(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
14 January 2010 (14.01.2010)

PCT

(10) International Publication Number
WO 2010/004338 A2

(51) International Patent Classification:
C07K 1/22 (2006.01) *C07K 14/33* (2006.01)

(21) International Application Number:
PCT/GB2009/050817

(22) International Filing Date:
9 July 2009 (09.07.2009)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
0812651.8 10 July 2008 (10.07.2008) GB

(71) Applicant *(for all designated States except US)*: **THE UNIVERSITY OF SHEFFIELD** [GB/GB]; Western Bank, Sheffield Yorkshire S10 2TN (GB).

(72) Inventor; and

(75) Inventor/Applicant *(for US only)*: **HORNBY, David Peter Joseph** [GB/GB]; 4 Downing Row, Hathersage, Hope Valley S32 1AT (GB).

(74) Agent: **HARRISON GODDARD FOOTE**; 106 Micklegate, York Yorkshire YO1 6JX (GB).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

— *of inventorship (Rule 4.17(iv))*

Published:

— *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*

(54) Title: AFFINITY TAGS

Figure 1

SEQ ID NO:1

MQSNKTFNLEKQNHTPRKHHQHHHQQQHHQQQQQQPP
PPPIPANGQQASSQNEGLTIDLK

(57) Abstract: The invention relates to an affinity tag comprising a mosaic of a first polypeptide tag and a second polypeptide tag.

WO 2010/004338 A2

Affinity Tags

The present invention relates to affinity tags. In particular the invention relates to affinity tags possessing affinity for two distinct capture reagents and methods of protein purification using said tags.

BACKGROUND

Bacteria such as *E.coli* encode around 3,000 proteins, whilst a human cell is capable of expressing at least 30,000 different polypeptides and with post-translational modification taken into account, possibly 10 to 100 fold more. At any given time and at any given location in the body, there may be only a subset of these proteins expressed, but nonetheless, it is clear that there are considerable challenges facing the biochemist in purifying a single protein species from a cell. In addition, it is important to appreciate that the expression levels (or abundance) of proteins expressed can vary by over 1000-fold. The most demanding of protein purification experiments would therefore involve the purification of a low abundance protein from a cell extract, where it may be necessary to enrich the target protein by 20,000 fold.

Proteins at high abundance, such as haemoglobin and serum albumin, were originally purified by solvent extraction and salt fractionation: such methods are too crude for the purification of the majority of cellular proteins, particularly when protein activity needs to be retained. During the 1950s and '60s, two modes of liquid chromatography were introduced by Pharmacia: Gel filtration, in which proteins are separated on the basis of their molecular sizes, provides one means of purifying proteins, but perhaps the most versatile and robust method is ion-exchange chromatography, in which polypeptides are separated on the basis of differences in surface charge.

These methods have proved invaluable for purifying many proteins of medium abundance and in particular those possessing extremes of charge (e.g. the highly basic histones) or size (e.g. myoglobin), but are rarely up to the task of purifying low abundance proteins (with some notable exceptions from the Kornberg laboratory (Rigaut G, et al (1999) Nature Biotechnology 17, 1030-1032).

It was the introduction of small ligand based affinity chromatography at approximately the same time as recombinant DNA technology that made possible the purification of low

abundance proteins from virtually any organism or cell type. By raising the level of expression of low abundance proteins, using recombinant methods, such proteins can be brought within the scope of most existing purification methods.

Recombinant expression systems in *E.coli,* capable of producing protein levels approaching 20% of the total cell protein, in conjunction with the rapid selectivity afforded by affinity chromatography has remained an enduring strategy for the purification of many proteins and has transformed modern biochemistry. Thus single polypeptides that require 1,000-10,000 fold enrichment from a natural source can now often be isolated as 1-10% of the total cell protein and is often via a single affinity chromatography step.

Latterly, developments in fusion protein technology, in which "universal" affinity tags (such as the poly Histidine tag or the Glutathione-S-transferase tag) have been added to target proteins through molecular genetics, have meant that affinity chromatography can be applied to virtually any target protein. Poly-Histidine, typically in the form of a hexa-His tag, appended to the N or C-terminus of a recombinant polypeptide (using standard molecular biology procedures), has been widely used to purify recombinant proteins using immobilised metal ion chromatography (IMAC). Immobilised Nickel is coordinated by the tag and subsequently displaced by an excess of imidazole in order to enrich the corresponding "tagged" recombinant species.

Whilst raising the abundance of a protein by genetic means is suitable for most single polypeptide proteins, it is unsuitable for the majority of protein complexes. The expression of multi-component protein assemblies is difficult, since the level of expression of each component needs to be balanced in order to retain the original proportions of each component. With the discovery that more and more proteins act in collaboration with a range of partners, there ensued a drive to develop technologies for their purification and analysis.

The use of site-specific recombination, a key component of TAP (tandem affinity purification) tagging, is used to eliminate the wild type copy of a target protein coding gene. This, coupled with the fusion to the target gene, a sequence encoding two high affinity protein binding domains (the Protein A domain, which interacts with IgG and the Calmodulin binding peptide which interacts with Calmodulin and which make stronger interactions than the small ligands used in earlier forms of affinity chromatography),

separated by a protease sensitive site, has greatly facilitated the isolation of many protein complexes. Using *Saccharomyces cerevisiae* as the test-bed organism, Seraphin's group at Heidelberg (Rigaut G, et al (1999) Nature Biotechnology 17, 1030-1032) successfully compensated for the low abundance of most tagged protein complexes by the use of highly selective tags and the elimination of the (competing) wild type polypeptide. In this way the issue of abundance is simply overcome by raising the volume of the starting engineered yeast strain. It is now possible, using this approach, to purify many yeast protein complexes and to determine their composition using mass spectrometry.

In the hands of experienced protein chemists, the TAP tagging method can be applied to the yeast proteome very successfully. However, whilst the method is intrinsically sound, it has a number of drawbacks, in particular, the size of the TAP tag can be in the region of 21kDa and the larger a tag is the greater the possibility that it will interfere with the function or structure of the polypeptide to be isolated, or alternatively be cleaved or digested by the polypeptide target.

Accordingly, there remains a need for an alternative affinity tag.

