


(List continued on next page.)

Primary Examiner—Karen Cochran Carlson
Assistant Examiner—Sheridan S Snedden
(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.
OTHER PUBLICATIONS


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

Fig. 2B

- mRNA-pep
- mRNA-pep/RNAse
- mRNA-pep/RNAse/ biotin

Biotin Added
% binding  SEQ  ID  NO:
FL  MDEKTGWRGHVVEGLAGELQRLARLEHPGQREPLVQEVEDVDEGLVQDLHGVAGLLDPVEKLL/TDFKKFKJNVS
C1 MDEKTGWRGHVVEGLAGELQRLARLEHPGQREPLVQEVEDVDEGLVQDLHGVAGLLDPVEKLL/TDFKKFKJNVS
C2 MDEKTGWRGHVVEGLAGELQRLARLEHPGQREPLVQEVEDVDEGLVQDLHGVAGLLDPVEKLL/TDFKKFKJNVS
C3 MDEKTGWRGHVVEGLAGELQRLARLEHPGQREPLVQEVEDVDEGLVQ
C4 MDEKTGWRGHVVEGLAGELQRLARLEHPGQREPLVQEVEDVDEGLVQ
M1 MDEKTGWRGHVVEGLAGELQRLARLEHPGQREPLVQEVEDVDEGLVQ
N1 MD  GHVVEGLAGELQRLARLEHPGQREPLVQEVEDVDEGLVQ
N2 MD  GHVVEGLAGELQRLARLEHPGQREPLVQEVEDVDEGLVQDLHGVAGLLDPVEKLL/TDFKKFKJNVS
N3 M  ELEQLARLEHPGQREPLVQEVEDVDEGLVQDLHGVAGLLDPVEKLL/TDFKKFKJNVS

Fig. 5
Fig. 6A (continued)
(SEQ ID No.:38)
MGIEEGKLV WINGDKGY NGLAEVG KKF EKDTGI KVTVEH PD KLEE KFPQVAATGDGP DIIFWAHD RFGGYA QSGLLAEI TPD KAFQDKLY PFTWDAVR YNGKLIAYPIAVEALS LIYNKL PLLPPKPK TWEIPAL DKE LKAGKSA LMFNQEPYFTWPL IAADGGYAFKYEN KYD IDK DVGVDNAGAKAGLTLFLVDLKNKHMNADTDS IAEAAFNKG ETA MINGPWAWSNI DTSKV VNYGVTVPFTKQPSKFVG VLSAGINAASPNE KELAEFLE NYI LTDEGLEA VNKD KPLGAVALKSYEEELAK DRI AATT MNA QGEIMPNIPQMQSAFWYAVRTAVINAASGRQTVDEALKDAQT NSSGSSG SGMDEKTTTG WRGGH VVEGLAGELE Q LRARLEHPQGQREPGS HHHHEFLVPGRSM DPCVKCKVAP RNW KVKNHLRI YNMCKTCF NNS IDI GDDTY HGHDD

Fig. 6B

(SEQ ID No.:39)
MDPCVKCKVAP RNW KVKNHLRI YNMCKTCF NNSIDIGDDTY HGHDD

Fig. 6C
**Fig. 8A**
RNA: TMV AUG FLAG ORF His6

**Fig. 8B**
AUG

**Fig. 8C**
AUG AUG

**Fig. 8D**
AUG deletion

**Fig. 8E**
Nickel purification + anti-FLAG purification

**Fig. 8F**
Type IIS Restriction Enzymes

RT-PCR
ATCCGAGATATAGTTCTCTTCTGCAGAAAAACCCCTGAGACCCGGTTTAGAGGACCCGCAAGGGTTATGCTAGTTATTGC
TCAGCGGTCGACACGCAGGCACCTCAGGTTCTTCGCTGAGCTGTTGAGTGTGGTGTGGTCTGCTG
GAGTGCGCAGCAAGCTTACGCTGAGGCTATAATTAGAATCTGTGATAGTGATGTGGAGGCGATCTGTTACGG
GACCTTTGGGTAGTGCTCGCTAATGCTGTCGACGAAGTTGTLCAATGTCGACCAGCTCTAGTCCACCTCGC
CAACCCAGTGCTCTTCTGCTGCTACCTGCTAGCTGGAATTAGCTGCTGCTGCTTTACGGCTTTCTC
GAGCAGCTGACAGCTCGTGGCGGTTGATACCGACACCGACGATACCGAAGAAGCGGACATCTGCGGGATGTGC
GCATGATTCTACCCCTCTCTGGGCTTTTTTCAT6TGCGCGGAATACGTTGATCTTTGGCGACAACTTC
CTTTGGCAAGCTCTTCTGGTGAGCTGCGGCGGTTATATACCTTGCTCAGCAGCAAGAAGGTTGATGGTAGCTAGCCCT
TGAAGTCTGGGAGTAGGTTAACCACTACATAAATCTACTTTGCGTGTAGTGTGGGACATGCCACGCGCTGTTAGTGC
ATCGCTTTGCTCCTTATTATAGCAGACTCTCTGCGATGGAATATTGCTGTCGATTTTTGTTTATACAG
GTCAACCAGGAAGGTCAGACCCGCCTTGGGCGGCGCTAATCCAGCCACGCTTTTAAATGTCGACTTCGCTGTTTAT
ACTGGAACAGTAAACCCGTCAGACGCAATACGGCAGGTAGATACGCTTCTCGAGTTGAAACATACGGCGCTC
TTACCTTTGCCGTCTTTATCTGCTATCGGCAAGCCTGCTTTCTTTTGGCGGTTGGCAGCATCTTGGTT
ATAAATACGGGATAAAGGCTCTCAAGCGCGGGATCTCTTTTCTGCGGTTAGAAGCAAATCAGCTCTCTGCTGGTGAACG
GTAACAGCTGTGCTGCAAGCGCTTGGCGGTTATACGCGTCGTGACGCAAGAAGGTTGATGGTAGCTAGCCCT
TGTCGCCAGAAGATACTGCGACGCTACGGCAACAGTGGGAGAATCTCTGCTCTCTGCTGATGTCG
AACGGCTCTTTCAGGATGTGTTTCTCTGCAATTTTACTGAGTTAAGGGATAGCTGTTTATCTGAGAATGGGA
TCCAGATATTGCTTCTTCTTCTTGCAATCTGTAATTTGCTGCTGCTGCTGCTGCTGCTGCTG
ATTGTATATCAGCTCTTTCCATACATGTAGCTGTTGAGTGGCAGATTCTGTGCTGCTGCTGATTCTCAGCCGACG
CATCGTGGCAGCAGCTACCTGCGGGCCACAGGTGGCGTGTGCTGCTGCTGCTGCTGCTGCTGCTG
ATCGCTCTTCTGCTGCTCTAGGCTGTGTTCTGGAGTGGTATGTGGGACCCGCGGCGGGGACGCTGTGTTGGCGCC
ATCTCTTGCTCATGCACCACCTCTTCTCGGCGGCGGTGCTCAACCGCTCAACTACTACTGCGGTCTCCTTAAATCGAGA
GTGCAATAGAGGAAGAGCTACAGGCAAGCGCAGCCACCACATCAGAATGCGGAAACCCTTCTCGGGATTAGCGATAGCGCCG
GAAGAGACTCAATTACAGTTGTTGAATGGAACAGTACATTACATACGATTGCGCAGATATGCGGGTGCTTATATCA
GACCCGTTCTCGGCTGTAACAGCGCCACGCGCAGTTCTCTTGCGAAGCGGAGAAATGGAAGGCGGGAGCTGCGAG
TGAAATTACATTTCAACCCGTCGTCGCAACACACTGCGCGAGTTAAAAATTCTGCGGCGGACGAACTGCGTTGGCGTCTCTG
GCCTTCAGCGCAGCTGCAATTTGCGCGGCGGATTAAATTCTGCGGCGGACGAATCTGCGTTGGCGTCTCTG
GTAGAAAGCACGCGCGTCCAGAAGCTGTTAAAGCGGGCCTGCACAATCTTCTCGGCAACGGCGTTCTGAGATCTA
Fig. 9A (continued)
(SEQ ID No.:41)

MGIEEGKLWINGDKGYGLAEVGKKEKDTGIKVTEHPDKLEEKFPQVAATGDGPDIIFWAHDRFGGYAQSGGLAEI
TPDKAFQDKLYPFTWDAVRYNGKLIAYPIAVEALSILIYNKDLLPNPPKTWEEIPALDKELKAKGK5ALMFNLQE
PYFTWP
LIAADGGYAFKYENGKDYIKDVGVDNAGAKAGLTFIDLKIHNMNADTDYSIAEAAAFNKEATMTINGPWAWSNI
DTSK
VNYGVTIVLPFTKIQPSKPVFVGVLASGINAASNKELAKEFLENYLLTDLEGLEAVNKDKPLGAVALKSY
EEELAKDPRIAA
TMENAQKGEIMPNIPQMSAFWYAVRTAVINASGRQTVDEALKDAQTNSSSGSGS55MDEKTTGWGHHVVEGL
AGELEQ
LRARLEHHPQGQRREPSCGHHHHHEF

