fold dilution in TBS-0.05% Tween-20/3% BSA, and allowed to incubate for 1 hour at room temperature. The blot was then washed 3 times with TBS/0.05% Tween-20, and then one time with TBS. The HRP substrate (3,3′,5,5′-tetramethybenzidene, Promega catalog number W4120) was then added according to the manufacturer’s instructions, and the blot was developed for approximately 1 minute.

These results are shown in FIG. 14. The streptavidin binding peptide-tagged protein is readily observed, and no proteins from the E. coli lysate are labeled, thus indicating the specificity of this interaction. Streptavidin-derivatized horseradish peroxidase and other streptavidin-derivatized reagents therefore provide many alternative methods of detecting streptavidin binding peptide-tagged proteins on membranes, plates, tissue sections, etc. cetera.

Yield of Fusion Proteins Containing Streptavidin-Binding Peptides

When scaling up protein expression to one milligram and above, the capacity and expense of the affinity matrix often begins to become important. To measure the capacity of the three matrices described above for their respective tag sequences, each matrix was overloaded by observing the multiply-tagged protein. Cells over-expressing this protein were split into three aliquots and re-suspended in different buffers according to the matrix that was to be used. The sample applied to immobilized streptavidin was extracted into streptavidin-binding buffer, and the samples for the other two purifications were extracted into the buffers recommended by the manufacturers of the two matrices, as described above.

The amount of protein in the flow-throughs, washes, and elution fractions was measured using the Bradford assay. We confirmed that each column was overloaded by observing the multiply-tagged protein in the flow-through fractions from each column. Purification using Ni-NTA agarose (His-tag/imidazole) yielded 12 mg of protein per ml of matrix. The amylose column yielded 4.4 mg of protein per ml of matrix. The immobilized streptavidin column yielded 0.5 mg of protein per ml of matrix (0.53 mg/ml with a standard deviation of 0.07 mg/ml, n=4). The overall capacity of immobilized streptavidin for the streptavidin binding peptide-tag was therefore lower than that of the Ni-NTA agarose or amylose matrices for their respective tags, but significantly greater than that of immobilized antibody matrices for purifying proteins upon the basis of epitope tags.

Biacore Analysis of the Affinity of a Fusion Protein for Streptavidin

The SB19-C4 streptavidin-binding fusion protein was expressed and purified from E. coli. This fusion protein contained, from the amino- to carboxy-terminus, maltose-binding protein, the first 38 amino acids of the SB19-C4 sequence, and a hexahistidine tag (FIG. 1A, SEQ ID No. 41). The plasmid (FIG. 9A, SEQ ID No. 40) encoding this fusion protein was constructed using standard molecular biology techniques and used to express the fusion protein in E. coli.

This data was used to calculate an upper limit of ~x~0~10~½ for the dissociation rate, k, of 5x10³ M⁻¹s⁻¹, as described previously (BIACORE X Instrument Handbook, version AA, Biacore AB, Uppsala Sweden, 1997). To measure the dissociation rate, a pulse of 23, 11.5, or 5.75 μM streptavidin in the buffer described above was administered, and then buffer without streptavidin was washed over the chip (FIG. 10A). This data was used to calculate an upper limit of 2x10⁻⁷ s⁻¹ for the dissociation rate, k, (BIACORE X Instrument Handbook, supra). Based on these calculated association and dissociation rates, the association constant, K, for the binding of streptavidin by this fusion protein was less than 40 nM. This result confirms the high affinity binding of the SB19-C4 peptide for streptavidin that was observed in the streptavidin column-binding assay and the EMSA assay (Table 1). Additionally, this result demonstrates that this
peptide maintains its high affinity for streptavidin when expressed as part of a fusion protein.

Spin-Filter Binding Inhibition Assay Analysis of the Affinity of a Fusion Protein for streptavidin

To measure the \( K_d \) of the peptide described above, a second and more general method, termed a Spin-filter Binding Inhibition Assay (SBIA), was also utilized. This method is especially appropriate for cases in which numerous proteins or peptides are derived from in vitro selection techniques using immobilized targets such as phage display (Smith, G. P. et al. *Chem. Rev.* 97:391–410, 1997), mRNA display (Roberts et al. *Proc. Natl. Acad. Sci. USA* 94:12297–12302, 1997; Liu et al. *Methods Enzymol.* 318:268–293, 2000; Keefe et al. *Current Protocols in Molecular Biology* 2001; Keefe et al. *Nature* 410:715–718, 2001) or ribosome display (Jermutus et al. *Curr. Opin. Biotechnol.* 9:391–410, 1998). The principal benefit of this method is that it allows the affinities and specificitites of interacting proteins to be assayed before investing the time required to over-express and purify them in the quantities required for conventional affinity determinations.

In general, SBIA utilizes low concentrations (<\( K_d \)) of \( ^{35} \)S-labeled protein generated in a cell-free translation system. The labeled protein is exposed to the bead-immobilized target (in this case immobilized streptavidin), and then the flow-through is collected by centrifugation in a 0.2 pm Durapore® spin-filter. The fraction of counts that pass through the filter (and therefore did not bind to the column) indicates the fraction of labeled peptide that did not bind to the matrix. To determine the affinity of the interaction, the labeled protein is first exposed to a range of concentrations of non-immobilized target. After this mixture has reached equilibrium, it is briefly (1 minute) exposed to the bead-immobilized target, and then spin-filtered as above. The initial incubation with the soluble target will compete with the binding of the radio-labeled protein to the immobilized target. The amount of inhibition is directly related to the fraction of labeled protein that was bound to the free target before this mixture was exposed to the immobilized target. By plotting the immobilized target-binding inhibition against the concentration of the free target, the \( K_d \) can easily be derived.

Because this titration is based on the interaction of the protein with the free (not immobilized) target, it may be more accurate than methods that quantify peptide binding to immobilized targets. However, SBIA may underestimate the affinity if there is significant dissociation of the complex during the brief incubation with immobilized target. The \( K_d \) corresponds to the concentration of free target that half-inhibits the binding to the immobilized target. As long as the concentration of the labeled binder is significantly lower than the \( K_d \), the concentration of free target may be approximated as the total concentration of added target. This method may be generally useful for the determination of \( K_d \) values for protein-protein, protein-peptide, and protein-small molecule complexes.

To measure the \( K_d \) of the SB19-C4 peptide, this peptide was fused to a FLAG-tag (streptavidin binding peptide-FLAG) (Wilson et al. *Proc. Natl. Acad. Sci. USA* 98:3750–3755, 2001). Streptavidin binding peptide-FLAG was then translated in the presence of \( ^{35} \)S-labeled methionine by in vitro translation in reticulocyte lysate according to the manufacturer’s instructions (Red Nova, Novagen, Madison, Wis.) using a template concentration of 400 nM, 1 mM Tris(hydroxymethyl) amino methane, 5 mM 2-mercaptoethanol, 2 mM EDTA, 0.1% Triton-X 100, pH 7.4). The peptide was then diluted into the same buffer and mixed with a range of different streptavidin concentrations to give a set of 50 µl samples in which the SB19-FLAG peptide was at 200 nM and the streptavidin concentration ranged from 30 pM to 1 µM. Each of these samples was then incubated for 2 hours at 0° C. and then subsequently incubated for 1 minute with 10 µl samples of the washed and dried immobilized streptavidin matrix (Ultralink Immobilized Streptavidin Plus, Pierce, Rockford IL). The flow-throughs were then immediately collected by centrifugation in a 0.2 µm Durapore® spin-filter (Millipore, Bedford Mass.), and these were counted in a scintillation counter. These data were iteratively fitted to the following equation:

\[
Y = b + c \frac{K_d}{K_d + x}
\]

where \( y \) is the number of radioactive decompositions detected per minute in each flow-through, \( K_d \) is the dissociation constant of the complex, \( x \) is the concentration of free streptavidin, \( b \) is the number of counts per minute not competent to bind the matrix under the assay conditions, and \( c \) is the number of counts per minute competent to bind the matrix under the assay conditions. \( K_d \), \( b \), and \( c \) were iteratively determined using the program Deltagraph 4.0 (SPSS, Chicago IL).

In this experiment, the binding of labeled streptavidin binding peptide-tagged peptide was inhibited by pre-incubation with free streptavidin (FIG. 12). In this experiment 48% of the counts bound to immobilized streptavidin in the 1 minute slurry incubation in the absence of free streptavidin competitor, and this binding was completely inhibited by high concentrations (>100 nM) of streptavidin. Analysis of the binding curve gave a \( K_d \) of 2.5±0.1 nM, which compared very favorably with the \( K_d \) determined by other methods. For example, the same peptide was purified on a FLAG-affinity matrix, then bound to streptavidin immobilized on a microtiter plate. The labeled streptavidin binding peptide-tagged peptide was incubated with a range of streptavidin concentrations for 1 hour before being transferred to a streptavidin-coated plate and incubated for five minutes. The labeled peptide not bound to the plate was pipetted off. The binding affinity was measured by competition with varying concentrations of streptavidin. Analysis of the binding curve gave a \( K_d \) of 2.4±0.1 nM (FIG. 15)(2.4 nM; Wilson et al. *Proc. Natl. Acad. Sci. USA* 98:3750–3755, 2001). The streptavidin binding peptide-tag therefore bound to streptavidin with a ~50,000-fold higher affinity than does the Strep-tag II (Schmidt et al. *J. Mol. Biol.* 255:753–66, 1996). This higher affinity accounts for the fact that, after extensive washing (with 80 column volumes), the yield of retained peptide using the streptavidin binding peptide-tag was 2,200-fold higher than the yield using the Strep-tag II (Wilson et al. *Proc. Natl. Acad. Sci. USA* 98:3750–3755, 2001). Both the streptavidin binding peptide-tag and the Strep-tag II contain the tripeptide motif HPQ, but the linking sequence in the streptavidin binding peptide-tag presumably provides additional favorable contacts with streptavidin.