BRIEF SUMMARY OF THE DISCLOSURE

According to a first aspect, the invention provides an affinity tag comprising a copolymer of a first monomeric unit and a second monomeric unit, wherein each of said first and second monomeric units has affinity for a distinct capture reagent, provided that said copolymer is not a block copolymer.

Preferably each of said first and second monomeric units are amino acid residues.

Preferably, said first monomeric unit has affinity for an immobilized metal affinity chromatography resin. More preferably, said immobilized ion is selected from the group consisting of Co^{2+}, Cu^{2+}, Ni^{2+}, Zn^{2+}, Ca^{2+} and Fe^{3+}. Still more preferably, said immobilised ion is Ni^{2+}. Preferably, said first monomeric unit is His.

Preferably, said second monomeric unit has affinity for an ion exchange resin, for example an anion exchange resin or a cation exchange resin. Preferably, second monomeric unit is an acidic amino acid, such as Arg or Lys. Alternatively, said second monomeric unit is Asp or Glu.

Alternatively said second monomeric unit is a basic amino acid, such as Arg or Lys.

Preferably, said first monomeric unit is His and said second monomeric unit is Arg.

In a further aspect, the invention provides a method of purifying a recombinant protein comprising:

a) preparing a peptide comprising an affinity tag as herein described;
b) purifying the peptide using a first purification step comprising binding the affinity tag to a first capture reagent selective for the affinity tag; and
c) further purifying the peptide of b) using a second purification step comprising binging the affinity tag to a second capture reagent selective for the affinity tag, wherein said second capture reagent is distinct from said first capture reagent.

Preferably, said first purification step comprises immobilized metal ion affinity chromatography.

Preferably, said first purification step comprises immobilized metal ion affinity chromatography and said second step comprises ion exchange chromatography.

Alternatively, said first purification step comprises ion exchange chromatography.

Preferably, said first purification step comprises ion exchange chromatography and said second step comprises immobilized metal ion affinity chromatography. When the purification step comprises immobilized metal ion affinity chromatography said capture reagent is an immobilized metal ion affinity chromatography resin. Preferably, said immobilized ion is selected from the group consisting of Co^{2+}, Cu^{2+}, Ni^{2+}, Zn^{2+}, Ca^{2+} and Fe^{3+}.

When the purification step comprises ion exchange chromatography, said capture reagent is ion exchange resin, more preferably an anion exchange resin or a cation exchange resin.

In a further aspect the invention provides, use of an affinity tag according to the invention for the purification of a recombinantly expressed polypeptide.

In a further aspect the invention provides a recombinant protein comprising an affinity tag according to the invention.

Preferably, the recombinant protein is cleavably linked to said affinity tag, either to the C-terminal end of said polypeptide or to the N-terminal end of said polypeptide.

In a further aspect the invention provides an isolated nucleic acid molecule encoding a protein according to the invention.

In a further aspect the invention provides an expression vector comprising a nucleic acid molecule according to the invention.

In a further aspect the invention provides a host cell comprising an expression vector according to the invention. The host cell may be a eukaryotic cell or a prokaryotic cell.

Throughout the description and claims of this specification, the words "comprise" and "contain" and variations of the words, for example "comprising" and "comprises", means "including but not limited to", and is not intended to (and does not) exclude other moieties, additives, components, integers or steps.

Throughout the description and claims of this specification, the singular encompasses the plural unless the context otherwise requires. In particular, where the indefinite article is used, the specification is to be understood as contemplating plurality as well as singularity, unless the context requires otherwise.

Features, integers, characteristics, compounds, chemical moieties or groups described in conjunction with a particular aspect, embodiment or example of the invention are to be understood to be applicable to any other aspect, embodiment or example described herein unless incompatible therewith.

BRIEF DESCRIPTION OF THE DRAWINGS

The invention will now be described, by way of example only, with reference to the accompanying drawings, in which:

Figure 1 provides the original amino acid sequence of an N-terminus p54 – tag (SEQ ID NO:1);

Figure 2 provides the amino acid sequence of a mosaic His/Arg tag of the invention (SEQ ID NO:2);

Figure 3a provides the nucleotide sequence (top oligo) encoding the His/Arg tag of SEQ ID NO:2 (SEQ ID NO:3), and figure 3b provides the nucleotide sequence (bottom oligo) encoding the His/Arg tag of SEQ ID NO:2 (SEQ ID NO:4);

Figure 4 is a schematic representation of the His/Arg tag construct as annealed in pQIS207, illustrating the translation of the encoding sequence and varies restriction sites;

Figure 5 is a restriction digest of pQSI207 with *NcoI* and *NdeI* to cut and recover the vector from the gel (lanes 1&2), a DNA marker hyperladder (Bioline) (lane 3) and annealed oligos for insertion into the vector (lane 4);

Figure 6 shows a restriction digest of the pQSI207 construct with *Sal1* and *Hinc11;*

Figure 7 is a schematic representation of the His/Arg affinity tag in relation to the target polypeptide, Rubridoxin, within the pQSI207 vector;

Figure 8 provides the amino acid sequence of the pQIS 216 translation product comprising the affinity tag and target polypeptide (SEQ ID NO:5);

Figure 9 is an SDS PAGE illustrating induction and expression of the affinity tagged polypeptide. Lane 1 shows a cell extract prior to induction of the tagged Rubredoxin protein, Lane 2 shows the induced protein (boxed) and finally Lane 3 contains a Molecular Weight marker sample.

Figure 10 illustrates the fractions (1 to 10) obtained following Ni-NTA purification and elution, tube 6 contains the peak sample (indicated by the dark red colour);

Figure 11 is an SDS PAGE illustrating size separation of fractions 2, 3, 4, 5, 6 and 7 of figure 10; and

Figure 12 is an SDS PAGE illustrating ion exchange purified fraction 6 of figure 10 following SP-Sepharose separation (Lanes 3-6, circled), Lane 1 contains a sample of independently purified Rubredoxin.

DETAILED DESCRIPTION

The present inventors have identified a single affinity tag that surprisingly enables a target protein to be purified via two complementary separation methods. The inventors have discovered that the monomeric units of two distinct homopolymeric protein affinity tags may be copolymerised, not as block co-polymers, so as to form a single affinity tag. The resulting affinity tag retains the binding affinity of each of the distinct homopolymeric tags. Accordingly, the single affinity tag of the present application exhibits affinity for two distinct capture reagents, thereby facilitating a two step purification strategy whereby the target protein is sequentially purified using different properties of the single tag.