Fig. 9B
Fig. 10B
Fig. 11
Fig. 13

Maltose Elution
Imidazole Elution
Biotin Elution
Soluble Fraction
Induced
Uninduced

83 kDa  62 kDa  47.5 kDa  32.5 kDa  25 kDa
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 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 µM (that is, binds streptavidin more tightly than a K_d of 10 µM) and which is not disulfide bonded or cyclized. Preferably, the dissociation constant is equal to or less than 5 µM, 1 µM, 100 nM, 50 nM, 25 nM, 10 nM, or even 5 nM. In one preferred embodiment, the dissociation constant is less than 5 µM, 1 µM, 100 nM, 50 nM, 25 nM, 10 nM, or even 5 nM. In another preferred embodiment, the dissociation constant is less than 0.01 nM, 0.1 nM, 1 nM, 5 nM, or 10 nM. In other related aspects, the invention provides nucleic acids encoding the peptides of the present invention, and vectors that include such nucleic acids.

In addition, standard gene fusion techniques may be used to generate fusion nucleic acids that encode fusion proteins which include a peptide of the present invention and a protein of interest. The fusion proteins may be purified, detected, or quantified based on the high affinity of the peptides for streptavidin.

Accordingly, in one such aspect, the invention provides a fusion protein including a protein of interest covalently linked to one of the following peptides: (a) a peptide which binds streptavidin with a dissociation constant less than 10 µM and which is not disulfide bonded or cyclized, (b) a peptide which binds streptavidin with a dissociation constant less than 10 µM and which does not contain an HPQ, HPM, HPN, HNP, or HQP motif, or (c) a peptide which binds streptavidin with a dissociation constant less than 23 nM. In preferred embodiments, the peptide is attached to the amino-terminus or the carboxy-terminus of the protein of interest, or the peptide is positioned between the amino and carboxy-termini of the protein of interest. Preferably, the peptide is linked to the protein of interest by a linker which includes a protease-sensitive site.

In related aspects, the invention provides nucleic acids encoding the fusion proteins of the present invention, and vectors that include these fusion nucleic acids.

In addition, the invention provides a method of producing a fusion protein of the present invention. This method includes transfecting a vector having a nucleic acid sequence encoding the fusion protein into a suitable host cell and culturing the host cell under conditions appropriate for expression of the fusion protein.

The fusion proteins described herein may be used in methods for purifying proteins of interest from samples. Such a method involves expressing the protein of interest as a fusion protein covalently linked to one of the following peptides: (a) a peptide which binds streptavidin with a dissociation constant less than 10 µM and which is not disulfide bonded or cyclized, (b) a peptide which binds streptavidin with a dissociation constant less than 10 µM and which does not contain an HPQ, HPM, HPN, or HQP motif, or (c) a peptide which binds streptavidin with a dissociation constant less than 23 nM. A sample containing the fusion protein is contacted with streptavidin under conditions that allow complex formation between the fusion protein and streptavidin. The complex is isolated, and the fusion protein is recovered from the complex, thereby purifying the protein.
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. An 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 that 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
another molecule. These compounds may, for example, be conjugated to streptavidin through gene fusion technology or protein synthesis and may, for example, be covalently linked to an enzyme, such as horseradish peroxidase or alkaline phosphatase, a fluorescent group, a radiolabel, or any other detectable group.

By “free target molecule” is meant a target molecule that is not bound to a solid support.

By “cyclized” is meant nonlinear. A peptide may be cyclized by the formation of a covalent bond between the N-terminal amino group of the peptide or the side-chain of a residue and the C-terminal carboxyl group or the side-chain of a residue. For example, a peptide lactam may be formed by the cyclization between the N-terminal amino group or an amino group of an amino acid side-chain and the C-terminal carboxyl group or a carboxyl or amide containing side-chain. Other possible cyclizations include the formation of a thioether by the reaction of a thiol in a cysteine side-chain with the N-terminal amino group, C-terminal carboxyl group, or the side-chain of another amino acid. A disulfide bond may also be formed between two cysteine residues. As used herein, a “non-cyclized peptide” is a linear peptide that does not have any of the above cyclizations.

By “dissociation constant” is meant the dissociation constant for binding streptavidin as measured using the electrophoretic mobility shift assay described herein. By “less than” a particular dissociation constant is meant capable of binding streptavidin more tightly than the strength of binding represented by a particular dissociation constant.

By “purifying” is meant separating a compound, for example, a protein, from other components that naturally accompany it. Typically, a protein is substantially pure when it is at least 50%, by weight, free from proteins and naturally-occurring organic molecules with which it is naturally associated. Preferably, the protein is at least 75%, more preferably, at least 90%, and most preferably, at least 99%, by weight, pure. In other preferred embodiments, the protein is at least 2, 5, 10, 25, 50, or 100 times as pure as the starting material. Purity may be assayed by any appropriate method, such as polyacrylamide gel electrophoresis, column chromatography, optical density, HPLC analysis, western analysis, or ELISA (see, for example, Ausubel et al., Current Protocols in Molecular Biology, John Wiley & Sons, New York, 2000).

By “recovered from the complex” is meant physically separated from the complex of streptavidin and the fusion protein. For example, the streptavidin-bound fusion protein may be incubated under conditions that reduce the affinity of the fusion protein for streptavidin (i.e., at low or high salt concentrations or at low or high pH values) or incubated in the presence of molecules that compete with the fusion protein for binding streptavidin. Preferably, either the fusion protein or the streptavidin that has been released from the complex is isolated using standard procedures, such as column chromatography, polyacrylamide gel electrophoresis, HPLC, or western analysis.

The present invention provides a number of advantages related to the detection and purification of proteins of interest. For example, because the present methods do not require the generation of an antibody or other affinity reagent that is specific for each protein of interest, these methods may be universally applied to any protein. In addition, if desired, the streptavidin-binding peptide may be connected to the protein of interest through a protease cleavable linker, allowing removal of the peptide after purification of the fusion protein.

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 (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 in the HPQ sequence. The number of times each sequence was observed is shown. The HPQ sequence is in bold type. Non-random sequences at the termini are underlined. The six C-terminal most nucleotides are not shown.
with these enzymes yields the full-length DNA library. The present methods stem from the discovery of peptides that prematurely stop codons and thus encode a full-length protein so that they can be used as a tool for screening libraries. The present methods involve the use of 11s restriction enzymes which are encoded in each cassette. Based on the presence of both protein affinity tags in the open reading frame encoding the two protein affinity tags tagged), a streptavidin-binding peptide (SEQ ID No.: 35), a hexahistidine tag, and another peptide called 2r18-19dN. FIG. 6C is the amino acid sequence of the SB19 (SEQ ID Nos.: 21–29). The full-length (FL), C-terminal deleted (C1–C4), N-terminal deleted (N1–N3), and point mutated (M1) peptide sequences are shown. The “% binding” refers to the performance of these peptides in the peptide-binding assay. FIG. 7A is the amino acid sequence of the plasmid used for expression of a fusion protein containing a streptavidin-binding peptide (SEQ ID No.: 35). FIG. 7B is a picture of a native gel showing the purity of the fusion protein after elution from the streptavidin column (lane 2) compared to the purity of the E. coli lysate that was applied to the column (lane 1).

Figs. 8A–8F are schematic illustrations of the pre-selection method. FIG. 8A is an illustration of an mRNA display template terminating in puromycin in which the tobacco mosaic virus translation enhancer sequence (TMV), the initiating methionine codon (AUG), and the sections of the open reading frame encoding the two protein affinity tags (FLAG and His) are labeled. FIG. 8B is an illustration of an mRNA display template that is free of frameshifts and premature stop codons and thus encodes a full-length protein containing both affinity tags. FIG. 8C illustrates an mRNA display template that has initiated internally and displays the corresponding truncated protein lacking the N-terminal TAG flag. FIG. 8D shows an mRNA display template that has a deletion in its open reading frame and thus displays the corresponding frameshifted protein lacking the C-terminal His tag. FIG. 8E illustrates the reverse transcription of the mRNA display template from FIG. 8B that was purified based on the presence of both protein affinity tags in the encoded protein. FIG. 8F shows the cleavage sites for Type IIS restriction enzymes which are encoded in each cassette. Ligation of pre-selected cassettes which have been cleaved with these enzymes yields the full-length DNA library.