Other Embodiments

From the foregoing description, it will be apparent that variations and modifications may be made to the invention described herein to adapt it to various usages and conditions. Such embodiments are also within the scope of the following claims.

All publications mentioned in this specification are herein incorporated by reference to the same extent as if each independent publication or patent application was specifically and individually indicated to be incorporated by reference.
<160> NUMBER OF SEQ ID NOS: 41
<210> SEQ ID NO 1
<211> LENGTH: 101
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: selected peptide

<400> SEQUENCE: 1
Met Asp Glu Lys Thr His Cys Thr Ile Ser Met Asn Gly Ala Val Pro
1 5 10 15
Leu Val Pro His His His Pro Gln Gly Asp Pro Leu Arg Leu Leu His
20 25 30
Arg Pro Gln Pro Ala Leu Val Arg His Pro Gln Gly Asp Leu Val
35 40 45
Ala Leu Val Glu His Glu Val Asp Arg Gly Leu Val Ala Leu
50 55 60
Pro Glu Leu His Ala Glu Glu Gln Pro Val Arg Asp Leu Val
65 70 75 80
Gln Gly Pro Val Glu Gln Val Gln Val Asp Ala Leu Val Trp
85 90 95
Arg Leu Pro Pro Ser
100

<210> SEQ ID NO 2
<211> LENGTH: 101
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: selected peptide

<400> SEQUENCE: 2
Met Asp Glu Lys Thr His Cys Phe His Pro Gly Asp His Leu Val Arg
1 5 10 15
Leu Val Glu Leu Glu Leu Ala Ala Gly Leu Gln Arg Gln Gly
20 25 30
Gly Arg Gln Pro His Arg Leu Pro Arg Arg Arg Pro His His Leu Gln
35 40 45
Leu Leu Leu Asp Glu Ala His Pro Gln Ala Gly Pro Leu Arg Glu Arg
50 55 60
Ala His Gln Val Asp Gly Arg Leu Leu Leu His Pro Gln Gly
65 70 75 80
Asp Arg Leu Leu Gln Pro Gln Asp His Pro Leu Glu Leu Val Trp
85 90 95
Arg Leu Pro Pro Ser
100

<210> SEQ ID NO 3
<211> LENGTH: 101
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: selected peptide

<400> SEQUENCE: 3
Met Thr Arg Arg Pro Thr Ala Ser Ser Ser Ser Cys Val Arg His Leu
1 5 10 15
Leu Leu Arg Gln Gly Glu His Gly His Gln Ala Leu Glu Asp Arg Asp
   20    25  30

Lys Ala Arg His Val Arg Leu Val Gly Asp Val Glu Val Leu Gly
   35  40  45

Gly Leu Asp Arg Leu Ala Arg Ala Arg His Glu Ala Leu His Pro Gln
   50   55  60

Ala Gly Leu Val His Leu His Leu Gly Asp Leu Gly Gly His
   65  70  75  80

Leu Arg Leu Val Leu Glu Ala His Pro Gln Gly Asp Arg Leu Gly Leu
   85   90  95

Ala Val His His
   100

<210> SEQ ID NO 4
<211> LENGTH: 102
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: selected peptide
<400> SEQUENCE: 4

Met Asp Glu Lys Thr His Trp Gly Ile Ser Thr Trp Arg Gly Glu Pro
  1   5  10  15

Leu Leu His His Pro Gln Ala Gly Arg Leu Pro Leu Asp Arg Arg Arg
  20  25  30

Ala Arg His Arg Arg Ile Leu Gly Ala Glu Pro Gly Gly Val Asp His
  35  40  45

Gly Leu Arg Leu Glu Leu Leu Asp Asp His Arg Pro Leu Val Pro Asp
  50   55  60

His His Pro Gln Arg Gly Pro Leu Gln Arg Gly Asp Leu Pro Gln Val
  65  70  75  80

Val Pro Leu Val Arg Leu Arg His Ala His Val Leu Gly Leu Gly Leu
  85   90  95

Ala Ala Ala Thr Ile Thr
  100

<210> SEQ ID NO 5
<211> LENGTH: 102
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: selected peptide
<400> SEQUENCE: 5

Met Asp Glu Lys Thr His Trp Asn Val Tyr His Pro Gln Gly Asp
  1   5  10  15

Leu Leu Val Arg Gly His Gly His Asp Val Glu Ala Leu His Asp Gln
  20  25  30

Gly Leu His Gln Leu Asp Leu Val Gly Pro Pro Gln Val Val
  35  40  45

Arg Ala Leu Arg Gly Val Leu Gly Leu Arg Arg Leu Val Pro
  50   55  60

Leu Asp His Pro Gln Gly Ala Leu Asp Gln Ala Arg Gln Arg Pro
  65   70  75  80

Gln His Leu Leu Glu Leu His Arg Ala Leu Pro Pro Ala Leu Val
  85   90  95

Trp Arg Leu Pro Pro Ser
  100
<210> SEQ ID NO 6
<211> LENGTH: 162
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: selected peptide

<400> SEQUENCE: 6

Met Asp Glu Lys Thr His Trp Leu Asn Ser Phe Glu Leu Leu Ala
1 5 10 15
Arg Leu Asp Gly Leu Arg Glu Gly Asp His Pro Leu Val Leu Arg
20 25 30
His His Gly Leu Leu Asp Gly Leu Asp Gln Pro Leu Gly Arg His
35 40 45
Arg Ala Leu Asp Gly Val Arg Glu Gly Asp Arg Pro Leu Asp Gln
50 55 60
Gly Gly Glu Leu Gly Ala Leu Val Asp Asp Gly Glu Val
65 70 75 80
Leu Asp Gly Leu Val His Val Gln Leu His Asp Pro Leu Val
85 90 95
Cys Gly Cys His His His
100

<210> SEQ ID NO 7
<211> LENGTH: 161
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: selected peptide

<400> SEQUENCE: 7

Met Asp Glu Lys Thr His Trp Phe Gly Thr Leu Asn Ser Phe Pro Thr
1 5 10 15
His Trp Met Ser Ala Val Gly Asn Gly Lys Ile Asp Cys Ser Phe Asn
20 25 30
Met Aan Leu Ser Leu Asn His Trp Leu Ser Ser Gln His Pro Aasp Gly
35 40 45
Ala Leu Asp Gln Leu His Pro Gln Gly Aaa Leu Val Gly Arg
50 55 60
Asp Gly Val Val Gln Ala Leu Arg Leu Glu Gly Aaa Gln His
65 70 75 80
Arg Arg Ala Gln Arg Arg Ala Asp Arg His Arg Gln Leu Val Trp
85 90 95
Arg Leu Pro Pro Ser
100

<210> SEQ ID NO 8
<211> LENGTH: 162
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: selected peptide

<400> SEQUENCE: 8

Met Asp Glu Lys Thr His Cys Thr Ile Glu Leu Asn Phe Ser Phe Thr
1 5 10 15
His Trp Lys Leu His His Pro Gln Gly Asp Ala Leu Leu Asp Aasp
20 25 30
Gly Val Arg Pro His Pro Leu Ala Asp Gly Gly Gly Gly Leu Asp
35 40 45
Gln Gly Leu Gly His Arg Arg Gly Val Val Ala Glu Arg Ala Arg
50 55 60
Arg Asp Pro Glu Val Leu Glu Gly Leu Val Glu Arg His Arg Gly Leu
65 70 75 80
Val Pro Arg Leu Arg His Gly Glu Arg His Ala Glu Pro Leu Val
85 90 95
Trp Arg Leu Pro Pro Ser
100

<210> SEQ ID NO 9
<211> LENGTH: 102
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: selected peptide
<400> SEQUENCE: 9
Met Asp Glu Lys Thr His Cys Asn Thr Gly Leu Tyr Asp Gly Ala Ala
1 5 10 15
Asp Cys Phe Asn Glu Leu Asn Lys Asp Val Ala Pro Leu Val Glu Gly
20 25 30
Arg His Asp Leu Val Glu Leu Leu Leu Arg His Pro Gln Gly
35 40 45
Asp Pro Leu Val Ala His Arg Gln Leu Val His Pro Leu Leu Gly
50 55 60
Arg Gly Glu Arg His Arg Ala Leu Val Pro Gln Gln Glu His Gln
65 70 75 80
Pro His Arg Leu Gln Pro Val Val Asp Leu Gly Arg Arg Arg Leu Val
85 90 95
Trp Arg Leu Pro Pro Ser
100

<210> SEQ ID NO 10
<211> LENGTH: 103
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: selected peptide
<400> SEQUENCE: 10
Met Asp Glu Lys Thr His Trp His Glu Arg Ala Gln Glu Leu Val Gly
1 5 10 15
Gly Leu Leu Leu His Asp His Pro Gln Arg Leu Leu Leu Glu Pro Arg
20 25 30
Gly Pro Arg Pro Leu Arg Gly Leu Val His Glu Arg Gly His Gln Pro
35 40 45
Gln Pro Leu Ala Gly Arg Val Glu Ala Asp Gly Gly Leu Leu Arg
50 55 60
Asp Gly Gly Gly Leu Leu Glu Pro Leu Val Arg Glu Gly Asp His
65 70 75 80
Leu Glu Pro Leu Asp Asp Glu Leu Ala Gly Pro Arg Gly Leu Val
85 90 95
Trp Arg Leu Pro His His
100

<210> SEQ ID NO 11
<211> LENGTH: 102
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: selected peptide