The use of a single affinity tag offers many advantages. Use of a single tag reduces the likelihood of interference with protein function. Furthermore, genetic incorporation of the sequence encoding the tag is simplified compared with the standard TAP tags.

In addition, use of the tag of the present invention can eliminate the need for a protease step as part of a two step purification strategy.

As used herein, the term “affinity tag” refers to a polypeptide sequence, which has affinity for a specific capture reagent and which can be separated from a pool of proteins and thus purified on the basis of its affinity for the binding partner. The affinity tag may form part of a target peptide sequence to be purified. Where the affinity tag forms part of the target polypeptide, the tag remains part of the target following purification and cannot be cleaved therefrom. In such circumstances the affinity tag does not affect the properties of the target polypeptide.

Alternatively, the affinity tag may be attached to the target peptide, for example the affinity tag may be covalently linked to the target peptide via an amino acid linker. The amino acid linker may comprise a cleavage site to enable detachment of the affinity tag from the target protein. Selective cleavage sites and methods of cleaving such sequences are well known in the art. The affinity tag may be attached to the target peptide at the C-terminal or the N-terminal end thereof.

An affinity tag in accordance with the present invention comprises or consists of a copolymer of a first monomeric unit and a second monomeric unit, wherein said first and second monomeric units are distinct. The resultant copolymer is not a block copolymer.

Preferably said first and second monomeric units are amino acid residues. Preferably, each of said first and second monomeric units can be formed into a first and second homopolymeric affinity tag respectively.

As used here in the term "copolymer" refers to a polymer derived from two or more monomeric species or units. Preferably the copolymer is a polypeptide and the monomeric units are amino acid residues. The copolymers of the invention may be classified on how the constituent monomeric units are arranged. The copolymer may be an alternating copolymer, with regular alternating monomeric units, for example:

-X-Y-X-Y-X-Y-X-Y-X-Y-

The copolymer may be a statistical copolymer, in which the occurrence of the monomeric units within the copolymer obeys known statistical rules. The copolymer may be a periodic copolymer, with the monomeric units are arranged in a repeating fashion, for example:

$(-X-Y-X-X-Y-Y-Y-X-X-Y-X-X)_n$

The copolymer may be a random copolymer with a random sequence of monomeric units, for example:

-X-Y-X-X-Y-Y-X-X-Y-

The copolymer is not a block copolymer, comprised of two or more homopolymeric subunits, for example:

-X-X-X-X-X-Y-Y-Y-Y-Y-

As used herein, the terms "homopolymer" and "homopolymeric affinity tag" refer to an affinity tag consisting of monomeric amino acid units of a single amino acid. Examples of such homopolymeric tags include polyHis-tags, polyArg-tags, polyCys-tags, polyPhe-tags, polyLys-tags, polyGlu-tags and polyAsp-tags.

As used herein, the term "monomeric unit" refers to a monomer from which a copolymer is formed. Preferably the monomeric units are amino acid residues having affinity for a capture reagent. Examples of monomeric units include Histidine, Arginine, Cysteine, Phenylalanine, Lysine, Glutamic acid and Aspartic acid residues.

Preferably, the copolymer comprises at least 3 monomeric units of a first monomeric unit. More preferably, the copolymer comprises at least 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25 or 30 monomeric units of a first monomeric unit.

Preferably, the copolymer comprises at least 3 monomeric units of a second monomeric unit. More preferably, the copolymer comprises at least 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25 or 30 monomeric units of a second monomeric unit.

Preferably, said first monomeric unit is a His and said second monomeric unit is an Arg and the resultant copolymer consists of histidine and arginine amino acid residues.

Preferably, said first monomeric unit is a His and said second monomeric unit is a Glu and the resultant copolymer consists of histidine and glutamic acid amino acid residues.

Preferably, said first monomeric unit is a His and said second monomeric unit is an Asp and the resultant copolymer consists of histidine and aspartic acid amino acid residues.

Preferably, said first monomeric unit is a His and said second monomeric unit is an Lys and the resultant copolymer consists of histidine and lysine amino acid residues.

Preferably, said first monomeric unit is a His and said second monomeric unit is a Cys and the resultant copolymer consists of histidine and cystine amino acid residues.

Preferably, said first monomeric unit is a His and said second monomeric unit is a Phe and the resultant copolymer consists of histidine and phenylalanine amino acid residues.

Other copolymer combinations of amino acid monomer are within the scope of the present invention

The affinity tag of the present application has substantially the same affinities as each of said first homopolymeric affinity tag and said second homopolymeric affinity tag from which it is derived.

As used herein, the term “target polypeptide” refers to a polypeptide of interest to be isolated using the affinity tag of the present application. A person of skill in the art would be aware of the recombinant molecular biological techniques required to expand the target polypeptide by several amino acids so as to attach thereto or incorporate therein an affinity tag of the invention to a target polypeptide. As such the affinity tags of the present invention may be used to purify a large range of recombinantly produced proteins.

Accordingly, the present application also relates to the provision of a recombinant protein comprising the affinity tag of the invention. Preferably, the recombinant protein is produced by expression of a nucleic acid corresponding to the protein in a suitable expression vector.

Accordingly, the present application also relates to the provision of an isolated nucleic acid molecule that encodes for a recombinant protein comprising the affinity tag of the invention. The nucleic acid may be a DNA (i.e. a cDNA) or RNA (i.e. an mRNA). The nucleic acid may be an analog of the DNA or RNA generated, e.g., by the use of nucleotide analogs. The nucleic acid molecules may be used to produce a target protein labeled with an affinity tag. The sequence of the nucleic acid molecule may be designed so as to link a target polypeptide to an affinity tag. In addition, the nucleic acid molecule may be engineered to introduce a cleavage site between the affinity tag and the target polypeptide.

The nucleic acid molecule may be incorporated into an expression vector for expression in a host cell.

A person of skill in the art will be aware of the molecular techniques available for the preparation of expression vectors.

Accordingly, the present application also relates to the provision of an expression vector containing a nucleic acid molecule of the invention.