FIG. 9A is the nucleotide sequence of the mRNA display template terminating in puromycin which contains the first 38 amino acids of the SB19-C4 peptide is underlined. FIG. 10A is a graph of the Biorec (Surface Plasmon Resonance) response units over various lengths of time for the dissociation of streptavidin from the fusion protein listed in FIG. 9B immobilized on a biacore chip. For line “a,” the streptavidin concentration is 25 μM; for line “b,” the concentration is 11.5 μM, and for line “c,” the concentration is 5.75 μM. This data was used to calculate an upper limit of 2x10^{-3}s for the dissociation rate, K_d. FIG. 10B is a graph showing the association and subsequent dissociation of streptavidin from the immobilized fusion protein. For lines “a” and “b,” the streptavidin concentrations are 1.6, 0.8, 0.4, 0.2, 0.1, and 0.05 μM, respectively. This data was used to calculate an association rate, k_a, of 5x10^{7}/M/s.

FIG. 11 is a map of the pTAG2 vector indicating the order of the domains in the multiply-tagged fusion protein it encodes.

FIG. 12 is a graph showing a K_d determination of the peptide used for expression of a fusion protein containing a streptavidin-binding peptide (SEQ ID No.: 35) used as an affinity tag for the purification of the fusion protein listed in FIG. 6B. This peptide contains the first 38 amino acids of the SB19-C4 peptide (FIG. 5). FIG. 12B is a graph of the Biorec (Surface Plasmon Resonance) response units over various lengths of time for the association of streptavidin from the fusion protein listed in FIG. 9B immobilized on a biacore chip. For line “a,” the streptavidin concentration is 25 μM; for line “b,” the concentration is 11.5 μM, and for line “c,” the concentration is 5.75 μM. This data was used to calculate an upper limit of 2x10^{-3}s for the dissociation rate, K_d. FIG. 10B is a graph showing the association and subsequent dissociation of streptavidin from the immobilized fusion protein. For lines “a” and “b,” the streptavidin concentrations are 1.6, 0.8, 0.4, 0.2, 0.1, and 0.05 μM, respectively. This data was used to calculate an association rate, k_a, of 5x10^{7}/M/s.

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 supernatants 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 right-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.
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>Standard deviation (n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-selected</td>
<td>HPQ 19 HPQ</td>
<td>0.16</td>
<td>Two HPQ motifs</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>4.8</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>
<tr>
<td>SB6</td>
<td>HPL 4 HPQ</td>
<td>49</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SB7</td>
<td>HPL 7 HPQ</td>
<td>28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SB8</td>
<td>HPL 12 HPL</td>
<td>27</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SB9</td>
<td>HPL 12 HP</td>
<td>64</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SB10</td>
<td>HPL 21 QPO</td>
<td>15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SB11</td>
<td>HPL 28 HPL</td>
<td>88</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SB12</td>
<td>HPL 30 EPO</td>
<td>73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SB13</td>
<td>HPL 32 EPO</td>
<td>64</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SB14</td>
<td>HPL 43 HPL</td>
<td>11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SB15</td>
<td>QPO 50 HPQ</td>
<td>44</td>
<td>92</td>
<td>16 (4)</td>
</tr>
<tr>
<td>SB16</td>
<td>HPL 74 HPQ</td>
<td>50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SB17</td>
<td>8.3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SB18</td>
<td>68</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SB19</td>
<td>85</td>
<td>10</td>
<td>1.8 (10)</td>
<td></td>
</tr>
<tr>
<td>SB19-C4</td>
<td>88</td>
<td>4.9</td>
<td>0.88 (10)</td>
<td></td>
</tr>
<tr>
<td>SB20</td>
<td>HPL</td>
<td>34</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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 either 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., 1999).

In one preferred embodiment of this method, mRNA display templates that contain a translation enhancer sequence operably-linked to an open reading frame and 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 displayed 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 reversed 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, 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 concatamerized 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 2.4×10^4 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 3' 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 reaction mixture was supplemented with 2 mCi 35S-methionine and a total methionine concentration of 10 μM. The reaction mixture also included 300 μM 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 1.2×10^9 polypeptides 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 polypeptides 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 mg 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 tRNA (Boehringer-Mannheim, Indianapolis, Ind.), and 5 mg/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 and de-salted using two successive NAP columns (Amersham-Pharmacia, Piscataway, N.J.) equilibrated in 1 mM Tris(hydroxymethyl)aminomethane, 0.01% Triton X-100, 50 µM EDTA, 0.5 mM 2-mercaptoethanol, 0.5 µg/mL tRNA (Boehringer-Mannheim, Indianapolis, Ind.), and 50 mg/mL BSA at pH 7.6.

The mRNA portion was then reverse transcribed using Superscript II (Gibco BRL, Rockville, Md.) according to the manufacturer's 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 1 µM of "splint" 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 mRNA displayed peptides were run on an SDS-PAGE gel, as shown in FIG. 1B.

### Selection of Streptavidin-Binding Peptides

For selection of peptides with high affinity for streptavidin, the above mRNA-displayed library was incubated with immobilized streptavidin (Ultralink immobilized streptavidin, Pierce, Rockford, Ill.) in streptavidin-binding buffer under reducing conditions (40 mM Tris(hydroxymethyl)aminomethane, 300 mM KCl, 2 mM EDTA, 0.1% Triton X-100, 5 mM 2-mercaptoethanol, 100 µg/mL BSA, and 1 µg/mL tRNA at pH 7.4). The amount of gel used was 0.5 mL in a total volume of 5.5 mL. After incubating for 20 minutes at room temperature, the contents were loaded onto a disposable chromatography column, drained, washed with 14 column volumes of streptavidin-binding buffer, and eluted with five successive aliquots, at 10 minutes intervals, of streptavidin-binding buffer plus 2 mM D-biotin (Sigma, St. Louis, Mo.) (FIG. 1A). The fraction of the library that survived this purification was 0.08%. Elution fractions were combined, de-salted on NAP columns, and then PCR-amplified to regenerate the double-stranded DNA library using the described conditions and primers in a 8 µL reaction mixture (Cho et al., supra).

This concluded the first round of selection, and the remaining six rounds followed the same protocol except that the translation was scaled down 10-fold, and the number of column volumes for washing the streptavidin column was increased (32 volumes for round 2; 40 volumes for rounds 3, 4 and 6; and 25 volumes for rounds 5 and 7). The streptavidin-binding selection for rounds 5 and 7 was performed directly on the streptavidin-column eluate from the preceding selection rounds, without intervening amplification (the biotin was removed by three successive passages through NAP columns). PCR products amplified after the seventh selection round were cloned and the TOPO TA cloning kit (Invitrogen, Carlsbad, Calif.), following the manufacturer's protocol. The fraction of the library that bound and eluted from the streptavidin column increased in each round, reaching 61% at round seven (FIG. 2A).

### Characterization of the Selected Library

The elute from the seventh round of selection was amplified by PCR. The resulting PCR DNA was used to synthesize a library of displayed peptides to confirm that the displayed peptides, rather than the RNA or DNA portion 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 library 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. Prior to 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 eliminate 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' primer as described for the library construction (Cho et al., supra), and a new 3' primer, ATAGCCGGTGCCAAGTTGGAGCGCCGACAGCA GT-3'; SEQ ID No. 30), which altered the 3' RNA sequence to ACUGGUCUGGCGGCUAGCAACGUUAGCCAC CGGCUAAU (SEQ ID No. 31). This sequence was designed to anneal to the photo-crosslinking linker, which has the sequence 5'-psoralen-TAGCCGGTTG-A17-CC-purymycin-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 WSSGCVHHHHHSSA (SEQ ID No. 32) to WSGGCCLGTY (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, results in half of the DTP being mobility-shifted. To determine the dissociation constants of the selected peptides for streptavidin are shown in FIG. 4A. Some clones showed some clones showed

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 irreversibility 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 (separated by 19 amino acids) were introduced into a control, unselected member of the streptavidin column-binding assay (FIG. 4B), and the data is graphed in FIG. 4C. The dissociation constants for the clones ranged from 110 nM to 4.3 µM. These surprisingly high affinities were determined using the equation $K_d = [\text{streptavidin}] R$, where 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$_{295}$ using the molar extinction coefficient of 57,000 per monomer.

Examples of these mobility shifts in the presence of streptavidin are shown in FIG. 4. A. 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...
sequence at various positions and also replaced two codons (encoding Asp and Trp) in the C-terminal constant region with methionine codons to increase the 35S-incorporation. Analogous primers were used for the N-terminal truncation analysis, except that no change was made in the N-terminal constant sequence.