<400> SEQUENCE: 11

Met Asp Glu Lys Thr His Trp His Glu Val His Leu Ala Asp
1 5 10 15
Gly Leu Glu Gln His Pro Gln Gly Gln Arg Arg Pro Leu Val Glu Arg
20 25 30
His Arg Gln Val Pro Arg Gly Leu Val Arg Glu Leu Gln His Glu Gly
35 40 45
Leu Pro Leu Glu His Pro Ala Gly Val His Val Ile Arg Leu His Gln
50 55 60
Gly Asp Asp Arg Asp Val Asp Gly Leu Val Asp Gly His Gly Arg Asp
65 70 75 80
Val Arg Gly Leu Arg Glu Val Gly Asp Gly Pro His Arg Leu Val
85 90 95
Trp Arg Leu Pro Pro Ser
100

<210> SEQ ID NO 12
<211> LENGTH: 101
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: selected peptide

<400> SEQUENCE: 12

Met Asp Lys Asp Pro Leu Leu Glu Glu Glu Leu Arg Glu Arg
1 5 10 15
Leu Val His His Pro Gln Gly Gly Leu Leu Pro Leu Arg Gly Gln Val
20 25 30
Gly His Ala Glu Arg Leu Gly Ala Glu Val Asp Leu Arg Gly
35 40 45
Gly Leu Leu Asp Glu Pro Gln Arg Ala Val Ala Gly Leu His His Val
50 55 60
Pro His Arg Val Gly Gln Arg Leu Val His Glu Val Arg Glu Leu Asp
65 70 75 80
Glu Gly Leu Leu Asp Glu Arg Asp Leu Arg Gln Arg Leu Val Trp
85 90 95
Arg Leu Pro Pro Ser
100

<210> SEQ ID NO 13
<211> LENGTH: 102
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: selected peptide

<400> SEQUENCE: 13

Met Glu Arg Glu Asp Pro Leu Asp Glu Gln Leu Arg Leu Arg Glu
1 5 10 15
Asp Leu Val His Pro Gln Gly Gly Ala Gin Ala Leu His Arg His
20 25 30
Asp Gly Gln His Val Pro Leu Arg Val Gin His Arg Leu Gin
35 40 45
<table>
<thead>
<tr>
<th>Peptide 1</th>
<th>Peptide 2</th>
<th>Peptide 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pro Gly Leu Gln His His Leu Glu Pro Gln Pro Leu Gly Leu Leu Gly</td>
<td>Glu Leu Gln Ala Arg Leu Gln Pro Leu Ala Gly Glu His Glu Asp</td>
<td>Gly Ala Gly Leu Gln Arg Val Pro Gly His Gln Gly Arg Arg Leu Val</td>
</tr>
<tr>
<td>50 55 60</td>
<td>65 70 75</td>
<td>85 90 95</td>
</tr>
</tbody>
</table>

Trp Arg Leu Pro Pro Ser
100

<table>
<thead>
<tr>
<th>Sequence 1</th>
<th>Sequence 2</th>
<th>Sequence 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Met Asp Glu Lys Thr His Arg Thr Leu Ser Val Ser Leu Ser Phe Asn</td>
<td>Asp Trp Leu Gly Gln Thr Lys Ala Cys Trp Arg Leu Val Glu Gly Leu</td>
<td>His Gly His Pro Gln Gly Leu Val Arg Glu His Glu Val Asp Val Leu</td>
</tr>
<tr>
<td>1 5 10 15</td>
<td>20 25 30</td>
<td>35 40 45</td>
</tr>
<tr>
<td>Pro Leu Ala Glu Glu Val Gln Val Val Gly Leu Ala Asp Gly</td>
<td>Val Glu Gln Pro Gly Gly Gly Leu Leu His His Asp Arg Gly Glu Pro Val</td>
<td>His Pro Leu Pro Asp His Ala Gly Gln Val Leu Gly Arg Leu Val Trp</td>
</tr>
<tr>
<td>50 55 60</td>
<td>65 70 75</td>
<td>85 90 95</td>
</tr>
<tr>
<td>Arg Leu Pro Pro Ser</td>
<td></td>
<td>100</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sequence 4</th>
<th>Sequence 5</th>
<th>Sequence 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Met Asp Glu Lys Thr His Trp Leu Glu Asp Leu Lys Gly Val Leu Lys</td>
<td>Asp Cys Leu Lys Asp Leu Met Asp Phe Thr Lys Asp Cys Arg Ser Pro</td>
<td>Arg Val Gln Pro Gln Pro Leu His His Arg Asp Gly Glu Pro Val</td>
</tr>
<tr>
<td>1 5 10 15</td>
<td>20 25 30</td>
<td>35 40 45</td>
</tr>
<tr>
<td>Pro Leu Leu Arg Ala Gly Arg Asp Leu Gly Leu Gly Pro Arg</td>
<td>Pro Leu Leu Arg Ala Gly Arg Asp Leu Gly Leu Gly Pro Arg</td>
<td>Ala Pro Arg Gln Ala Arg Pro Leu His His Gly Arg His Asp Arg</td>
</tr>
<tr>
<td>50 55 60</td>
<td>65 70 75</td>
<td>85 90 95</td>
</tr>
<tr>
<td>Glu Pro Leu Val Leu Gln Asp His Pro Gin Gly Gly Pro Leu Val Cys</td>
<td>Gly Cys His His</td>
<td>100</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sequence 7</th>
<th>Sequence 8</th>
<th>Sequence 9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Met Asp Glu Lys Thr His Trp Leu Glu Asp Leu Lys Gly Val Leu Lys</td>
<td>Asp Cys Leu Lys Asp Leu Met Asp Phe Thr Lys Asp Cys Arg Ser Pro</td>
<td>Arg Val Gln Pro Gln Pro Leu His His Arg Asp Gly Glu Pro Val</td>
</tr>
<tr>
<td>1 5 10 15</td>
<td>20 25 30</td>
<td>35 40 45</td>
</tr>
<tr>
<td>Pro Leu Leu Arg Ala Gly Arg Asp Leu Gly Leu Gly Pro Arg</td>
<td>Pro Leu Leu Arg Ala Gly Arg Asp Leu Gly Leu Gly Pro Arg</td>
<td>Ala Pro Arg Gln Ala Arg Pro Leu His His Gly Arg His Asp Arg</td>
</tr>
<tr>
<td>50 55 60</td>
<td>65 70 75</td>
<td>85 90 95</td>
</tr>
<tr>
<td>Glu Pro Leu Val Leu Gln Asp His Pro Gin Gly Gly Pro Leu Val Cys</td>
<td>Gly Cys His His</td>
<td>100</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sequence 10</th>
<th>Sequence 11</th>
<th>Sequence 12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Met Asp Glu Lys Thr His Trp Leu Glu Asp Leu Lys Gly Val Leu Lys</td>
<td>Asp Cys Leu Lys Asp Leu Met Asp Phe Thr Lys Asp Cys Arg Ser Pro</td>
<td>Arg Val Gln Pro Gln Pro Leu His His Arg Asp Gly Glu Pro Val</td>
</tr>
<tr>
<td>1 5 10 15</td>
<td>20 25 30</td>
<td>35 40 45</td>
</tr>
<tr>
<td>Pro Leu Leu Arg Ala Gly Arg Asp Leu Gly Leu Gly Pro Arg</td>
<td>Pro Leu Leu Arg Ala Gly Arg Asp Leu Gly Leu Gly Pro Arg</td>
<td>Ala Pro Arg Gln Ala Arg Pro Leu His His Gly Arg His Asp Arg</td>
</tr>
<tr>
<td>50 55 60</td>
<td>65 70 75</td>
<td>85 90 95</td>
</tr>
<tr>
<td>Glu Pro Leu Val Leu Gln Asp His Pro Gin Gly Gly Pro Leu Val Cys</td>
<td>Gly Cys His His</td>
<td>100</td>
</tr>
</tbody>
</table>
<220> FEATURE:
<223> OTHER INFORMATION: selected peptide

<400> SEQUENCE: 16

Met Asp Glu Lys Thr His Trp Val Leu Gln Leu His Pro Gln Gly Asp
  1   5  10  15
Arg Leu Gly Pro Arg His Gly Gly Asp Asp Val Arg Leu Val Gly Gin
 20  25  30
Gly Glu Val Leu Glu Gly Leu Asp Arg Arg Arg Arg Arg Arg Arg Arg
 35  40  45
His Arg Leu Pro Arg Glu His Val Arg Ala Leu Val Ase
 50  55  60
Gln Val Arg Ase Leu Ala Gln Leu Val Glu Val Asp Gly Gly
 65  70  75  80
Val Glu Ala Leu Arg His Leu Gly Leu Pro Gin Asp Glu Pro Arg Ser
 85  90  95
Gly Gly Cys His His His
100

<210> SEQ ID NO 17
<211> LENGTH: 102
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: selected peptide

<400> SEQUENCE: 17

Met Asp Glu Lys Thr His Trp Val Gly Asp Leu Gln Gln Pro Leu Gly
  1   5  10  15
Pro Leu His Gly Val Gly Val Gly Leu Val Leu Arg
 20  25  30
His His Pro Gin Arg Ase Arg Leu Val Gly Val Gly Pro His Gly
 35  40  45
Arg Ala Leu Ala Arg Arg Pro His Arg Val Val Gln Leu His His
 50  55  60
Leu Leu Gin Arg Gly GLU Arg Leu Pro Ase Pro Ase Pro Arg Gin
 65  70  75  80
Leu Gly Leu Leu Gly Leu Asp Arg Ala Ase Pro Ala Leu Val
 85  90  95
Trp Arg Leu Pro Pro Ser
100

<210> SEQ ID NO 18
<211> LENGTH: 101
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: selected peptide

<400> SEQUENCE: 18

Met Asp Glu Lys Thr His Cys Ala Val Asn Val Asn Val Gly Leu Thr
  1   5  10  15
His Trp Cys His Arg Val Ala His Leu Gin Pro Leu Ase Pro His Pro
 20  25  30
Gln Gly Ase His Leu Arg Leu Gin Pro Leu Gin Leu His Ala Leu Val Ase
 35  40  45
Pro Leu Val Gin Gln Gly Val Val Arg Pro Leu Gin Leu Ase
 50  55  60
Val Gly Val Gln Arg Val Ala Leu Val Glu Gin Val Ala Gin Val Gly
65 70 75 80
Glu Gly Leu Asp His Glu Ala Gly Gin Ala His Gly Ala Leu Val Trp
85 90 95
Arg Leu Pro Pro Pro Ser
100