As used herein, the term “vector” refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been operably linked. The vector can be capable of autonomous replication or it can integrate into a host DNA. The vector may include restriction enzyme sites for insertion of recombinant DNA and may include one or more selectable markers, together with regulatory sequences, such as promoters. The vector can be a nucleic acid in the form of a plasmid, a bacteriophage or a cosmid. Preferably the vector is suitable for bacterial expression, e.g. for expression in *E. coli.*

The design of the expression vector depends on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, and the like.

Preferably the vector comprises those genetic elements which are necessary for purification of a target polypeptide from a host cell. The elements required include a promoter, a coding region for the target polypeptide incorporating or linked to the affinity tag of the invention and a transcriptional terminator.

The nucleic acid molecule for incorporation into the expression vector of the invention, as described above, can be prepared by synthesizing nucleic acid molecules using mutually priming oligonucleotides and the nucleic acid sequences described herein.

In order to isolate and/or modify the nucleic acid molecule of interest for insertion into the chosen plasmid, it is preferable to use PCR. Appropriate primers for use in PCR preparation of the sequence can be designed to isolate the required coding region of the nucleic acid molecule and add a region encoding an affinity tag of the invention.

In a preferred embodiment a nucleic acid molecule for incorporation into an expression vector of the invention, is prepared by the use of the polymerase chain reaction as disclosed by Saiki et al (1988) *Science* 239, 487-491, using appropriate oligonucleotide primers. The coding region is amplified, whilst the primers themselves become incorporated into the amplified sequence product. In a preferred embodiment the amplification primers contain restriction endonuclease recognition sites which allow the amplified sequence product to be cloned into an appropriate vector.

The expression vectors of the invention can contain a single copy of the nucleic acid molecule described previously, or multiple copies of the nucleic acid molecule described previously.

The expression vectors of the invention can be introduced into host cells to thereby produce target polypeptides comprising affinity tags, encoded by nucleic acids as described herein. Accordingly, the present application also relates to the provision of a host cell for expression of a target polypeptide comprising the affinity tag of the invention, said cell comprising an expression vector, comprising a nucleic acid molecule described herein.

The affinity tag of the invention facilitates isolation and purification of the target polypeptide from the cell or from the surrounding medium should sufficient polypeptide

be secreted. The expressed polypeptides may be separated from the cells by conventional methods, including separating the cells from the medium via filtration or centrifugation. The separated cells can then be disrupted, by for example cell lysis and the polypeptide isolated and purified from the lysate.

Accordingly, the present invention also provides a method of purifying a recombinant target polypeptide.

High purification of a target polypeptide comprising an affinity tag of the invention is achieved using a tandem affinity strategy. The use of a two step strategy greatly reduces non-specific background binding and thus guarantees highly pure samples.

The single affinity tag of the present invention possesses affinity for two distinct capture reagents. Accordingly, a target polypeptide comprising the tag of the invention may be purified using a first purification step selective for the affinity tag; and
a second purification step, selective for the affinity tag, wherein said second purification step is distinct from said first purification step.

The target polypeptide comprising the affinity tag may be obtained using any suitable method, preferably a method as herein described.

As used herein, the term “capture reagent” refers to any agent that preferentially binds to an affinity tag of the present invention. Any suitable capture reagent can be used in the method of the present invention. The properties of the affinity tag of the invention will necessarily determine the choice of capture reagent used. A person of skill in the art will be aware of the how the binding properties of the affinity tag will dictate the selection of capture reagent. The affinity tag of the invention will possess the unique binding affinities of each of the first and second monomeric units of which the affinity tag is a mosaic, allowing the tag to bind to two capture reagents.

As used herein, the term "binding" or “binds” refers to an interaction between the affinity tag and capture reagent, such that the tag is immobilised and retained by the reagent. Immobilization may be as a result of any physical or chemical mechanism, for example immobilization by bonding, such as ionic bonding, polar bonding, hydrogen bonding or co-ordinate bonding, or by way of covalent interaction, ionic or electrostatic interactions, or any combination thereof.

The first and / or second purification step may comprise ion exchange chromatography. For example, but not by way of limitation, the capture reagent may be an ion exchange matrix, for example an anion or a cation exchange matrix. The amino acid composition of the affinity tag of the invention can significantly alter the overall charge and pI of the target polypeptide and thus serve to distinguish the target form background proteins. This distinguishing feature can be used to isolate the target protein comprising the affinity via a combination of adsorption of the affinity tag to an ion exchange matrix and subsequent elution of the bound protein by alteration of the salt gradient pH. The pH or salt gradient required to elute a bound target polypeptide can be determined by a person of skill in the art via routine experimentation.

Alternatively, the first and / or second purification step may comprise affinity chromatography. For example, but not by way of limitation, the capture reagent may be an immobilised metal affinity chromatography resin, for example an immobilised transition metal affinity chromatography resin. Preferably the immobilised ion is Co^{2+}, Cu^{2+}, Ni^{2+}, Zn^{2+}, Ca^{2+} or Fe^{3+}. Suitable resins for the immobilisation of ions include Iminodiacetic acid (chelating sepharose, Amersham Biosciences), nitrilotriacetic acid agarose (Ni-NTA resin, Qigen) or carboxymethylaspartate agarose (Talon resin, Clonetech).

The incorporation in the affinity tag, which consists of amino acid residues or polypeptides having high affinity for, and thus high binding strength to, the aforementioned capture reagents, serve to distinguish the affinity tag and accordingly the target polypeptide, form other host cell proteins. This distinguishing feature allows specific binding between the tag and the capture reagent and can be used to purify the target protein comprising the affinity via a combination of binding of the affinity tag an affinity matrix and subsequent elution of the bound protein.

Preferably, when said first or second homopolymeric affinity tag is a polyHis-tag one of said first or second purification steps utilizes an immobilised metal affinity chromatography resin as a capture reagent, more preferably an Ni-IMAC. Preferably the purification step utilizes imidazole or low pH as an elution agent.

Preferably, when said first or second homopolymeric affinity tag is a polyArg-tag one of said first or second purification steps utilizes a cation exchange resin as a capture

reagent, for example SP Sepharose. Preferably the purification step utilizes a high pH salt gradient as an elution agent.

Preferably, when said first or second monomeric unit is an Asp one of said first or second purification steps utilizes an anion exchange resin as a capture reagent. Preferably the purification step utilizes a low-neutral pH salt gradient as an elution agent.