Deletion of up to 56 residues had no observable effect on the binding strength. Peptide SB19-C4 retained only the first 38 residues from the selected construct (plus the C-terminal MMSSGGCKLG sequence, SEQ ID No. 36). Mutating the HPQ motif to HGA reduced the activity by three orders of magnitude (compare construct C4 to M1). Results from the N-terminal truncation constructs (N1-N3) suggested that binding determinants were spread throughout the N-terminal 38 residues of peptide SB19. Of the peptides tested, SB19-C4 was therefore the minimal peptide retaining full activity in this assay. EMSA analysis of peptide SB19-C4 confirmed high affinity streptavidin-binding, but a fraction (13%) of the peptide was inactive even at streptavidin concentrations >1 μM. The majority (87%), however, had an apparent Kd of 4.9 nM after correction for the amount of inactive peptide. Purification of Fusion Protein Containing Streptavidin-Binding Peptide

A fusion protein containing the first 38 amino acids of the SB19-C4 streptavidin-binding peptide (FIG. 7A) was expressed in E. coli and then purified from the cell lysate. For the expression of the fusion protein, BL21 (DE3) cells were transformed with a plasmid containing a Malose Binding Protein—Streptavidin-binding Peptide—His-tag—Protein of Interest insert (FIG. II) which encodes a fusion protein containing, from the amino- to carboxy-terminus, the first 38 residues of the SB19-C4 sequence, a hexahistidine tag, and another peptide called 2R18-19DN (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 × g for 20 minutes. The pelleted cells were 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 × g for 20 minutes at 4°C. Purification using the His-tag was also carried out in an analogous manner to the streptavidin binding peptide-tag procedure with the same amounts of cells and in the same volumes. The soluble fraction was prepared in His-tag binding buffer (300 mM NaCl, 50 mM sodium phosphate, 0.25% Triton X-100, 10 mM imidazole, pH 8.0). The sample was applied directly to the Ni-column (Ni—NTA, Qiagen, Valencia Calif.) 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. Purification using the maltose-binding protein sequence was carried out in an analogous manner to the streptavidin binding peptide-tag procedure with the same amounts of cells and in the same volumes.

More stringent conditions eliminated the contaminating protein of higher molecular weight. The soluble fraction of lysed induced cells was prepared in Streptavidin-binding buffer (SBB: 300 mM KCl, 40 mM Tris(hydroxymethyl) amino methane, 5 mM 2-mercaptoethanol, 2 mM EDTA, 0.1% Triton-X 100, pH 7.4) as described above. This sample (e.g., 79 mg net weight cells in 1 ml SBB) 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 were analyzed as described above.

For detection, ten picomoles of the loaded and eluted samples from the immobilized streptavidin column, the Ni-column, and the amylose column was loaded onto a 12% SDS-Tricine PAGE gel. In the adjacent lane, a whole bacterial extract from BL21 (DE3) cells was loaded. The whole cell extract was prepared by growing the cells to saturation in LB, removing the media by centrifugation, and re-suspending them in 10% of the original culture volume with the SDS-PAGE protein-loading buffer. Six microliters of this extract was run on the gel. These two samples and a molecular weight marker were run side by side on the gel in duplicate. After running the gel, it was stained with Coomassie Brilliant Blue.
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 Escherichia 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, 300 mM NaCl, 2.68 mM KCl, 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 RPN1231) 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′-tetramethylbenzidine, Promega catalog number W4121) 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.

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. 9B, 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 as described above. This fusion protein was purified from the E. coli extract using amylose resin to bind the maltose-binding protein portion of the fusion protein and then Ni—NTA resin to bind the hexahistidine tag.

To measure the affinity of the fusion protein for streptavidin, the fusion protein was immobilized on a biacore chip through the crosslinking of free amino groups in the fusion protein to the biacore chip. Buffer containing streptavidin was washed over the chip, allowing streptavidin to bind the immobilized fusion protein (FIG. 10B). This resulted in an increase in the biacore response units which are proportional to the amount of streptavidin adhering to the biacore chip. Then buffer without streptavidin was washed over the chip, and the biacore response units decreased as streptavidin dissociated from the immobilized fusion protein (FIGS. 10A and 10B). To measure the association rate for the binding of streptavidin to the fusion protein, streptavidin concentrations of 0.6, 0.8, 1.0, 2.0, 4.0, and 8.0 μM, (lines “a” to “f” in FIG. 10B, respectively) were washed over the biacore chip. The buffer also contained 40 mM Tri(hydroxyethyl)aminomethane, 300 mM KCl, 2 mM EDTA, 0.1% w/v Triton X-100, and 5 mM 2-mercaptoethanol at pH 7.4. This data was used to calculate an association rate, kₐ, of 5x10⁶ 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 for the dissociation rate, kₐ (BIACORE X Instrument Handbook, supra). Based on these calculated association and dissociation rates, the dissociation constant, K_d, 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 $K_p$ 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 specificities 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 ($<K_p$) 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 μm 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 $K_p$ 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 $K_p$ 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 $K_p$, 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 $K_p$ values for protein-protein, protein-peptide, and protein-small molecule complexes.

To measure the $K_p$ 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 extra MgCl$_2$ and 100 mM extra KCl. This mixture was then purified successively upon the basis of the FLAG-tag and then the His-tag (using the manufacturer’s instructions; denaturing conditions were used for the Ni—NTA purification). The resultant purified peptide was then dialyzed into streptavidin-binding buffer (SBB: 300 mM KCl, 40 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 pM 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 $y=\frac{b}{x+(\frac{K_p}{x}+c)}$ in which $y$ was the number of radioactive decompositions detected per minute in each flow-through, $K_p$ is the dissociation constant of the complex, $x$ was the concentration of free streptavidin, $b$ was the number of counts per minute not competent to bind the matrix under the assay conditions, and $c$ was the number of counts per minute competent to bind the matrix under the assay conditions. $K_p$, $b$, 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 $K_p$ of 2.5 nM, which compared very favorably with the $K_p$ determined by other methods. For example, the same peptide was purified on a FLAG-affinity matrix, then bound to streptavidin immobilized on a microparticle 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 ~5,000-fold higher affinity than does the Strep-tag II (Schmidt et al. *J. Mol. Biol.* 255:753–766, 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 flanking 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 adopt it to various uses 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
Leu Val Pro His His Pro Gln Gly Asp Pro Leu Arg Leu Leu His
Arg Pro Gln Pro Ala Leu Val Arg His Pro Gln Gly Asp Leu Val
Leu Val Glu His His Glu Val Asp Arg Gln Leu Val Ala Leu
Pro Glu Leu His Ala Glu Leu Gly Glu Pro Val Asp Ala Leu Val
Gln Gly Pro Val Glu Gln Val Gly Val Val Asp Ala Leu Val Trp
Arg Leu Pro Pro Ser

<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
Leu Val Glu Leu Gln Ala Leu Ala Gly Leu Gln Arg Gln Gly
Gly Arg Gln Pro His Arg Pro Arg Arg Pro His His Leu Gln
Leu Leu Asp Glu Ala His Pro Gln Ala Gly Pro Leu Arg Glu Arg
His Gln Val Asp Gly Arg Leu Leu Gln His Pro Gln Gly
Arg Leu Leu Gln Pro Gln Asp His Pro Leu Glu Leu Val Trp
Arg Leu Pro Pro Ser

<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
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 Pro Leu His 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 Gly Ile Ser Thr Trp Arg Gly Glu Asp
1 5 10 15
Leu Leu Val Arg Gly His Gly Asp Val Glu Ala Leu His Asp Gln
20 25 30
Gly Leu His Gly Leu Leu Asp Leu Val Gly Pro Pro Gly Val Val
35 40 45
Arg Ala Leu Arg Gly Val Leu Gly Leu Arg Arg Pro Val Leu Pro
50 55 60
Leu His Pro Gln Gly Glu Ala Asp Ala 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: 102
<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 Asp Phe Glu Leu Leu Ala
   1      5      10      15
Arg Leu Asp Gly Leu Arg Gly Gly Asp His Pro Leu Val Leu Arg
   20     25     30
His His Pro Gln Gly Asp Gly Leu Asp Gln Pro Leu Gly Arg His
   35     40     45
Arg Ala Leu Asp Gly Leu Gly Asp Asp Leu Gly Asp Leu Asp Gln
   50     55     60
Gly Gly Glu Leu Gly Leu Val Asp Asp Gly Glu Glu Asp His Pro Leu
   65     70     75     80
Leu Asp Gly Leu Val His Val Gly Val His Val Asp Pro Leu Val
   85     90     95
Cys Gly Cys His His His
   100