<210> SEQ ID NO 19
<211> LENGTH: 101
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: selected peptide
<400> SEQUENCE: 19
Met Asp Glu Lys Thr Thr Gly Trp Arg Gly Gly His Val Val Glu Gly
1 5 10 15
Leu Ala Gly Leu Glu Gin Leu Arg Ala Arg Leu Glu His His Pro
20 25 30
Gln Gly Gin Arg Glu Pro Leu Val Gln Glu Val Glu Aep Val Asp Glu
35 40 45
Gly Leu Val Gin Asp Leu His Gly Val Val Ala Gly Leu Leu Asp Pro
50 55 60
Val Glu Leu Leu Thr Asp Trp Phe Lys Lys Leu Asn Val Ser
65 70 75 80
Lys Asp Cys Lys Met Thr Phe Tyr Leu Glu Met Tyr Asp Trp Ser Gly
85 90 95
Gly Cys His His His
100

<210> SEQ ID NO 20
<211> LENGTH: 102
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: selected peptide
<400> SEQUENCE: 20
Met Asn Glu Lys Thr His Cys Leu Asn Phe Lys Val Asn Ile Ala
1 5 10 15
Asp Trp Leu Ala Glu Phe His Gly Gly Gin Gly Gin Leu Leu Gly Arg
20 25 30
Arg Asp Gly Val Val Gin Arg Leu Val Asp Gly Val Gin Glu Arg Val
35 40 45
Glu Arg Leu Aep Arg Asp Pro Gly Leu Gly Leu Arg Leu Gly Leu
50 55 60
His His Arg Asp His Arg Leu Arg Gin Gly Glu His Leu Arg
65 70 75 80
Asp His Pro Leu Glu Pro Asp His Leu Val Val Gly Gly Leu Val
85 90 95
Trp Arg Leu Pro Pro Ser
100

<210> SEQ ID NO 21
<211> LENGTH: 101
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: selected peptide
<400> SEQUENCE: 21
Met Asp Glu Lys Thr Thr Gly Trp Arg Gly Gly His Val Val Glu Gly
1 5 10 15
Leu Ala Gly Glu Leu Glu Gln Leu Arg Ala Arg Leu Glu His His Pro
20 25 30
Gln Gly Gln Arg Glu Pro Leu Val Gln Glu Val Glu Asp Val Asp Glu
35 40 45
Gly Leu Val Gln Asp Leu His Gly Val Val Ala Gly Leu Leu Asp Pro
50 55 60
Val Glu Lys Leu Leu Thr Asp Trp Phe Lys Lys Lys His Val Ser
65 70 75 80
Lys Asp Cys Lys Met Thr Phe Tyr Leu Glu Met Tyr Asp Trp Ser Gly
85 90 95 100

<210> SEQ ID NO 22
<211> LENGTH: 89
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: selected peptide

<400> SEQUENCE: 22
Met Asp Glu Lys Thr Thr Gly Trp Arg Gly Gly His Val Val Glu Gly
1 5 10 15
Leu Ala Gly Glu Leu Glu Gln Leu Arg Ala Arg Leu Glu His His Pro
20 25 30
Gln Gly Gln Arg Glu Pro Leu Val Gln Glu Val Glu Asp Val Asp Glu
35 40 45
Gly Leu Val Gln Asp Leu His Gly Val Val Ala Gly Leu Leu Asp Pro
50 55 60
Val Glu Lys Leu Leu Thr Asp Trp Phe Lys Lys Lys Asn Val Ser
65 70 75 80
Met Met Ser Gly Gly Cys Lys Leu Gly
85

<210> SEQ ID NO 23
<211> LENGTH: 75
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: selected peptide

<400> SEQUENCE: 23
Met Asp Glu Lys Thr Thr Gly Trp Arg Gly Gly His Val Val Glu Gly
1 5 10 15
Leu Ala Gly Glu Leu Glu Gln Leu Arg Ala Arg Leu Glu His His Pro
20 25 30
Gln Gly Gln Arg Glu Pro Leu Val Gln Glu Val Glu Asp Val Asp Glu
35 40 45
Gly Leu Val Gln Asp Leu His Gly Val Val Ala Gly Leu Leu Asp Pro
50 55 60
Val Glu Met Met Ser Gly Gly Cys Lys Leu Gly
65 70 75

<210> SEQ ID NO 24
<211> LENGTH: 61
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: selected peptide

<400> SEQUENCE: 24

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

<210> SEQ ID NO 25
<211> LENGTH: 47
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: selected peptide

<400> SEQUENCE: 25

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

<210> SEQ ID NO 26
<211> LENGTH: 47
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: selected peptide

<400> SEQUENCE: 26

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

<210> SEQ ID NO 27
<211> LENGTH: 39
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: selected peptide

<400> SEQUENCE: 27

Met Asp Gly His Val Val Glu Gly Leu Ala Gly Glu Glu Gln Leu
 1  5  10  15
Arg Ala Arg Leu Glu His His Pro Gln Gly Gln Arg Glu Pro Met Met
 20  25  30
Ser Gly Gly Cys Lys Leu Gly
 35

<210> SEQ ID NO 28
<211> LENGTH: 89
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: selected peptide

<400> SEQUENCE:

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

<210> SEQ ID NO 29
<211> LENGTH: 29
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: selected peptide

<400> SEQUENCE:

Met Glu Leu Glu Gln Leu Arg Ala Arg Leu Glu His His Pro Gln Gly 1 5 10 15
Gln Arg Glu Pro Met Ser Gly Gly Cys Lys Leu Gly 20 25

<210> SEQ ID NO 30
<211> LENGTH: 35
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: primer

<400> SEQUENCE: 30

atagcgytg cccagttgc agcgcgccaga ccagt 35

<210> SEQ ID NO 31
<211> LENGTH: 35
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: primer

<400> SEQUENCE: 31

acuggcugg cggcugcaag cuuggcaccg gcuau 35

<210> SEQ ID NO 32
<211> LENGTH: 14
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: selected peptide

<400> SEQUENCE: 32

Trp Ser Gly Gly Cys His His His His Ser Ser Ala 1 5 10
<210> SEQ ID NO 33
<211> LENGTH: 11
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: designed peptide

<400> SEQUENCE: 33
Trp Ser Gly Cys Lys Leu Gly Thr Gly Tyr
 1 5 10

<210> SEQ ID NO 34
<211> LENGTH: 44
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: designed peptide

<400> SEQUENCE: 34
Met Asp Glu Ala His Pro Gln Ala Gly Pro Val Asp Gln Ala Asp Ala
 1 5 10 15
Arg Leu Val Gln Gly Ala Leu Gln His His Pro Gln Gly Asp Arg
20 25 30
Met Met Ser Gly Cys Lys Leu Gly Thr Gly Tyr
35 40

<210> SEQ ID NO 35
<211> LENGTH: 38
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: selected peptide

<400> SEQUENCE: 35
Met Asp Glu Lys Thr Thr Gly Trp Arg Gly Gly His Val Val Glu Gly
 1 5 10 15
Leu Ala Gly Glu Leu Gln Leu Leu Arg Ala Arg Leu Glu His His Pro
20 25 30
Gln Gly Gln Arg Glu Pro
35

<210> SEQ ID NO 36
<211> LENGTH: 9
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: designed peptide

<400> SEQUENCE: 36
Met Met Ser Gly Cys Lys Leu Gly
 1 5

<210> SEQ ID NO 37
<211> LENGTH: 6688
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: plasmid