Preferably, when said first or second monomeric unit is a Cys one of said first or second purification steps utilizes a thiopropyl sepharose, such as DTT or beta ME, as capture reagent. Preferably the purification step utilizes a thiol-containing reducing agent, e.g. DTT or beta ME, as an elution agent.

Preferably, when said first or second monomeric unit is a Phe one of said first or second purification steps utilizes a phenyl sepharose as capture reagent. Preferably the purification step utilizes ethylene glycol as an elution agent.

In a preferred embodiment the capture reagent is in a solid phase, for example, present on the surface of a carrier material. The capture reagent may be attached to the surface of an affinity column, such that when a sample containing a target polypeptide comprising an affinity tag is passed through said column, the tagged polypeptide is retained in the column by the capture reagent, whilst the background polypeptides continues through the column. Preferably the column is a capillary column, such as a capillary column described in US2007/0196833, comprising an extraction capillary channel having an inner surface, a substantial portion of which is coated with a solid phase extraction surface. The solid phase extraction surface is affixed to the inner surface by a magnetic field. Preferably said capillary column is copper.

Alternatively, the capture reagent is provided on the surface of a bead, for example a magnetic bead. The beads can be added to a sample containing an affinity tagged target peptide, where the capture reagent will bind to the affinity tag. The beads may then be separated from the sample by, for example, centrifugation or, where the beads are magnetic, by use of a magnetic field.

Following binding of the affinity tag by the capture reagent, the affinity tag and target polypeptide may be eluted form the capture reagent by use of an appropriate elution agent.

The first purification step comprises binding of the affinity tag of the invention to a first binding partner, for example binding of His-residues to an immobilised metal affinity chromatography resin under suitable binding conditions. The step may then include an optional washing step, to remove any unbound polypeptides. The bound tag and target polypeptide are then separated from the capture reagent. Separation may be affected by elution of the bound tag using an appropriate elution agent, for example imidazole.

Proteins eluted by the first purification step are then further purified by a second purification step. The second purification step comprises binding of the affinity tag of the invention to a second binding partner, for example binding of Arg-residues to a cation exchange resin under suitable binding conditions. The step may then include an optional washing step, to remove any unbound polypeptides. The bound tag and target polypeptide are then separated from the capture reagent. Separation may be affected by elution of the bound tag using an appropriate elution agent, for example a high pH salt gradient.

Examples

Affinity tag construction

Oligonucleotides (SEQ ID NO:3 and 4) were designed and synthesized in order to effect the modification of the Poly-His rich N-Terminal region of the p54 tag (SEQ ID NO:1). The tag was modified with Glutamine codons changed to Arginine codons and with appropriate Restriction sites introduced to facilitate subcloning (SEQ ID NO:2). The aim was to generate a plasmid construct for the expression of HR-tagged Rubridoxin, a conveniently detected, approximately14 kDa Red protein

Affinity tag-polypeptide construct

Newly synthesized oligonucleotides were annealed for subcloning into pET28a-Rubridoxin or pET28a-p54-Rubridoxin (pQIS207) at *Nco*I – *Nde*I sites to make a new construct designated pQIS 216 (figure 4).

pQIS 207 was subject to restriction digestion with *Nco*I and *Nde*I to cut and recover the vector from the gel (Figure 5, lanes 1 & 2). DNA Marker Hyperladder (Bioline) is shown

in Lane 3 of figure 5 and Lane 4 shows the annealed oligonucleotides ready for insertion in the vector.

pQIS 207 was subject to restriction digest with *Sal*I and *Hinc*II (for confirmation purposes). The results are illustrated in figure 5, where the sizes shown are expected DNA fragments. The lanes represent the following disgests:

1. pQIS 207 + *Sal*I = 5648 bp
2. pQIS 207 + *Sal*I = 5648 bp
3. HyperLadder I (Bioline)
4. New construct Miniprep 1 + *Sal*I ~ 5318 + 219 bp
5. New construct miniprep 2 + *Sal*I ~ 5318 + 219 bp
6. New construct Miniprep 3 + *Sal*I ~ 5318 + 219 bp
7. New construct Miniprep 4 + *Sal*I ~ 5318 + 219 bp
9. pQIS 207 + *Hinc*II = 3921 + 1727 bp
10. pQIS 207 + *Hinc*II = 3921 + 1727 bp
11. HyperLadder I (Bioline)
12. New construct Miniprep 1 + *Hinc*II ~ 3921 + 1397 + 219 bp
13. New construct miniprep 2 + *Hinc*II ~ 3921 + 1397 + 219 bp
14. New construct Miniprep 3 + *Hinc*II ~ 3921 + 1397 + 219 bp
15. New construct Miniprep 4 + *Hinc*II ~ 3921 + 1397 + 219 bp

The digest results demonstrate that the first three recombinant plasmids (1, 2 and 3) had the correct insert. These plasmids were used for the overexpression of protein in Escherichia coli Rosetta (Novagen®).

Cell transformation

The newly constructed plasmid was named as pQIS 216 and it was used to transform Rosetta (*E. coli)* cells. Two clones pQIS 216.1 and pQIS 216.2 were used for this transformation. Glycerol stocks of pQIS 216 in DH5α compatible and Rosetta strains were made and stored at -80^{o}C. Rosetta cells carrying the plasmid pQIS 216 were grown in the presence of Kanamycin + Chloramphenicol and induced when OD_{600} was 0.75. Cells were further grown for 4 to 5 hours at 30^{o}C before harvesting and stored at -20^{o}C (for cell lysis and protein purification).

Fusion protein

The resultant fusion protein consists of 114 amino acids, has a Molecular Weight of 13367.9, an isoelectric Pt (pI) 9.44. The expected colour of the protein was red. The amino acid sequence of the fusion protein is illustrated in SEQ ID NO:5.
The protein was expressed after induction with 1 mM IPTG (final concentration). 20 ml culture was harvested in universal tubes to check the expression level. Cells were lysed with Bug Buster (Novagene®) and soluble fraction was loaded on the SDS gel. (There was no clear evidence for high level expression, but this is not uncommon). The resulting SDS gel illustrating induction and expression of the affinity tagged polypepetide is illustrated in figure 9, where the lanes contain as follows:

1. Un-induced Soluble Fraction
2. Induced Soluble Fraction
3. Protein Marker (kDa) (MBI Fermentas)

The fusion protein was expressed after induction with 1 mM IPTG in a culture volume of 400 ml. The cells were harvested, cells lysed with Bug Buster and soluble fraction loaded on to a Ni-Sepharose column.