<210> SEQ ID NO 7
<211> LENGTH: 101
<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 Asp Gly Ile Asp Cys Ser Phe Asn
   20     25     30
Met Asn Leu Ser Leu Asn His Trp Leu Ser Ser Gly His Pro Asp Gly
   35     40     45
Ala Leu Asp Glu Leu His Pro Gln Gly Asp Ala Leu Val Gly Arg
   50     55     60
Asp Asp Gly Val Val Gly Ala Leu Arg Leu Glu Gly Gln His Gln His
   65     70     75     80
Arg Arg Leu 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: 102
<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 Asp
   20     25     30
Gly Val Arg Pro His His Pro Leu Ala Asp Gly Gly Gly Gly Leu Asp
35 40 45
Gln Gly Leu Gly His Arg Arg Gly 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 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 Asn Leu Asn Lys Asp Val Ala Pro Leu Val Glu Gly
20 25 30
Arg His Asp Leu Val Glu Leu Leu Arg His Pro Gln Gly
35 40 45
Asp Pro Leu Val Ala Asp Arg Glu Arg Leu 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 Glu 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 Gly Leu Pro Leu Val Arg Gly Glu 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
LENGTH: 102
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: selected peptide

SEQUENCE: 11
Met Asp Glu Lys Thr His Trp His Glu Val His Leu Ala Asp
1 5 10 15
Gly Leu Glu Gin His Pro Gln Gly Gin Arg Gin Pro Leu Val Glu Gin
20 25 30
His Arg Gin Val Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
35 40 45
Leu Pro Leu Glu Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
50 55 60
Gly Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
65 70 75 80
Val Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
85 90 95
Trp Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
100

SEQ ID NO 12
LENGTH: 101
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: selected peptide

SEQUENCE: 12
Met Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
1 5 10 15
Leu Val Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
20 25 30
Gly Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
35 40 45
Gly Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
50 55 60
Pro Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
65 70 75 80
Glu Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
85 90 95
Arg Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
100

SEQ ID NO 13
LENGTH: 102
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: selected peptide

SEQUENCE: 13
Met Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
1 5 10 15
Ala Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
20 25 30
Asp Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
35 40 45
Pro Gly Leu Gln His His Leu Glu Pro Gln Pro Leu Gly Leu Leu Gly
50 55 60
Glu Leu Gln Ala Arg Leu Gln Pro Leu Ala Gly Glu His Glu Gly Asp
65 70 75 80
Gly Ala Gly Leu Gln Arg Val Pro Gly His Gly Arg Arg Leu Val 85 90 95
Trp Arg Leu Pro Pro Ser 100

<210> SEQ ID NO 14
<211> LENGTH: 101
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: selected peptide
<400> SEQUENCE: 14
Met Asp Glu Lys Thr His Arg Thr Leu Ser Val Ser Leu Ser Phe Asn
1 5 10 15
Asp Trp Leu Gly Gln Thr Lys Ala Cys Trp Arg Leu Val Glu Gly Leu
20 25 30
His Gly His Pro Gln Gly Leu Val Arg Glu His Glu Val Asp Val Leu
35 40 45
Pro Leu Ala Gln Glu Val Gln Val Val Gly Leu Ala Asp Gly
50 55 60
Val Glu Gln Pro Gly Gly Leu Leu His Arg Ala Gln Arg Val Asp
65 70 75 80
His Pro Leu Pro Asp His Ala Gly Gln Val Leu Gly Arg Leu Val Trp
85 90 95
Arg Leu Pro Pro Ser 100

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

<210> SEQ ID NO 16
<211> LENGTH: 102
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<table>
<thead>
<tr>
<th>Feature</th>
<th>Other Information</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td>selected peptide</td>
<td>Met Asp Glu Lys Thr His Trp Val Leu Gln Leu His Pro Gln Gly Asp 1 5 10 15</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Arg Leu Gly Pro Arg His Gly Asp Asp Val Arg Leu Val Gly Asp 20 25 30</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Gly Glu Val Leu Glu Gly Leu Asp Gly Arg Pro Arg Arg Arg Arg 35 40 45</td>
</tr>
<tr>
<td></td>
<td></td>
<td>His Arg Leu Pro Arg Glu Asp Glu His Arg Val Ala Leu Val Asp 50 55 60</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Gln Val Arg Asp Leu Ala Glu Arg Leu Val Glu Val Asp Gly Gly 65 70 75 80</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Val Glu Ala Leu Arg His Leu Gly Leu Pro Gln Asp Glu Pro Arg Ser 85 90 95</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Gly Gly Cys His His His 100</td>
</tr>
</tbody>
</table>

| 17      | selected peptide                | Met Asp Glu Lys Thr His Trp Val Gly Asp Leu Gln Glu Pro Leu Gly 1 5 10 15 |
|         |                                 | Pro Leu His Gly Gly Val Gly Val Pro Gly Gly Leu Val Leu Arg 20 25 30 |
|         |                                 | His His Pro Gln Arg Asp Arg Leu Val Asp Gly Val Gly Pro His Gly 25 40 45 |
|         |                                 | Arg Ala Leu Ala Arg Arg Pro His Arg Val Val Glu Gly Leu His His 50 55 60 |
|         |                                 | Leu Leu Gln Arg Gly Gly Glu Arg Leu Pro Asp Gly Pro Arg Gln 65 70 75 80 |
|         |                                 | Leu Gly Leu Gly Gly Leu Asp Arg Ala Asp Pro Ala Leu Val 85 90 95 |
|         |                                 | Trp Arg Leu Pro Pro Ser 100 |

| 18      | selected peptide                | 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 Gln Pro Leu Asp Pro His Pro 20 25 30 |
|         |                                 | Gln Gly Asp His Leu Arg Leu Glu Pro Leu Gly His Ala Leu Val Asp 35 40 45 |
|         |                                 | Pro Leu Val Gln Gly Val Glu Val Val Arg Pro Leu Gln Leu Asp 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 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 Glu Leu Glu Gin Leu Arg Ala Arg Leu Glu His His Pro
20 25 30
Gln Gly Gin Arg Glu Pro Leu Val Gin Glu Val Glu Aep Val Asp Glu
35 40 45
Gly Leu Val Gin Aep Leu His Gly Val Val Ala Gey Val Leu Leu Aep Pro
50 55 60
Val Glu Lys Leu Thr Leu Asp Trp Phe Lys Lys Phe Lys 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 Gin Leu Asp Gin Leu Arg Leu Glu
50 55 60
His His Arg Asp Gin Leu Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
65 70 75 80
Asp His Pro Leu Glu Pro Asp His Leu Val Gin 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 Phe Lys 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 Lys Val Gly
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 Phe 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
**SEQUENCE:** 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 Glu Met Ser Gly Gly Cys Lys Leu Gly
50 55 60

**SEQ ID NO:** 25

**LENGTH:** 47

**TYPE:** PRT

**ORGANISM:** Artificial Sequence

**FEATURE:**

**OTHER INFORMATION:** selected peptide

**SEQUENCE:** 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

**SEQ ID NO:** 26

**LENGTH:** 47

**TYPE:** PRT

**ORGANISM:** Artificial Sequence

**FEATURE:**

**OTHER INFORMATION:** selected peptide

**SEQUENCE:** 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

**SEQ ID NO:** 27

**LENGTH:** 39

**TYPE:** PRT

**ORGANISM:** Artificial Sequence

**FEATURE:**

**OTHER INFORMATION:** selected peptide

**SEQUENCE:** Met Asp Gly His Val Val Glu Gly Leu Ala Gly Glu Leu 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

**SEQ ID NO:** 28

**LENGTH:** 89
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: selected peptide

<400> SEQUENCE: 28
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 Val Glu Glu Val Glu
20 25 30
Asp Val Asp Glu Leu Val Gln Asp Leu His Gly 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: 29
Met Glu Leu Glu Glu 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
ataagccggtg ccaagcttgc agccgccaga ccagt

<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 cccagcggcag guccgaccg gcuau