<400> SEQUENCE: 37
atccggatat agttcctcct ttcagcaaaa aacccctcaa gacccgttta gaggccccaa 60
ggggttatgc tagttattgc tcagcggtgg cagcagccaa ctcagcttcc tttcgggctt 120
tttgtagcag cggatctcag tggtggtggt ggtggtgctc gagtgcggcc gagcagctttt 180
agtcgtcatt ctcagcagtg gtgtggtacct ggcggcagc 240
taccagcttg tttgaggttc ttcagcctag cggccctttt 300
tgcatttccc acmaaggtgg cttggggtcct gcggatgttc 360
gggcgcgct cctggtccag cgacagtgac acgcgtcttg 420
gtccagctg agttttgaaa aacagaaaag gcgctttggt 480
cggcgtgttc tcccggtctc gctggtttgt cttcgtgttt 540
catacagttctggcctgcc cggggcgctt gccggtgttt 600
agcagcaggtt cggccctttt ggcgggcttg cggatgtttt 660
cggcgtgttc tcccggtctc gctggtttgt cttcgtgttt 720
cgggctttgt ctcagcagtg gtgtggtacct ggcggcagc 780
acccacaggg gacagagcgc ggttcgcagc cggccctttt 840
agcagcaggtt cggccctttt ggcgggcttg cggatgtttt 900
cggcgtgttc tcccggtctc gctggtttgt cttcgtgttt 960
tatttcagctt ctcagcagtg gtgtggtacct ggcggcagc 1020
tcgccagcc cggccctttt ggcgggcttg cggatgtttt 1080
tgtttaatgt ctcagcagtg gtgtggtacct ggcggcagc 1140
gggcgcggct ctcagcagtg gtgtggtacct ggcggcagc 1200
gtccaggttc ttcagcagtg gtgtggtacct ggcggcagc 1260
taatagctct ctcagcagtg gtgtggtacct ggcggcagc 1320
gggcgcggct ctcagcagtg gtgtggtacct ggcggcagc 1380
taatagctct ctcagcagtg gtgtggtacct ggcggcagc 1440
taatagctct ctcagcagtg gtgtggtacct ggcggcagc 1500
taatagctct ctcagcagtg gtgtggtacct ggcggcagc 1560
taatagctct ctcagcagtg gtgtggtacct ggcggcagc 1620
taatagctct ctcagcagtg gtgtggtacct ggcggcagc 1680
taatagctct ctcagcagtg gtgtggtacct ggcggcagc 1740
taatagctct ctcagcagtg gtgtggtacct ggcggcagc 1800
taatagctct ctcagcagtg gtgtggtacct ggcggcagc 1860
taatagctct ctcagcagtg gtgtggtacct ggcggcagc 1920
taatagctct ctcagcagtg gtgtggtacct ggcggcagc 1980
taatagctct ctcagcagtg gtgtggtacct ggcggcagc 2040
taatagctct ctcagcagtg gtgtggtacct ggcggcagc 2100
taatagctct ctcagcagtg gtgtggtacct ggcggcagc 2160
taatagctct ctcagcagtg gtgtggtacct ggcggcagc 2220
taatagctct ctcagcagtg gtgtggtacct ggcggcagc 2280
taatagctct ctcagcagtg gtgtggtacct ggcggcagc 2340
taatagctct ctcagcagtg gtgtggtacct ggcggcagc 2400
taatagctct ctcagcagtg gtgtggtacct ggcggcagc 2460
taatagctct ctcagcagtg gtgtggtacct ggcggcagc 2520
taatagctct ctcagcagtg gtgtggtacct ggcggcagc 2580
gaagacgta cggaactggg cgtaagcact cggccagcat cggtaacca cgcacatcg 2640
cgtttagcgg gcgcctaaag tctgtctcgc gcgcgtctgc gcgaagcag 2700
taccccac gcaacaaatg ccagcagcag cgcagcagt ggcggtcgcc 2760
tcgggtttc aaacaccaact gcgcagcactg agtcagggca tcgggttcatc 2820
gtggcaacag atagcagagg gtcgccttc gcctggctgg ctggcataaa 2880
gtgggtgcgg atactcgcgg gcgcatacag gcgcatacag atgtagacta 2940
cgccgcttaa ccccccctct ccccccctct ccccccctct 3000
tcgcgtcacc tctctccaggg ccggcgtcacc cggcgtcacc 3060
tggaanaa gaaaccaccc ggcggcccagg cgcgtctctg cctctttct 3120
gattccattc tcagcacggt cggcggtcacc cgcgtctctg 3180
cgcctacatt gcgcagcttc gcgcgtctct cggcgtctctg 3240
tgctcagtcgc gcgccggagg gcgcgtctct cggcgtctctg 3300
tatgtctgct tcccttttcgg cccacccctct cccacccctct 3360
ttcccttttc ggccccctct cgcgccggagg gcgcgtctctg 3420
cgcctctccg cccacccctct cccacccctct cccacccctct 3480
cgcgcgtctct cggcgtctctg gcgcgtctctg gcgcgtctctg 3540
ttcgcgcag tgcggagcgc gtcgccttc ggccgggtgc 3600
gtcgcgcag tgcggagcgc gtcgccttc ggccgggtgc 3660
gggcgcag tgcggagcgc gtcgccttc ggccgggtgc 3720
gggcgcag tgcggagcgc gtcgccttc ggccgggtgc 3780
gggcgcag tgcggagcgc gtcgccttc ggccgggtgc 3840
gggcgcag tgcggagcgc gtcgccttc ggccgggtgc 3900
gggcgcag tgcggagcgc gtcgccttc ggccgggtgc 3960
tcccacattcg ggcgggcagc ggcggggcag cgcgtctctg 4020
ttgctcggag gcgcagcttc ggcgggcagc ggcggggcag cgcgtctctg 4080
ggggagcag atcagcagc ggcgcggccc gcgcgtctctg gcgcgtctctg 4140
ggggagcag atcagcagc ggcgcggccc gcgcgtctctg gcgcgtctctg 4200
ggggagcag atcagcagc ggcgcggccc gcgcgtctctg gcgcgtctctg 4260
ggggagcag atcagcagc ggcgcggccc gcgcgtctctg gcgcgtctctg 4320
ggggagcag atcagcagc ggcgcggccc gcgcgtctctg gcgcgtctctg 4380
ggggagcag atcagcagc ggcgcggccc gcgcgtctctg gcgcgtctctg 4440
ggggagcag atcagcagc ggcgcggccc gcgcgtctctg gcgcgtctctg 4500
ggggagcag atcagcagc ggcgcggccc gcgcgtctctg gcgcgtctctg 4560
ggggagcag atcagcagc ggcgcggccc gcgcgtctctg gcgcgtctctg 4620
ggggagcag atcagcagc ggcgcggccc gcgcgtctctg gcgcgtctctg 4680
ggggagcag atcagcagc ggcgcggccc gcgcgtctctg gcgcgtctctg 4740
ggggagcag atcagcagc ggcgcggccc gcgcgtctctg gcgcgtctctg 4800
ggggagcag atcagcagc ggcgcggccc gcgcgtctctg gcgcgtctctg 4860
ggggagcag atcagcagc ggcgcggccc gcgcgtctctg gcgcgtctctg 4920
**US 6,841,359 B2**

**-continued**

```
acccgtaaag acacgactta tcgccactgg cagcagccac tggtaacagg attagcagag
cgaggtatgt aggcggtgct acagagttct tgaagtggtg gcctaactac ggctacacta
gaaggacagt atttggtatc tgcgctctgc tgaagccagt taccttcgga aaaagagttg
gtagctcttg atccggcaaa caaaccaccg ctggtagcgg tggttttttt gtttgcaagc
gagcagattac gcgcagaaaa aaaggatctc aagaagatcc tttgatcttt tctacggggt
ctgacgctca gtggaacgaa aactcacgtt aagggatttt ggtcatgaac aataaaactg
tctgcttaca taaacagtaa tacaaggggt gttatgagcc atattcaacg ggaaacgtct
ctgctctaggc cgcgattaaa ttccaacatg gatgctgatt tatatgggta taaatgggct
tcgggcaatc aggtgcgaca atctatcgat tgtatgggaa gcccgatgcg ccagagttgt
cttgaaaca tggcaaaggt agcgttgcca atgatgttac agatgagatg gtcagactaa
actggctgac ggaatttatg cctcttccga ccatcaagca ttttatccgt actcctgatg
atgcatggtt actcaccact gcgatccccg ggaaaacagc attccaggta ttagaagaat
atcctgattc aggtgaaaat attgttgatg cgctggcagt gttcctgcgc cggttgcatt
cgattcctgt ttgtaattgt ccttttaaca gcgatcgcgt atttcgtctc gcctaggcgc
aatcacgaat gaataacggt ttggttgatg cgagtgattt tgatgacgag cgtaatggct
ggctgttga acaagtctgg aaagaaatgc ataaactttt gccattctca ccggattcag
tcgtcactca tggtgatttc tcacttgata accttatttt tgacgagggg aaattaatag
gttgtattga tgttggacga gtcggaatcg cagaccgata ccaggatctt gccatcctat
agaactgcct cggtgagttt tctccttcat tacagaaacg gctttttcaa aaatatggta
ttgataatcc tgatatgaat aaattgcagt ttcatttgat gctcgatgag tttttctaag
aattaattca tgagcggata catatttgaa tgtatttaga aaaataaaca ataggggtt
ccgcgcacat ttccccgaaa agtgccacct gaaattgtaa acgttaatat ttttgttaaa
ttcgcgttaa atttttgtta aatcagctca ttttttaacc aataggccga atccgcaaa
atcccttata aatccaaaga atagacccag atagggttga gtggtggtcc aagttggaac
agsagtaacc tagtaactag ccagacctcc cagctcaagg ggcaaaaaa acgtcttagg
catcagctg cactactgta acatcacccts tataaagaagc tgatactacc cggttgggtc
cgtgtgtgctt ctgctgatgc tggctgattt gagaataggt cagtgtagtg ccaggggag
ccggagggaa ccggagggaa ccggagggaa ccggagggaa ccggagggaa ccggagggaa
cggagggaa ccggagggaa ccggagggaa ccggagggaa ccggagggaa ccggagggaa
```