The column was washed with PBS buffer pH 8.0 and then with PBS containing 50 mM Imidazole.

A red protein started migrating through the column during the wash (approx. 15 – 20 ml of 50 mM imidazole).

When the protein reached the foot of the column, the wash was stopped and the protein was eluted with 250 mM Imidazole and fractions of approximately 0.4ml collected (see figure 10). The peak fractions are obvious by their colour.

SDS gel illustrating size separation of fractions 2, 3, 4, 5, 6 and 7 of figure 10 is illustrated in figure 11, where the lanes are numbered according to the fractions in Figure 10.

As illustrated by figures 11 and 12 the red protein migrated at a higher than predicted molecular weight approximately 28 kDa (this is a dimer as shown by mass spectrometry). Contaminant proteins were present in almost all fractions. Hence a less stringent wash with 10 mM Imidazole was needed before elution.

The protein collected in Fraction 6 was diluted 20-fold with HEPES buffer (pH 8.2), DTT 5mM, EDTA 1mM (without NaCl) and loaded onto an SP Sepharose column and was washed with 10 column volumes of the same buffer.

HEPES buffer (pH 8.2) DTT 5mM, EDTA 1mM, with 200 mM NaCl was then used to elute the bound protein. Finally, the bound column was stripped with 750 mM NaCl and 2 to 3 ml fractions collected.

Having established conditions for elution, the protein was then diluted again in 50 ml of HEPES buffer (pH 8.2) DTT 5mM, EDTA 1mM (without NaCl) and loaded on the column and washed with same wash buffer before step washing with NaCl containing buffer 50, 100, 150, and the 200 mM.

Since 50 mM NaCl wash insufficient to displace the bound protein from the cation exchanger, the salt concentration was then increased to 100 mM (HEPES buffer (pH 8.2)) and the bound red band started to elute: the protein was efficiently eluted with 250 mM NaCl in four 0.5ml fractions, which were subsequently analysed by SDS PAGE.

As illustrated in figure 12 Rubredoxin migrated as an approximately 28 kDa dimer as well as 14 kDa monomer. Contaminant proteins (observed in lanes 6 – 10) were removed by the cation exchange step (1-3).

SDS gel illustrating size separation of ion exchange purified fraction 6 of figure 10 following SP-Sepharose separation is illustrated in figure 12, where the lanes contain as follows:

1. Ni-Sepharose Fraction 6 (5 µl)
2. Pre-stained protein marker Broad range (MBI) (7 µl)
3. SP Sepharose Fraction 1 (50 µl)
4. SP Sepharose Fraction 2 (50 µl)
5. SP Sepharose Fraction 3 (50 µl)
6. SP Sepharose Fraction 4 (50 µl)

As illustrated in figure 12 Rubredoxin migrated as an approximately 28 kDa dimer as well as 14 kDa monomer. Contaminant proteins (observed in lane 1) were now removed by the cation exchange step (see lanes 3 – 6).

The reader's attention is directed to all papers and documents which are filed concurrently with or previous to this specification in connection with this application and which are open to public inspection with this specification, and the contents of all such papers and documents are incorporated herein by reference.

All of the features disclosed in this specification (including any accompanying claims, abstract and drawings), and/or all of the steps of any method or process so disclosed, may be combined in any combination, except combinations where at least some of such features and/or steps are mutually exclusive.

Each feature disclosed in this specification (including any accompanying claims, abstract and drawings), may be replaced by alternative features serving the same, equivalent or similar purpose, unless expressly stated otherwise. Thus, unless expressly stated otherwise, each feature disclosed is one example only of a generic series of equivalent or similar features.

The invention is not restricted to the details of any foregoing embodiments. The invention extends to any novel one, or any novel combination, of the features disclosed in this specification (including any accompanying claims, abstract and drawings), or to any novel one, or any novel combination, of the steps of any method or process so disclosed.

CLAIMS

1. An affinity tag comprising a copolymer of a first monomeric unit and a second monomeric unit, wherein each of said first and second monomeric units has affinity for a distinct capture reagent, provided that said copolymer is not a block copolymer.

2. An affinity tag according to claim 1, wherein each of said first and second monomeric units are amino acid residues.

3. An affinity tag according to claim 1 or 2, wherein said first monomeric unit has affinity for an immobilized metal affinity chromatography resin.

4. An affinity tag according to claim 3, wherein said immobilized ion is selected from the group consisting of Co^{2+}, Cu^{2+}, Ni^{2+}, Zn^{2+}, Ca^{2+} and Fe^{3+}.

5. An affinity tag according to claim 4, wherein said immobilised ion is Ni^{2+}.

6. An affinity tag according to any one of the preceding claims, wherein said first monomeric unit is Histidine.

7. An affinity tag according to any one of the preceding claims, wherein said second monomeric unit has affinity for an ion exchange resin.

8. An affinity tag according to claim 7 wherein said ion exchange resin is an anion exchange resin.

9. An affinity tag according to claim 8, wherein said second monomeric unit is an acidic amino acid.

10. An affinity tag according to claim 9, wherein said acidic amino acid is Aspartic acid or a Glutamic acid.

11. An affinity tag according to claim 8, wherein said second monomeric unit is a basic amino acid.

12. An affinity tag according to claim 11, wherein said second monomeric unit is Arginine or Lysine.

13. An affinity according to any one of the preceding claims wherein said first monomeric unit is Histidine and said second monomeric unit is Arginie.

14. A method of purifying a recombinant protein comprising:
    a) preparing a peptide comprising an affinity tag according to any one of claims 1 to 13;
    b) purifying the peptide using a first purification step comprising binding the affinity tag to a first capture reagent selective for the affinity tag; and
    c) further purifying the peptide of b) using a second purification step comprising binging the affinity tag to a second capture reagent selective for the affinity tag, wherein said second capture reagent is distinct from said first capture reagent.

15. A method according to claim 14, wherein said first purification step comprises immobilized metal ion affinity chromatography.