<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 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 Glu Gln 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
ggggtatat gcagtcgtgc taacgcgttg cagcagccaa ctcagcttcc tttcgggctt
120
tgtagcagc cgatotcag tggtagtgtgtgttggtgtgcgaaggcggg 180
agtcgtcagt cccatgatag gttgtgctcc cgatcctacat gctatggtaa aagcaggtct 240
taaccgtgt atgcggttcc aagtgggtgt ttccacccct cccgtgctcgg ggagccaccc 300
tggatcctcc aacaggggtc atgtaaacct ggggttaccc gaaatgtgtaa tgggtgtgat 360
ggtcgcaggg ttcctgcgctc cgtgcggtgtct gtcggtctct ttcaggggtt 420
gttacagtt cccagccagt ctttacaacc cgtgacccgg cttgaaacca tgcgtctctct 480
cytcacatcc tggcgtgcga cctctgcggtct cgcgatcctg ttcagggctc 540
catagacatt cgcgcgaggg tgggtgaggt gttttttaa tgcggttggc 600
aagcgcgat cgcgcgaggg tgggtgaggt gttttttaa tgcggttggc 660
cgcagatcc tgcgcgaggg tgggtgaggt gttttttaa tgcggttggc 720
ggcggttgta tggggtgaggt gttttttaa tgcggttggc 780
actcttggtga cagcggtgcag gtcggttggc 840
cgcagatcc tgcgcgaggg tgggtgaggt gttttttaa tgcggttggc 900
tggcgcagtt cgcgcgaggg tgggtgaggt gttttttaa tgcggttggc 960
tatagacatt cgcgcgaggg tgggtgaggt gttttttaa tgcggttggc 1020
ttcagttgat cgcgcgaggg tgggtgaggt gttttttaa tgcggttggc 1080
atagatcgta cagcggtgcag gtcggttggc 1140
goacgggtta tggggtgaggt gttttttaa tgcggttggc 1200
gttcttgta cgcgcgaggg tgggtgaggt gttttttaa tgcggttggc 1260
ttgcttcgga cagcgggtga cggggttgcgg cttccgggtc 1320
aagctgaggg gtcggttggc 1380
ttcagttgat cgcgcgaggg tgggtgaggt gttttttaa tgcggttggc 1440
ttgcttcgga cagcgggtga cggggttgcgg cttccgggtc 1500
ttcagttgat cgcgcgaggg tgggtgaggt gttttttaa tgcggttggc 1560
ttgcttcgga cagcgggtga cggggttgcgg cttccgggtc 1620
ttacattctt tcctttataa aaccaaaatcattctcagag gaaatgtgtaa tgggtgtgat 1680
aatcacttc aatgggtgtgg gagagttttttc cctagggcgtg 1740
gcggctcgac cgtggtgtgg ctcgggtggtt cttgggtgttt 1800
aatcacttc aatgggtgtgg gagagttttttc cctagggcgtg 1860
ttgcttcgga cagcgggtga cggggttgcgg cttccgggtc 1920
ttgcttcgga cagcgggtga cggggttgcgg cttccgggtc 1980
aatcacttc aatgggtgtgg gagagttttttc cctagggcgtg 2040
aatcacttc aatgggtgtgg gagagttttttc cctagggcgtg 2100
aatcacttc aatgggtgtgg gagagttttttc cctagggcgtg 2160
aatcacttc aatgggtgtgg gagagttttttc cctagggcgtg 2220
aatcacttc aatgggtgtgg gagagttttttc cctagggcgtg 2280
aatcacttc aatgggtgtgg gagagttttttc cctagggcgtg 2340
aatcacttc aatgggtgtgg gagagttttttc cctagggcgtg 2400
aatcacttc aatgggtgtgg gagagttttttc cctagggcgtg 2460
aatcacttc aatgggtgtgg gagagttttttc cctagggcgtg 2520
aatcacttc aatgggtgtgg gagagttttttc cctagggcgtg 2580
gaagacggyta cggcagctggg cgtaggcacat cggtgtcata tgggtcaaca gcgaatgcgg 2640
cgtaggcacc gcccataaag ttctgtcctcg gcgcgtctgc gtctggctgg ctggcataaa 2700
tacotcaaca acggtgaccgc tgggagcag tctcctctcg ggcggcgggg gcggcgcggc 2760
tcgggggggtctgttttcc gctgttttcgcgcgttgaaaa ggtggtgcaag ggcgccacaag 2820
tcgggggggtctgttttcc gctgttttcgcgcgttgaaaa ggtggtgcaag ggcgccacaag 2880
tcgggggggtctgttttcc gctgttttcgcgcgttgaaaa ggtggtgcaag ggcgccacaag 2940
tcgggggggtctgttttcc gctgttttcgcgcgttgaaaa ggtggtgcaag ggcgccacaag 3000
ttcttcacc gcccataaag ttctgtcctcg gcgcgtctgc gtctggctgg ctggcataaa 3060
tgcccataaa acgggatggtc cggcacaggtt cgcggccatt tgcgatgctg gttgccacag 3120
tgcccataaa acgggatggtc cggcacaggtt cgcggccatt tgcgatgctg gttgccacag 3180
tgcccataaa acgggatggtc cggcacaggtt cgcggccatt tgcgatgctg gttgccacag 3240
tgcccataaa acgggatggtc cggcacaggtt cgcggccatt tgcgatgctg gttgccacag 3300
tgcccataaa acgggatggtc cggcacaggtt cgcggccatt tgcgatgctg gttgccacag 3360
ttcttcacc gcccataaag ttctgtcctcg gcgcgtctgc gtctggctgg ctggcataaa 3420
tgcgatgctg gttgccacag 3480
tgcgatgctg gttgccacag 3540
tgcgatgctg gttgccacag 3600
tgcgatgctg gttgccacag 3660
tgcgatgctg gttgccacag 3720
tgcgatgctg gttgccacag 3780
tgcgatgctg gttgccacag 3840
tgcgatgctg gttgccacag 3900
tgcgatgctg gttgccacag 3960
tgcgatgctg gttgccacag 4020
tgcgatgctg gttgccacag 4080
tgcgatgctg gttgccacag 4140
tgcgatgctg gttgccacag 4200
tgcgatgctg gttgccacag 4260
tgcgatgctg gttgccacag 4320
tgcgatgctg gttgccacag 4380
tgcgatgctg gttgccacag 4440
tgcgatgctg gttgccacag 4500
tgcgatgctg gttgccacag 4560
tgcgatgctg gttgccacag 4620
tgcgatgctg gttgccacag 4680
tgcgatgctg gttgccacag 4740
tgcgatgctg gttgccacag 4800
tgcgatgctg gttgccacag 4860
tgcgatgctg gttgccacag 4920
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 Lys Phe Glu Lys Asp Thr
20 25 30