**SEQ ID NO 38**

**LENGTH:** 479

**ORIGIN:** Artificial Sequence

**DESCRIPTION:** Designed protein

**SEQUENCE:** Met Gly Ile Glu Glu Gly Lys Leu Val Ile Trp Ile Asn Gly Asp Lys 1 5 10 15
Gly Tyr Asn Gly Leu Ala Glu Val Gly Lys Phe Glu Lys Asp Thr 20 25 30
Gly Ile Lys Val Thr Val Glu His Pro Asp Lys Leu Glu Gly Lys Phe 35 40 45
<table>
<thead>
<tr>
<th>Pro</th>
<th>Gln</th>
<th>Val</th>
<th>Ala</th>
<th>Ala</th>
<th>Thr</th>
<th>Gly</th>
<th>Asp</th>
<th>Gly</th>
<th>Pro</th>
<th>Asp</th>
<th>Ile</th>
<th>Ile</th>
<th>Phe</th>
<th>Trp</th>
<th>Ala</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td></td>
<td></td>
<td></td>
<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>His</td>
<td>Asp</td>
<td>Arg</td>
<td>Phe</td>
<td>Gly</td>
<td>Gly</td>
<td>Tyr</td>
<td>Ala</td>
<td>Gln</td>
<td>Ser</td>
<td>Gln</td>
<td>Gly</td>
<td>Leu</td>
<td>Leu</td>
<td>Ala</td>
<td>Glu</td>
</tr>
<tr>
<td>65</td>
<td></td>
<td></td>
<td></td>
<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>Thr</td>
<td>Pro</td>
<td>Asp</td>
<td>Lys</td>
<td>Ala</td>
<td>Phe</td>
<td>Gln</td>
<td>Asp</td>
<td>Lys</td>
<td>Leu</td>
<td>Tyr</td>
<td>Pro</td>
<td>Phe</td>
<td>Thr</td>
<td>Trp</td>
<td>Asp</td>
</tr>
<tr>
<td>85</td>
<td></td>
<td></td>
<td></td>
<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>Ala</td>
<td>Val</td>
<td>Arg</td>
<td>Tyr</td>
<td>Asn</td>
<td>Gly</td>
<td>Lys</td>
<td>Leu</td>
<td>Ile</td>
<td>Ala</td>
<td>Tyr</td>
<td>Pro</td>
<td>Ile</td>
<td>Ala</td>
<td>Val</td>
<td>Glu</td>
</tr>
<tr>
<td>100</td>
<td></td>
<td></td>
<td></td>
<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>Ala</td>
<td>Leu</td>
<td>Ser</td>
<td>Leu</td>
<td>Ile</td>
<td>Tyr</td>
<td>Asn</td>
<td>Lys</td>
<td>Asp</td>
<td>Leu</td>
<td>Leu</td>
<td>Pro</td>
<td>Pro</td>
<td>Pro</td>
<td>Pro</td>
<td>Lys</td>
</tr>
<tr>
<td>115</td>
<td></td>
<td></td>
<td></td>
<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>Thr</td>
<td>Trp</td>
<td>Glu</td>
<td>Glu</td>
<td>Ile</td>
<td>Pro</td>
<td>Ala</td>
<td>Leu</td>
<td>Asp</td>
<td>Lys</td>
<td>Leu</td>
<td>Lys</td>
<td>Lys</td>
<td>Ala</td>
<td>Lys</td>
<td>Gly</td>
</tr>
<tr>
<td>139</td>
<td></td>
<td></td>
<td></td>
<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>Lys</td>
<td>Ser</td>
<td>Ala</td>
<td>Leu</td>
<td>Met</td>
<td>Phe</td>
<td>Asn</td>
<td>Leu</td>
<td>Gln</td>
<td>Glu</td>
<td>Pro</td>
<td>Tyr</td>
<td>Phe</td>
<td>Thr</td>
<td>Trp</td>
<td>Pro</td>
</tr>
<tr>
<td>145</td>
<td></td>
<td></td>
<td></td>
<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>Leu</td>
<td>Ile</td>
<td>Ala</td>
<td>Ala</td>
<td>Asp</td>
<td>Gly</td>
<td>Gly</td>
<td>Tyr</td>
<td>Ala</td>
<td>Phe</td>
<td>Lys</td>
<td>Tyr</td>
<td>Glu</td>
<td>Asn</td>
<td>Gly</td>
<td>Lys</td>
</tr>
<tr>
<td>165</td>
<td></td>
<td></td>
<td></td>
<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>Tyr</td>
<td>Asp</td>
<td>Ile</td>
<td>Lys</td>
<td>Asp</td>
<td>Val</td>
<td>Gly</td>
<td>Val</td>
<td>Asp</td>
<td>Asn</td>
<td>Ala</td>
<td>Gly</td>
<td>Ala</td>
<td>Lys</td>
<td>Ala</td>
<td>Gly</td>
</tr>
<tr>
<td>180</td>
<td></td>
<td></td>
<td></td>
<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>Leu</td>
<td>Thr</td>
<td>Phe</td>
<td>Leu</td>
<td>Val</td>
<td>Asp</td>
<td>Leu</td>
<td>Ile</td>
<td>Lys</td>
<td>Asn</td>
<td>Lys</td>
<td>His</td>
<td>Met</td>
<td>Asn</td>
<td>Ala</td>
<td>Asp</td>
</tr>
<tr>
<td>195</td>
<td></td>
<td></td>
<td></td>
<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>Thr</td>
<td>Asp</td>
<td>Tyr</td>
<td>Ser</td>
<td>Ile</td>
<td>Ala</td>
<td>Glu</td>
<td>Ala</td>
<td>Asp</td>
<td>Asn</td>
<td>Lys</td>
<td>Gly</td>
<td>Glu</td>
<td>Gly</td>
<td>Thr</td>
<td>Ala</td>
</tr>
<tr>
<td>210</td>
<td></td>
<td></td>
<td></td>
<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>Met</td>
<td>Thr</td>
<td>Ile</td>
<td>Asn</td>
<td>Gly</td>
<td>Pro</td>
<td>Trp</td>
<td>Ala</td>
<td>Trp</td>
<td>Ser</td>
<td>Asn</td>
<td>Ile</td>
<td>Asp</td>
<td>Thr</td>
<td>Ser</td>
<td>Lys</td>
</tr>
<tr>
<td>225</td>
<td></td>
<td></td>
<td></td>
<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>Val</td>
<td>Aan</td>
<td>Tyr</td>
<td>Gly</td>
<td>Val</td>
<td>Thr</td>
<td>Val</td>
<td>Leu</td>
<td>Pro</td>
<td>Thr</td>
<td>Phe</td>
<td>Lys</td>
<td>Gly</td>
<td>Glu</td>
<td>Pro</td>
<td>Ser</td>
</tr>
<tr>
<td>245</td>
<td></td>
<td></td>
<td></td>
<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>Lys</td>
<td>Pro</td>
<td>Phe</td>
<td>Val</td>
<td>Gly</td>
<td>Val</td>
<td>Leu</td>
<td>Ser</td>
<td>Ala</td>
<td>Gly</td>
<td>Ile</td>
<td>Asn</td>
<td>Ala</td>
<td>Ala</td>
<td>Ser</td>
<td>Pro</td>
</tr>
<tr>
<td>260</td>
<td></td>
<td></td>
<td></td>
<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>Asn</td>
<td>Lys</td>
<td>Glu</td>
<td>Leu</td>
<td>Ala</td>
<td>Ala</td>
<td>Lys</td>
<td>Glu</td>
<td>Phe</td>
<td>Leu</td>
<td>Glu</td>
<td>Asn</td>
<td>Tyr</td>
<td>Leu</td>
<td>Leu</td>
<td>Thr</td>
</tr>
<tr>
<td>275</td>
<td></td>
<td></td>
<td></td>
<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>Glu</td>
<td>Gly</td>
<td>Leu</td>
<td>Glu</td>
<td>Ala</td>
<td>Val</td>
<td>Asn</td>
<td>Lys</td>
<td>Asp</td>
<td>Pro</td>
<td>Leu</td>
<td>Gly</td>
<td>Val</td>
<td>Ala</td>
<td>Val</td>
<td>Ala</td>
</tr>
<tr>
<td>290</td>
<td></td>
<td></td>
<td></td>
<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>Leu</td>
<td>Lys</td>
<td>Ser</td>
<td>Tyr</td>
<td>Gly</td>
<td>Leu</td>
<td>Glu</td>
<td>Ala</td>
<td>Lys</td>
<td>Asp</td>
<td>Pro</td>
<td>Arg</td>
<td>Ile</td>
<td>Ala</td>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>305</td>
<td></td>
<td></td>
<td></td>
<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>Thr</td>
<td>Met</td>
<td>Glu</td>
<td>Aes</td>
<td>Leu</td>
<td>Ala</td>
<td>Gly</td>
<td>Glu</td>
<td>Ile</td>
<td>Met</td>
<td>Pro</td>
<td>Aen</td>
<td>Ile</td>
<td>Pro</td>
<td>Glu</td>
<td></td>
</tr>
<tr>
<td>325</td>
<td></td>
<td></td>
<td></td>
<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>Met</td>
<td>Ser</td>
<td>Ala</td>
<td>Phe</td>
<td>Trp</td>
<td>Tyr</td>
<td>Ala</td>
<td>Val</td>
<td>Arg</td>
<td>Thr</td>
<td>Ala</td>
<td>Val</td>
<td>Ile</td>
<td>Asn</td>
<td>Ala</td>
<td>Ala</td>
</tr>
<tr>
<td>340</td>
<td></td>
<td></td>
<td></td>
<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>Ser</td>
<td>Gly</td>
<td>Arg</td>
<td>Gln</td>
<td>Thr</td>
<td>Val</td>
<td>Asp</td>
<td>Glu</td>
<td>Ala</td>
<td>Leu</td>
<td>Lys</td>
<td>Asp</td>
<td>Ala</td>
<td>Gln</td>
<td>Thr</td>
<td>Asn</td>
</tr>
<tr>
<td>355</td>
<td></td>
<td></td>
<td></td>
<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>Ser</td>
<td>Ser</td>
<td>Ser</td>
<td>Gly</td>
<td>Gly</td>
<td>Ser</td>
<td>Gly</td>
<td>Ser</td>
<td>Gly</td>
<td>Met</td>
<td>Asp</td>
<td>Glu</td>
<td>Lys</td>
<td>Thr</td>
<td>Thr</td>
<td>Gly</td>
</tr>
<tr>
<td>370</td>
<td></td>
<td></td>
<td></td>
<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>Trp</td>
<td>Arg</td>
<td>Gly</td>
<td>Gly</td>
<td>His</td>
<td>Val</td>
<td>Val</td>
<td>Glu</td>
<td>Gly</td>
<td>Leu</td>
<td>Ala</td>
<td>Gly</td>
<td>Leu</td>
<td>Glu</td>
<td>Glu</td>
<td></td>
</tr>
<tr>
<td>385</td>
<td></td>
<td></td>
<td></td>
<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>Leu</td>
<td>Arg</td>
<td>Ala</td>
<td>Arg</td>
<td>Leu</td>
<td>Glu</td>
<td>His</td>
<td>Pro</td>
<td>Gln</td>
<td>Gly</td>
<td>Glu</td>
<td>Arg</td>
<td>Glu</td>
<td>Pro</td>
<td>Gly</td>
<td></td>
</tr>
<tr>
<td>405</td>
<td></td>
<td></td>
<td></td>
<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>Ser</td>
<td>Gly</td>
<td>His</td>
<td>His</td>
<td>His</td>
<td>His</td>
<td>His</td>
<td>Glu</td>
<td>Phe</td>
<td>Leu</td>
<td>Val</td>
<td>Pro</td>
<td>Arg</td>
<td>Gly</td>
<td>Ser</td>
<td></td>
</tr>
<tr>
<td>420</td>
<td></td>
<td></td>
<td></td>
<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>Met</td>
<td>Asp</td>
<td>Pro</td>
<td>Cys</td>
<td>Val</td>
<td>Lys</td>
<td>Cys</td>
<td>Val</td>
<td>Ala</td>
<td>Pro</td>
<td>Arg</td>
<td>Aen</td>
<td>Lys</td>
<td>Thr</td>
<td>Lys</td>
<td>Ser</td>
</tr>
<tr>
<td>435</td>
<td></td>
<td></td>
<td></td>
<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>Lys</td>
<td>Aan</td>
<td>Lys</td>
<td>His</td>
<td>Leu</td>
<td>Arg</td>
<td>Ile</td>
<td>Tyr</td>
<td>Aen</td>
<td>Met</td>
<td>Cys</td>
<td>Lys</td>
<td>Thr</td>
<td>Cys</td>
<td>Phe</td>
<td>Asn</td>
</tr>
<tr>
<td>450</td>
<td></td>
<td></td>
<td></td>
<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>Asn</td>
<td>Ser</td>
<td>Ile</td>
<td>Asp</td>
<td>Ile</td>
<td>Gly</td>
<td>Asp</td>
<td>Asp</td>
<td>Thr</td>
<td>Tyr</td>
<td>His</td>
<td>Gly</td>
<td>His</td>
<td>Asp</td>
<td>Asp</td>
<td></td>
</tr>
</tbody>
</table>
<210> SEQ ID NO 39
<211> LENGTH: 47
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Selected peptide