16. A method according to claim 15, said first capture reagent is an immobilized metal ion affinity chromatography resin

17. A method according to claim 15 or 16, wherein said second step comprises ion exchange chromatography.

18. A method according to claim 17, wherein said second capture reagent is ion exchange resin.

19. A method according to claim 14, wherein said first purification step comprises ion exchange chromatography.

20. A method according to claim 19, wherein said first capture reagent is ion exchange resin.

21. A method according to claim 19 or 20, wherein said second step comprises immobilized metal affinity chromatography.

22. A method according to claim 21, wherein said second capture reagent is an immobilized metal affinity chromatography resin.

23. A method according to claim 16 or 22, wherein said immobilized ion is selected from the group consisting of Co^{2+}, Cu^{2+}, Ni^{2+}, Zn^{2+}, Ca^{2+} and Fe^{3+}.

24. A method according to claim 18 or 21, wherein said ion exchange resin is an anion exchange resin or a cation exchange resin.

25. Use of an affinity tag according to any one of claims 1 to 13 for the purification of a recombinantly expressed polypeptide.

26. A recombinant protein comprising an affinity tag according to any one of claims 1 to 13.

27. A recombinant protein according to claim 26, wherein the recombinant protein is cleavably linked to said affinity tag.

28. A recombinant polypeptide according to claim 26 or claim 27 wherein said affinity tag is attached to the C-terminal end of said polypeptide.

29. A recombinant polypeptide according to claim 26 or claim 27, wherein said affinity tag is attached to the N-terminal end of said polypeptide.

30. An isolated nucleic acid molecule encoding a protein according to any one of claims 26 to 29.

31. An expression vector comprising a nucleic acid molecule according to claim 30.

32. A host cell comprising an expression vector according to claim 31.

33. A host cell according to claim 32, wherein said cell is a eukaryotic cell.

34. A host cell according to claim 33, wherein said cell is a prokaryotic cell.

35. An affinity tag substantially as described herein with reference to the accompanying drawings.

36. A method of purifying a recombinant protein substantially as described herein with reference to the accompanying drawings.

Figure 1

SEQ ID NO:1

M Q S N K T F N L E K **Q** N H T P R K H H **Q** H H H **Q Q Q** H H **Q Q Q Q Q Q** P P
P P P I P A N G Q Q A S S Q N E G L T I D L K

Figure 2

SEQ ID NO:2

M E **R** N H T P R K H H **R** H H H **R R R** H H **R R R R R R** P P P P P I P

Figure 3a

SEQ ID NO:3

p54 Q-R (Top) oligo

CAT GGA ACG AAA CCA TAC TCC AAG AAA GCA TCA TCG ACA TCA CCA CCG GCG GCG GCA CCA CCG GCG TCG ACG GCG GAG GCC TCC ACC ACC GCC AAT ACC TCA = 102 bps

Figure 3b

SEQ ID NO:4

P54 Q-R (Bottom) oligo

TAT GAG GTA TTG GCG GTG GTG GAG GCC TCC GCC GTC GAC GCC GGT GGT GCC GCC GCC GGT GGT GAT GTC GAT GAT GCT TTC TTG GAG TAT GGT TTC GTT C = 100 bps

Figure 4

```
                                                                  BanI
                                                                 ┌────┐
1   catggaacga aaccatactc caagaaaagca tcatcgacat caccaccggc ggcggcacca
      cttgct ttggtatgag gttctttcgt agtagctgta gtggtggccg ccgccgtggt
       m  e  x    n  h  t    p  x  K    h  h  x  h    h  h  x    x  x  h

                HincII
                 SalI                StuI
                ┌────┐              ┌────┐
61  ccgggcgtcga cggcggaggc ctccaccacc gccaatacct ca
    ggccgcagct gccgcctccg gaggtgggtg cggttatgga gtat
    h  x  x  x    x  x  x    p  p  p    p  p  i  p
```