Gly Ile Lys Val Thr Val Glu His Pro Asp Lys Leu 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 Gin Ser Gly Leu Leu Ala Glu Ile
65 70 75 80
Thr Pro Asp Lys Ala Phe Gin Asp Lys Leu Tyr Pro Phe Thr Trp Asp
85 90 95
 Ala Val Arg Tyr Aen Gly Lys Leu Ile Ala Tyr Pro Ile Ala Val Glu
100 105 110
 Ala Leu Ser Leu Ile Tyr Aen Lys Aep Leu Leu Pro Aen Pro Pro Lys
115 120 125
Thr Trp Glu Glu Ile Pro Ala Leu Aep Lys Leu Lys Ala Lys Gly
130 135 140
 Lys Ser Ala Leu Met Phe Aen Leu Gin Glu Pro Tyr Phe Thr Trp Pro
145 150 155 160
 Leu Ile Ala Asp Gly Gyl Tyr Ala Phe Lys Tyr Glu Aen Gly Lys
165 170 175
 Tyr Aep Ile Lys Asp Val Gly Val Aep Aen Ala Gly Ala Lys Ala Gly
180 185 190
 Leu Thr Phe Leu Val Aep Leu Ile Lys Aen Lys His Met Aen Ala Aep
195 200 205
Thr Aep Tyr Ser Ile Ala Glu Ala Ala Phe Aen Lys Gly Gyl Thr Ala
210 215 220
 Met Thr Ile Aen Gly Pro Trp Ala Trp Ser Aen Ile Aep Thr Ser Lys
225 230 235 240
 Val Aen Tyr Gly Val Thr Val Leu Pro Thr Phe Lys Gly Gin Pro Ser
245 250 255
 Lys Pro Phe Val Gly Val Leu Ser Ala Gly Ile Aen Ala Ala Ser Pro
260 265 270
Aen Lys Glu Leu Ala Lys Gyl Phe Leu Glu Aen Tyr Leu Leu Thr Aep
275 280 285
 Glu Gly Leu Glu Ala Val Aen Lys Aep Lys Pro Leu Gly Glu Val Ala
290 295 300
Leu Lys Ser Tyr Glu Glu Leu Ala Lys Aep Pro Arg Ile Ala Ala
305 310 315 320
 Thr Met Glu Aen Ala Gin Gly Gyl Gin Ile Met Aen Ile Pro Gin
325 330 335
 Met Ser Ala Phe Trp Tyr Ala Val Arg Thr Ala Val Ile Aen Ala Ala
340 345 350
 Ser Gly Arg Gin Thr Val Asp Glu Ala Leu Lys Asp Ala Gin Thr Aen
355 360 365
Ser Ser Ser Gly Gly Ser Gly Met Aep Glu Lys Thr Gly
370 375 380
Trp Arg Gly Gin Val Val Glu Leu Ala Gyl Gly Leu Glu Glu
385 390 395 400
Leu Gin Ala Arg Leu Gin His Pro Gin Gly Gin Arg Glu Pro Gly
405 410 415
 Ser Gly His Gin His Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
420 425 430 435 440 445
Met Aep Pro Cy1 Val Lys Cy1 Val Ala Pro Aen Trp Lys Val
450 455 460
Lys Aen Lys His Leu Arg Ile Tyr Aen Met Cy1 Lys Thr Cy1 Phe Aen
<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
Aen Ser Ile Asp Ile Gly Asp Aen 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
1 20
gggtttagt agttcctcct ttcagcaaaa aacccctcaa gacccgttta gaggccccaa
1 20
tgtagaaco cggatcctc ttcagcaaaa aacccctcaa gacccgttta gaggccccaa
20 40
tgcggaac ttcagccct ttcagcaaaa aacccctcaa gacccgttta gaggccccaa
40 60
gacccgttta agttcctcct ttcagcaaaa aacccctcaa gacccgttta gaggccccaa
60 80
caccccgttta ttcagcaaaa aacccctcaa gacccgttta gaggccccaa
80 100
caccccgttta ttcagcaaaa aacccctcaa gacccgttta gaggccccaa
100 120
caccccgttta ttcagcaaaa aacccctcaa gacccgttta gaggccccaa
120 140
caccccgttta ttcagcaaaa aacccctcaa gacccgttta gaggccccaa
140 160
caccccgttta ttcagcaaaa aacccctcaa gacccgttta gaggccccaa
160 180
caccccgttta ttcagcaaaa aacccctcaa gacccgttta gaggccccaa
180 200
caccccgttta ttcagcaaaa aacccctcaa gacccgttta gaggccccaa
200 220
caccccgttta ttcagcaaaa aacccctcaa gacccgttta gaggccccaa
220 240
caccccgttta ttcagcaaaa aacccctcaa gacccgttta gaggccccaa
240 260
caccccgttta ttcagcaaaa aacccctcaa gacccgttta gaggccccaa
260 280
caccccgttta ttcagcaaaa aacccctcaa gacccgttta gaggccccaa
280 300
caccccgttta ttcagcaaaa aacccctcaa gacccgttta gaggccccaa
300 320
caccccgttta ttcagcaaaa aacccctcaa gacccgttta gaggccccaa
320 340
caccccgttta ttcagcaaaa aacccctcaa gacccgttta gaggccccaa
340 360
caccccgttta ttcagcaaaa aacccctcaa gacccgttta gaggccccaa
360 380
caccccgttta ttcagcaaaa aacccctcaa gacccgttta gaggccccaa
380 400
caccccgttta ttcagcaaaa aacccctcaa gacccgttta gaggccccaa
400 420
caccccgttta ttcagcaaaa aacccctcaa gacccgttta gaggccccaa
420 440
caccccgttta ttcagcaaaa aacccctcaa gacccgttta gaggccccaa
440 460
caccccgttta ttcagcaaaa aacccctcaa gacccgttta gaggccccaa
460 480
caccccgttta ttcagcaaaa aacccctcaa gacccgttta gaggccccaa
480 500
caccccgttta ttcagcaaaa aacccctcaa gacccgttta gaggccccaa
500 520
caccccgttta ttcagcaaaa aacccctcaa gacccgttta gaggccccaa
520 540
caccccgttta ttcagcaaaa aacccctcaa gacccgttta gaggccccaa
540 560
caccccgttta ttcagcaaaa aacccctcaa gacccgttta gaggccccaa
560 580
caccccgttta ttcagcaaaa aacccctcaa gacccgttta gaggccccaa
580 600
caccccgttta ttcagcaaaa aacccctcaa gacccgttta gaggccccaa
600 620
caccccgttta ttcagcaaaa aacccctcaa gacccgttta gaggccccaa
620 640
caccccgttta ttcagcaaaa aacccctcaa gacccgttta gaggccccaa
640 660
caccccgttta ttcagcaaaa aacccctcaa gacccgttta gaggccccaa
660 680
caccccgttta ttcagcaaaa aacccctcaa gacccgttta gaggccccaa
680 700
caccccgttta ttcagcaaaa aacccctcaa gacccgttta gaggccccaa
700 720
caccccgttta ttcagcaaaa aacccctcaa gacccgttta gaggccccaa
720 740
caccccgttta ttcagcaaaa aacccctcaa gacccgttta gaggccccaa
740 760
caccccgttta ttcagcaaaa aacccctcaa gacccgttta gaggccccaa
760 780
caccccgttta ttcagcaaaa aacccctcaa gacccgttta gaggccccaa
780 800
caccccgttta ttcagcaaaa aacccctcaa gacccgttta gaggccccaa
800 820
caccccgttta ttcagcaaaa aacccctcaa gacccgttta gaggccccaa
820 840
caccccgttta ttcagcaaaa aacccctcaa gacccgttta gaggccccaa
840 860
caccccgttta ttcagcaaaa aacccctcaa gacccgttta gaggccccaa
860 880
caccccgttta ttcagcaaaa aacccctcaa gacccgttta gaggccccaa
880 900
caccccgttta ttcagcaaaa aacccctcaa gacccgttta gaggccccaa
900 920
caccccgttta ttcagcaaaa aacccctcaa gacccgttta gaggccccaa
920 940
caccccgttta ttcagcaaaa aacccctcaa gacccgttta gaggccccaa
940 960
caccccgttta ttcagcaaaa aacccctcaa gacccgttta gaggccccaa
960 980
caccccgttta ttcagcaaaa aacccctcaa gacccgttta gaggccccaa
980 1000
caccccgttta ttcagcaaaa aacccctcaa gacccgttta gaggccccaa
1000 1020
caccccgttta ttcagcaaaa aacccctcaa gacccgttta gaggccccaa
1020 1040
caccccgttta ttcagcaaaa aacccctcaa gacccgttta gaggccccaa
1040 1060
caccccgttta ttcagcaaaa aacccctcaa gacccgttta gaggccccaa
1060 1080
caccccgttta ttcagcaaaa aacccctcaa gacccgttta gaggccccaa
1080 1100
caccccgttta ttcagcaaaa aacccctcaa gacccgttta gaggccccaa
1100 1120
caccccgttta ttcagcaaaa aacccctcaa gacccgttta gaggccccaa
1120 1140
caccccgttta ttcagcaaaa aacccctcaa gacccgttta gaggccccaa
1140 1160
caccccgttta ttcagcaaaa aacccctcaa gacccgttta gaggccccaa
1160 1180
caccccgttta ttcagcaaaa aacccctcaa gacccgttta gaggccccaa
1180 1200
caccccgttta ttcagcaaaa aacccctcaa gacccgttta gaggccccaa
1200 1220
caccccgttta ttcagcaaaa aacccctcaa gacccgttta gaggccccaa
1220 1240
caccccgttta ttcagcaaaa aacccctcaa gacccgttta gaggccccaa
1240 1260
caccccgttta ttcagcaaaa aacccctcaa gacccgttta gaggccccaa
1260 1280
caccccgttta ttcagcaaaa aacccctcaa gacccgttta gaggccccaa
1280 1300
caccccgttta ttcagcaaaa aacccctcaa gacccgttta gaggccccaa
1300 1320
caccccgttta ttcagcaaaa aacccctcaa gacccgttta gaggccccaa
1320 1340
caccccgttta ttcagcaaaa aacccctcaa gacccgttta gaggccccaa
1340 1360
caccccgttta ttcagcaaaa aacccctcaa gacccgttta gaggccccaa
1360 1380