<400> SEQUENCE: 39

Met Asp Pro Cys Val Lys Cys Val Ala Pro Arg Aen Trp Lys Val
1  5 10 15
Lys Aen Lys His Leu Arg Ile Tyr Aen Met Cys Lys Thr Cys Phe Aen
20 25 30
Asn Ser Ile Asp Ile Gly Asp Asp Thr Tyr His Gly His Asp Aen
35 40 45

<210> SEQ ID NO 40
<211> LENGTH: 6544
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: plasmid

<400> SEQUENCE: 40

atccggatat agttcctcct ttcagcaaaa aacccctcaa gacccgttta gaggccccaa
60
gggtttattg tattattgc tggcctgctg cacaccctca ctcagcincc tttcggcct
120
tgttaagacgc ggcatactgc tcgtgctgct gggctggtgc gcaagctgct gcattcctg
180
gctgagcatg tttagcttct tgtcctggtg ggtgctgctt ggcagtctgct ggttcgctg
240
gacctcgggt gcgcctgctg gcgcgctgctg gcgcgctgctg gcgcgctgctg gcgcgctgctg
300
cataacgctg ctgcgtgcgtg cgcgtgcgtg gcgcgtgcgtg gcgcgtgcgtg gcgcgtgcgtg
360
cgaagctgc aatagctgc aatagctgc aatagctgc aatagctgc aatagctgc aatagctgc
420
cgaagctgc aatagctgc aatagctgc aatagctgc aatagctgc aatagctgc aatagctgc
480
cgaagctgc aatagctgc aatagctgc aatagctgc aatagctgc aatagctgc aatagctgc
540
cgaagctgc aatagctgc aatagctgc aatagctgc aatagctgc aatagctgc aatagctgc
600
cgaagctgc aatagctgc aatagctgc aatagctgc aatagctgc aatagctgc aatagctgc
660
gacgcaagc gtcctgctg gcgcgtgcgtg gcgcgtgcgtg gcgcgtgcgtg gcgcgtgcgtg
720
cgaagctgc aatagctgc aatagctgc aatagctgc aatagctgc aatagctgc aatagctgc
780
cgaagctgc aatagctgc aatagctgc aatagctgc aatagctgc aatagctgc aatagctgc
840
cgaagctgc aatagctgc aatagctgc aatagctgc aatagctgc aatagctgc aatagctgc
900
cgaagctgc aatagctgc aatagctgc aatagctgc aatagctgc aatagctgc aatagctgc
960
cgaagctgc aatagctgc aatagctgc aatagctgc aatagctgc aatagctgc aatagctgc
1020
cgaagctgc aatagctgc aatagctgc aatagctgc aatagctgc aatagctgc aatagctgc
1080
cgaagctgc aatagctgc aatagctgc aatagctgc aatagctgc aatagctgc aatagctgc
1140
cgaagctgc aatagctgc aatagctgc aatagctgc aatagctgc aatagctgc aatagctgc
1200
cgaagctgc aatagctgc aatagctgc aatagctgc aatagctgc aatagctgc aatagctgc
1260
cgaagctgc aatagctgc aatagctgc aatagctgc aatagctgc aatagctgc aatagctgc
1320
cgaagctgc aatagctgc aatagctgc aatagctgc aatagctgc aatagctgc aatagctgc
1380
cgaagctgc aatagctgc aatagctgc aatagctgc aatagctgc aatagctgc aatagctgc
1440
tccagattac cgtttacct tcctcgattc ccagttcataa ctttcgattc ccatggtata tctccttctt aaagttaaac
aaaattattt ctagagggga attgttatcc gctcacaatt cccctatagt gagtcgtatt
aatttcgcgg gatcgagatc tcgatcctct acgccggacg catcgtggcc ggcatcaccg
gccgccacagg tgcggttgct ggcgcctata tcgccgacat caccgatggg gaagatcggg
cctgcgactt cgggctcatg agcgcttgtt tcggcgtggg tatggtggca ggccccgtgg
cgggggact gttgggcgcc atctccttgc atgcaccatt ccttgcggcg gcggtgctca
acggcctcaa cctactactg ggctgcttcc taatgcagga gtcgcataag ggagagctc
gagatoccyg aacccatcya atgygcaaaa accctcqcgc gatgygcatg atagocccg
gaagagagtc aattcgcgg gcggagctcc aaccgcgtta ctttgcgcgct tgcggttcat
ccgaaaagat cgggagagct cagcaggttc ctagccttcg cctgttcgct cgttccgtgt
tcatggtcgg cctgggtctt ctagccttcg cttcctgctgg cggagagctg cttcatgttc
tcgcgttcgt cttctgccgg cttggctggtt tccgacatga agagttctgt cggagagctg
tcggggtgtt ctctgcgttg ctgtggtcgt cctggtggtt ttcgaggtact gccggtggtt
tgatggcggg acgcagagtc atagcgtcgg cggagagctc attctggctg tttcggaggt
cctgggtctt ctagccttcg cttcctgctgg cggagagctg cttcatgttc
ccgaaaagat cgggagagct cagcaggttc ctagccttcg cctgttcgct cgttccgtgt
tcatggtcgg cctgggtctt ctagccttcg cttcctgctgg cggagagctg cttcatgttc
tcgcgttcgt cttctgccgg cttggctggtt tccgacatga agagttctgt cggagagctg
tcggggtgtt ctctgcgttg ctgtggtcgt cctggtggtt ttcgaggtact gccggtggtt
tgatggcggg acgcagagtc atagcgtcgg cggagagctc attctggctg tttcggaggt
cctgggtctt ctagccttcg cttcctgctgg cggagagctg cttcatgttc
ccgaaaagat cgggagagct cagcaggttc ctagccttcg cctgttcgct cgttccgtgt
tcatggtcgg cctgggtctt ctagccttcg cttcctgctgg cggagagctg cttcatgttc
tcgcgttcgt cttctgccgg cttggctggtt tccgacatga agagttctgt cggagagctg
tcggggtgtt ctctgcgttg ctgtggtcgt cctggtggtt ttcgaggtact gccggtggtt
tgatggcggg acgcagagtc atagcgtcgg cggagagctc attctggctg tttcggaggt
cctgggtctt ctagccttcg cttcctgctgg cggagagctg cttcatgttc
ccgaaaagat cgggagagct cagcaggttc ctagccttcg cctgttcgct cgttccgtgt
tcatggtcgg cctgggtctt ctagccttcg cttcctgctgg cggagagctg cttcatgttc
tcgcgttcgt cttctgccgg cttggctggtt tccgacatga agagttctgt cggagagctg

tccagattac cgtttacct tcctcgattc ccagttcataa ctttcgattc ccatggtata tctccttctt aaagttaaac
aaaattattt ctagagggga attgttatcc gctcacaatt cccctatagt gagtcgtatt
aatttcgcgg gatcgagatc tcgatcctct acgccggacg catcgtggcc ggcatcaccg
gccgccacagg tgcggttgct ggcgcctata tcgccgacat caccgatggg gaagatcggg
cctgcgactt cgggctcatg agcgcttgtt tcggcgtggg tatggtggca ggccccgtgg
ccgggggact gttgggcgcc atctccttgc atgcaccatt ccttgcggcg gcggtgctca
acggcctcaa cctactactg ggctgcttcc taatgcagga gtcgcataag ggagagctc
gagatoccyg aacccatcya atgygcaaaa accctcqcgc gatgygcatg atagocccg
gaagagagtc aattcgcgg gcggagctcc aaccgcgtta ctttgcgcgct tgcggttcat
ccgaaaagat cgggagagct cagcaggttc ctagccttcg cctgttcgct cgttccgtgt
tcatggtcgg cctgggtctt ctagccttcg cttcctgctgg cggagagctg cttcatgttc
acggcctcaa cctactactg ggctgcttcc taatgcagga gtcgcataag ggagagctc
gagatoccyg aacccatcya atgygcaaaa accctcqcgc gatgygcatg atagocccg
gaagagagtc aattcgcgg gcggagctcc aaccgcgtta ctttgcgcgct tgcggttcat
ccgaaaagat cgggagagct cagcaggttc ctagccttcg cctgttcgct cgttccgtgt
tcatggtcgg cctgggtctt ctagccttcg cttcctgctgg cggagagctg cttcatgttc
acggcctcaa cctactactg ggctgcttcc taatgcagga gtcgcataag ggagagctc
gagatoccyg aacccatcya atgygcaaaa accctcqcgc gatgygcatg atagocccg