caccccgttta ttcagcaaaa aacccctcaa gacccgttta gaggccccaa
1380 1400
caccccgttta ttcagcaaaa aacccctcaa gacccgttta gaggccccaa
1400
tcagattac cagttactct ctctccttctt aaagtttaacc 1500
aaattttatc tctcggtttg gctcagact cctttaaggc tctttgctctt tttaaggc 1560
aatattcctct cctcgatagt gctttaaaact tttaaggc tctttaaggc tctttgctctt 1620
gacattccag gacagttgaac cacttactct cagtttaac gctttaaaact tttaaggc 1680
cctttggtct cctcttctcctt cggatggc tttctctcctt tttctctcctt cggatggc 1740
cacatttca gacattttt tttctcttctt cgttctcttctt tttctcttctt cgttctcttctt 1800
gactttttaa tattctattct ctttcttctt ctttcttctt ctttcttctt ctttcttctt 1860
gacagcttcc acacagcagc acacagcagc acacagcagc acacagcagc acacagcagc 1920
gacaggttgt cctttttttttt ctttcttctt ctttcttctt ctttcttctt ctttcttctt 1980
gaccttttctt ctttcttctt ctttcttctt ctttcttctt ctttcttctt ctttcttctt 2040
gacgcacggt gacacagctt ctttcttctt ctttcttctt ctttcttctt ctttcttctt 2100
gacgcacggt gacacagctt ctttcttctt ctttcttctt ctttcttctt ctttcttctt 2160
gacgcacggt gacacagctt ctttcttctt ctttcttctt ctttcttctt ctttcttctt 2220
gacgcacggt gacacagctt ctttcttctt ctttcttctt ctttcttctt ctttcttctt 2280
gacgcacggt gacacagctt ctttcttctt ctttcttctt ctttcttctt ctttcttctt 2340
gacgcacggt gacacagctt ctttcttctt ctttcttctt ctttcttctt ctttcttctt 2400
gacgcacggt gacacagctt ctttcttctt ctttcttctt ctttcttctt ctttcttctt 2460
gacgcacggt gacacagctt ctttcttctt ctttcttctt ctttcttctt ctttcttctt 2520
gacgcacggt gacacagctt ctttcttctt ctttcttctt ctttcttctt ctttcttctt 2580
gacgcacggt gacacagctt ctttcttctt ctttcttctt ctttcttctt ctttcttctt 2640
gacgcacggt gacacagctt ctttcttctt ctttcttctt ctttcttctt ctttcttctt 2700
gacgcacggt gacacagctt ctttcttctt ctttcttctt ctttcttctt ctttcttctt 2760
gacgcacggt gacacagctt ctttcttctt ctttcttctt ctttcttctt ctttcttctt 2820
gacgcacggt gacacagctt ctttcttctt ctttcttctt ctttcttctt ctttcttctt 2880
gacgcacggt gacacagctt ctttcttctt ctttcttctt ctttcttctt ctttcttctt 2940
gacgcacggt gacacagctt ctttcttctt ctttcttctt ctttcttctt ctttcttctt 3000
gacgcacggt gacacagctt ctttcttctt ctttcttctt ctttcttctt ctttcttctt 3060
gacgcacggt gacacagctt ctttcttctt ctttcttctt ctttcttctt ctttcttctt 3120
gacgcacggt gacacagctt ctttcttctt ctttcttctt ctttcttctt ctttcttctt 3180
gacgcacggt gacacagctt ctttcttctt ctttcttctt ctttcttctt ctttcttctt 3240
gacgcacggt gacacagctt ctttcttctt ctttcttctt ctttcttctt ctttcttctt 3300
gacgcacggt gacacagctt ctttcttctt ctttcttctt ctttcttctt ctttcttctt 3360
gacgcacggt gacacagctt ctttcttctt ctttcttctt ctttcttctt ctttcttctt 3420
gacgcacggt gacacagctt ctttcttctt ctttcttctt ctttcttctt ctttcttctt 3480
gacgcacggt gacacagctt ctttcttctt ctttcttctt ctttcttctt ctttcttctt 3540
gacgcacggt gacacagctt ctttcttctt ctttcttctt ctttcttctt ctttcttctt 3600
gacgcacggt gacacagctt ctttcttctt ctttcttctt ctttcttctt ctttcttctt 3660
gacgcacggt gacacagctt ctttcttctt ctttcttctt ctttcttctt ctttcttctt 3720
gacgcacggt gacacagctt ctttcttctt ctttcttctt ctttcttctt ctttcttctt 3780
gacgcacggt gacacagctt ctttcttctt ctttcttctt ctttcttctt ctttcttctt 3840
gacccaaacg gaaaaaccg cccuttaacat ggcoccgctt tttcagacgct agaaggtaacc 3900
gcttcttgag aacactacacg gctctgagcg ggctgaacgct ggtggagagcg 3960
tgcgaacgaat ctgcttgtgc gggtgaacgcc caatgccttt acgtgcttatct 4020
aaccctgcta ccacatgaacat cgccttttc tgtcttgtctt cgctgctcgc 4080
ggctgaacgct gccctgcagc tcgggttgacgcc ggtggagagcg 4140
gaccaggctt caaaggtgac gcggtgcttc agactgatgcgt acagctgttat 4200
attgtacgr ggctggaacc cactagcagt gcgtgaacgcc gcagctgttat 4260
aacattcgc gggaggcctt ttcgctcgtgc cctgtgatga cgggtgttggc 4320
gggctgctct ccgcttactt acggctgtgc cctgtgatga cgggtgttggc 4380
ggcctgctct ccgcttcgat ccttggtctt tcgcttcgat ccttggtctt 4440
agcgcgcttg gcggcctgtt tcgccggtgt gcggcctgtt tcgccggtgt 4500
agcgcgcttg gcggcctgtt tcgccggtgt gcggcctgtt tcgccggtgt 4560
tcgaggaa cctgtgtgtg ggcgcggtgt gcggcctgtt tcgccggtgt 4620
ctttttcctc tcgggaacgc gcggcctgtt tcgccggtgt gcggcctgtt 4680
cgggtggctt gtcggcgttg ggcgcggtgt gcggcctgtt tcgccggtgt 4740
getgtgtgtt atccgtggttc agtggcgtgg gcggcctgtt tcgccggtgt 4800
cctttcctccc gcgtctggct gcgtctggct gcgtctggct gcgtctggct 4860
agtcgcggtc ctctggctgc gcgtctggct gcgtctggct gcgtctggct 4920
cgtgtgtgt gcgcgccttt cccgctggac ggcgcggtgt gcggcctgtt 4980
cgcgtgctg gtcggcgttg gcgcgccttt cccgctggac ggcgcggtgt 5040
gcgcgccttt cccgctggac ggcgcggtgt gcggcctgtt tcgccggtgt 5100
cgcgcgccttt cccgctggac ggcgcggtgt gcggcctgtt tcgccggtgt 5160
agcctgggttt gccgcggtgt gcgcgccttt cccgctggac ggcgcggtgt 5220
ctgctgcgtt gccgcggtgt gcgcgccttt cccgctggac ggcgcggtgt 5280
gcgcgccttt cccgctggac ggcgcggtgt gcggcctgtt tcgccggtgt 5340
agcgcgcttg gcgcgccttt cccgctggac ggcgcggtgt gcggcctgtt 5400
ttcctggaga aggcaggttt ttcctggaga aggcaggttt ttcctggaga 5460
ttcctggaga aggcaggttt ttcctggaga aggcaggttt ttcctggaga 5520
ggcgcggtgt gcgcgccttt cccgctggac ggcgcggtgt gcggcctgtt 5580
atggctttgc ttc ggctggcat ctctggctgc gcgtctggct gcgtctggct 5640
agcgcggtc ctctggctgc gcgtctggct gcgtctggct gcgtctggct 5700
gcgtgctt gcgcgccttt cccgctggac ggcgcggtgt gcggcctgtt 5760
ggcgcggtgt gcgcgccttt cccgctggac ggcgcggtgt gcggcctgtt 5820
ggcgcggtgt gcgcgccttt cccgctggac ggcgcggtgt gcggcctgtt 5880
ggcgcggtgt gcgcgccttt cccgctggac ggcgcggtgt gcggcctgtt 5940
ttcctgcgta aggcaggttt ttcctggaga aggcaggttt ttcctggaga 6000
ttcctgcgta aggcaggttt ttcctggaga aggcaggttt ttcctggaga 6060
ttcctgcgta aggcaggttt ttcctggaga aggcaggttt ttcctggaga 6120
ttcctgcgta aggcaggttt ttcctggaga aggcaggttt ttcctggaga 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 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 Glu 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 Gln 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 Leu Leu Pro Phe Glu Asp Leu Pro Leu
115 120 125
Thr Trp Glu Glu Ile Pro Ala Leu Asp Lys Leu Lys Ala Lys Gly
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 Asp Ile Asp Lys Asp Lys Asp
180 185 190
Leu Thr Phe Leu Val Asp Lys Ile Lys Asn Lys 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 Asn Tyr Gly Val Thr Val Leu Pro Thr Phe Lys Gly Gln Pro Ser
245 250 255
Lys Pro Phe Val Gly Val Leu Ser Ala Gly Ile Asn Ala Ala Asp Pro
260 265 270
Asn Lys Glu Leu Ala Lys Gly Phe Leu Glu Asn Tyr Lys Leu Thr Asp
275 280 285
Glu Gly Leu Glu Ala Val Asp Asp Lys Pro Lys Glu Ala Val Ala

---continued---
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 HPQ, HPM, HPN, or HQP motif;
2. 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 the amino acid sequence of said peptide does not contain an HPQ, HPM, HPN, or HQP motif;
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.
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 HPQ, 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 80% identity to the first 38 amino acids of 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.