US 6,841,359 B2

-continued
gacctaacg gaaaaaacg cctttttcagt gccccgttttg ttcgagacgct ggaatcaatttg 3900
ggctctgccg acactcagcg agcctgaaggc ggctgacggtg ggtggtgatcg 3960
ttcgacagc gctccgagcg gctcttggct cgtgcatctgt gctttcggcgct 4020
aacattcagt cggcgcaatcg cgcggagctca gctccgatggc 4080
gacctgacg gctctgtaaga cgtccgagctc ggaactacgct 4140
gacggagac gctctgtaaga cgtccgagctc ggaactacgct 4200
attgtaactg gcctgtatttt cgtcgtcctag gtctgcttgc 4260
aatcaccgac gcgggcttttz cgtcgagctc ggaactacgct 4320
agcctgggtcg gctctgtaaga cgtccgagctc ggaactacgct 4380
gggtagacg gacgttgaaac gcggttaagct cggagatcgtc 4440
aggtcctggcg gcatccgagctc ggaactacgct 4500
gcgtcgtctag cgcggagctc ggaactacgct 4560
tgtgtaaactg gcctgtatttt cgtcgtcctag 4620
ttttctcttt tccgttgatgg gcgtcctatttt 4680
cggagcttcg gctctgtaaga cgtccgagctc ggaactacgct 4740
getcctggcg gctctgtaaga cgtccgagctc ggaactacgct 4800
cactcggagac gacgttgaaac gcggttaagct cggagatcgtc 4860
agtccgttggcg gctctgtaaga cgtccgagctc ggaactacgct 4920
cgtcctggcg gctctgtaaga cgtccgagctc ggaactacgct 4980
eacoagctgg tagcggagctc ggaactacgct 5040
gactatcggagac gacgttgaaac gcggttaagct cggagatcgtc 5100
eactcggagac gacgttgaaac gcggttaagct cggagatcgtc 5160
aggggtggtta tggacgccactc gtttgcgtttg ttcgagacgct ggaatcaatttg 5220
aacattcagt cggcgcaatcg cgcggagctca gctccgatggc 5280
gccccatctc actacggagc gctctgtaaga cgtccgagctc ggaactacgct 5340
aanacctagg cccccgtttg ttgacgccactc gtttgcgtttg ttcgagacgct ggaatcaatttg 5400
aacattcagt cggcgcaatcg cgcggagctca gctccgatggc 5460
acccggcgacg gcctgtatttt cgtcgtcctag 5520
gaaatattg tttgctgctc gcgggatcgtc gctctgtaaga cgtccgagctc ggaactacgct 5580
aatgtaactg gcctgtatttt cgtcgtcctag 5640
aacggtttgg cggctgctc gcgggatcgtc gctctgtaaga cgtccgagctc ggaactacgct 5700
gtctgtagac gctctgtaaga cgtccgagctc ggaactacgct 5760
gactatcggagac gacgttgaaac gcggttaagct cggagatcgtc 5820
aacattcagt cggcgcaatcg cgcggagctca gctccgatggc 5880
gactatcggagac gacgttgaaac gcggttaagct cggagatcgtc 5940
actatcggagac gacgttgaaac gcggttaagct cggagatcgtc 6000
cctcctgtaaactc gtttgcgtttg ttcgagacgct ggaatcaatttg 6060
cgagccactc gtttgcgtttg ttcgagacgct ggaatcaatttg 6120
tttgtaacctg gccggatcgtc gctctgtaaga cgtccgagctc ggaactacgct 6180
<210> SEQ ID NO 41
<211> LENGTH: 426
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: designed protein

<400> SEQUENCE: 41

Met Gly Ile Glu Glu Gly Lys Leu Val Ile Asp Lys 1  5 10  15
Gly Tyr Asn Gly Leu Ala Glu Val Gly Lys Phe Glu Lys Asp Thr 20  25 30
Gly Ile Lys Val Thr Val Glu His Pro Asp Lys Leu Glu Glu Lys Phe 35  40 45
Pro Gln Val Ala Ala Thr Gly Asp Gly Pro Asp Ile Ile Phe Trp Ala 50  55 60
His Asp Arg Phe Gly Gly Tyr Ala Gln Ser Gly Leu Leu Ala Glu Ile 65  70  75 80
Thr Pro Asp Lys Ala Phe Glu Asp Lys Leu Tyr Pro Phe Thr Trp Asp 85  90  95
Ala Val Arg Tyr Asn Gly Lys Leu Ile Ala Tyr Pro Ile Ala Val Glu 100  105 110
Ala Leu Ser Leu Ile Tyr Asn Lys Asp Leu Pro Asp Pro Pro Lys 115 120 125
Thr Trp Glu Glu Ile Pro Ala Leu Asp Lys Leu Lys Ala Gly Ile 130 135 140
Lys Ser Ala Leu Met Phe Asn Leu Gln Glu Pro Tyr Phe Thr Trp Pro 145 150 155 160
Leu Ile Ala Ala Asp Gly Gly Tyr Ala Phe Lys Tyr Glu Asn Gly Lys 165 170 175
Tyr Asp Ile Lys Asp Val Gly Val Asp Asn Ala Gly Ala Lys Ala Gly 180 185 190
Leu Thr Phe Leu Val Asp Leu Ile Lys Asn Asp His Met Asn Ala Asp 195 200 205
Thr Asp Tyr Ser Ile Ala Glu Ala Phe Asn Lys Gly Glu Thr Ala 210 215 220
Met Thr Ile Asn Gly Pro Trp Ala Trp Ser Asn Ile Asp Thr Ser Lys 225 230 235 240
Val Ala Tyr Glu Val Thr Leu Pro Thr Phe Lys Glu Gly Pro Ser 245 250 255
Lys Pro Phe Val Gly Val Leu Ser Ala Gly Ile Asn Ala Ala Ser Pro 260 265 270
Asn Lys Glu Leu Ala Lys Glu Phe Leu Glu Asn Tyr Leu Leu Thr Asp 275 280 285
Glu Gly Leu Glu Ala Val Asn Lys Asp Lys Pro Leu Gly Ala Val Ala 290 295
What is claimed is:
1. A peptide which binds streptavidin with a dissociation constant less than 10 μM and comprises an amino acid sequence having at least 80% identity to the first 38 amino acids of SEQ ID NO:25, wherein said peptide does not contain an HQP, HPM, HPN, or HQP motif;
2. A peptide which binds streptavidin with a dissociation constant less than 23 nM and comprises an amino acid sequence having at least 80% identity to the first 38 amino acids of SEQ ID NO:25, wherein said protein is not disulfide bonded or cyclized;
3. A peptide which binds streptavidin with a dissociation constant less than 23 nM and comprises an amino acid sequence having at least 80% identity to the first 38 amino acids of SEQ ID NO:25, wherein said peptide is not disulfide bonded or cyclized;
4. The peptide of claim 1 or 2, wherein said dissociation constant is less than 5 μM;
5. The peptide of claim 1 or 2, wherein said dissociation constant is less than 1 μM;
6. The peptide of claim 5, wherein said dissociation constant is less than 100 nM;
7. The peptide of claim 6, wherein said dissociation constant is less than 50 nM;
8. The peptide of claim 3, wherein said dissociation constant is less than 10 nM;
9. The peptide of claim 8, wherein said dissociation constant is less than 5 nM;
10. A fusion protein comprising a protein of interest covalently linked to:
(a) a peptide which binds streptavidin with a dissociation constant less than 10 μM and comprises an amino acid sequence having at least 80% identity to the first 38 amino acids of SEQ ID NO:25, wherein said peptide does not contain an HQP, HPM, HPN, or HQP motif;
(b) a peptide which binds streptavidin with a dissociation constant less than 10 μM and comprises an amino acid sequence having at least 95% sequence identity to SEQ ID NO:25, wherein said peptide is not disulfide bonded or cyclized;
(c) a peptide which binds streptavidin with a dissociation constant less than 23 nM and comprises an amino acid sequence having at least 80% identity to the first 38 amino acids of SEQ ID NO:25.
11. The fusion protein of claim 10, wherein said peptide is attached to the amino terminus or the carboxy terminus of said protein of interest, or wherein said peptide is positioned between the amino and carboxy termini of said protein of interest.
12. The fusion protein of claim 10, wherein said peptide is linked to said protein of interest by a linker comprising a protease-sensitive site.
13. The peptide of claim 1, wherein said peptide comprises at least the first 38 amino acids of SEQ ID NO:25.
14. The peptide of claim 1, wherein said peptide has at least 90% sequence identity to SEQ ID NO:25.
15. The peptide of claim 1, wherein said peptide has at least 95% sequence identity to SEQ ID NO:25.
16. The peptide of claim 2, wherein said peptide has at least 90% sequence identity to SEQ ID NO:25.
17. The peptide of claim 2, wherein said peptide has at least 95% sequence identity to SEQ ID NO:25.
18. The peptide of claim 3, wherein said peptide has at least 90% sequence identity to SEQ ID NO:25.
19. The peptide of claim 3, wherein said peptide has at least 95% sequence identity to SEQ ID NO:25.
20. The peptide of claim 10, wherein said peptide has at least 90% sequence identity to SEQ ID NO:25.
21. The peptide of claim 10, wherein said peptide has at least 95% sequence identity to SEQ ID NO:25.

* * * * *