Figure 5

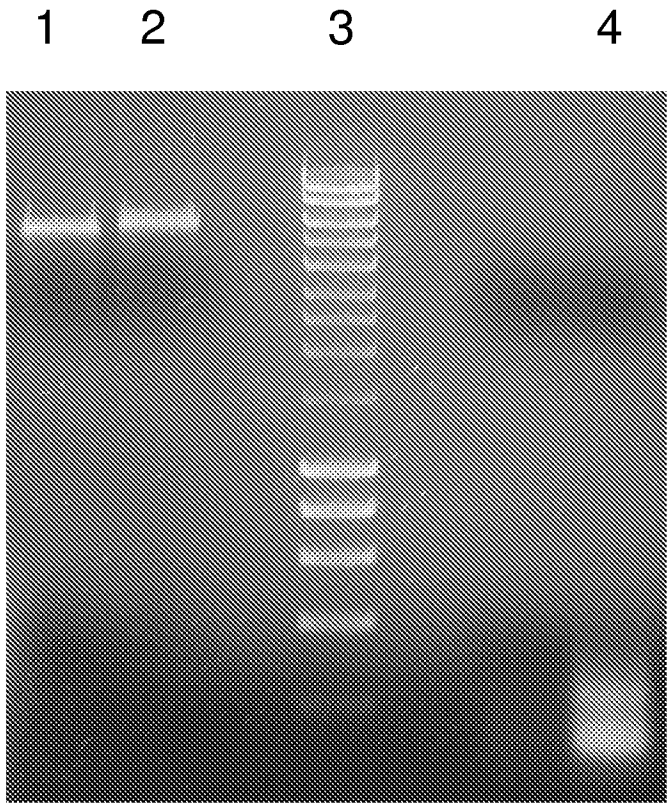

Figure 6

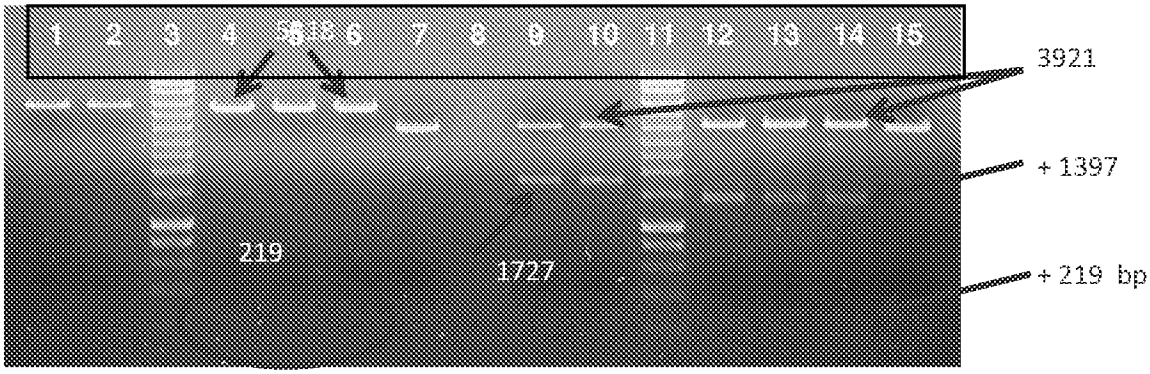

Figure 7

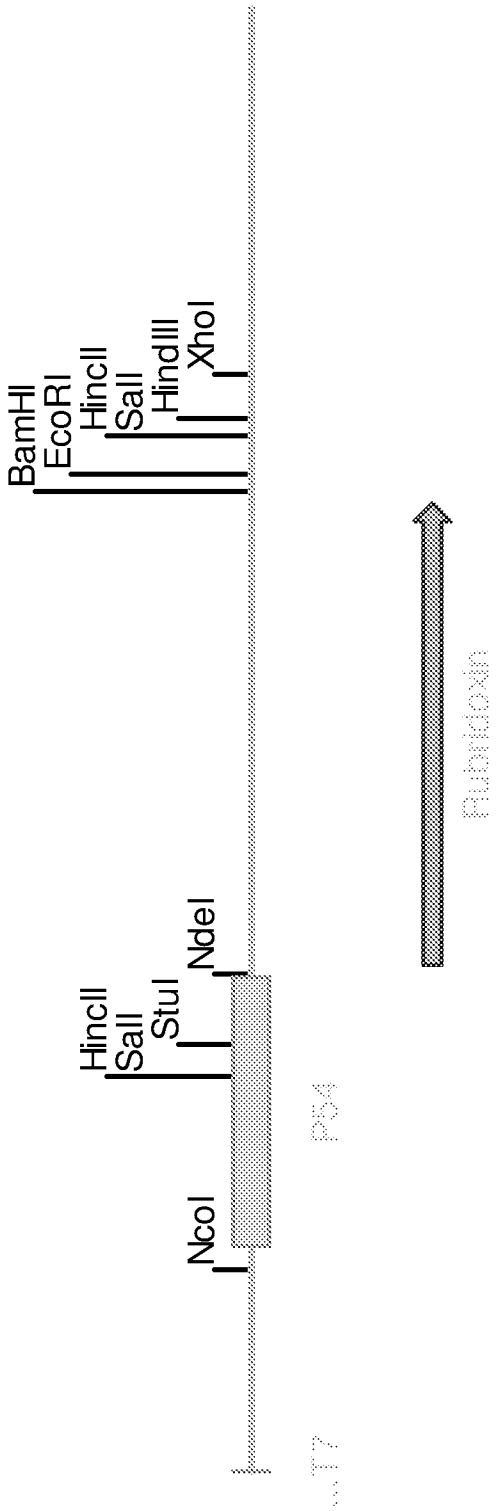
BamHI
EcoRI
HincII
SalI
HindIII
XhoI
NdeI
HincII
SalI
StuI
NcoI
P04
T7
Rubredoxin

Figure 8

SEQ ID NO:5

MERNHTPRKHHRHHHRRRHHRRRRRRPPPPPIPHMKKYRCKLCGYIYDPEQGDPDSG
IEPGTPFEDLPDDWVCPLCGASKEDFEPVEGGSEFELRRQACGRTRAPPPPPLRSGC

Figure 9

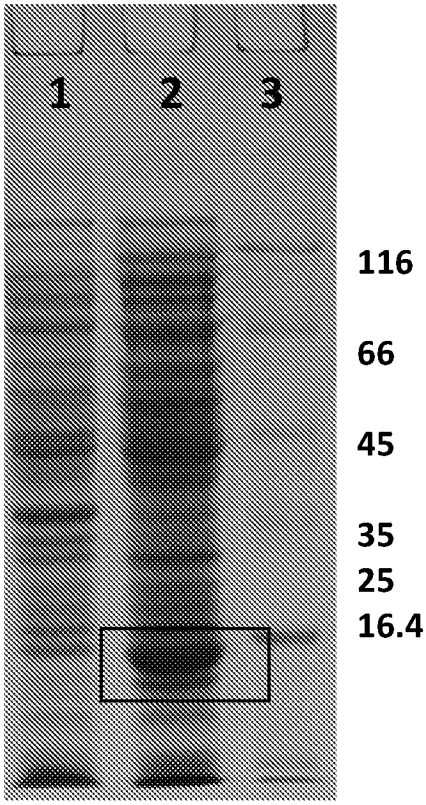

Figure 10

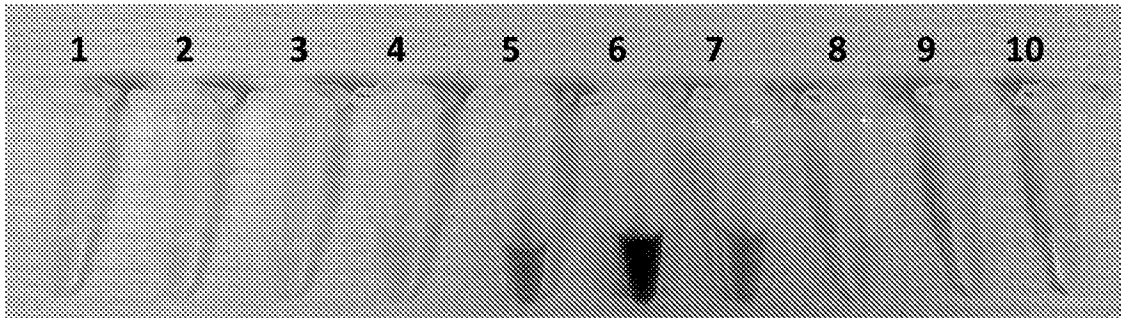

Figure 11

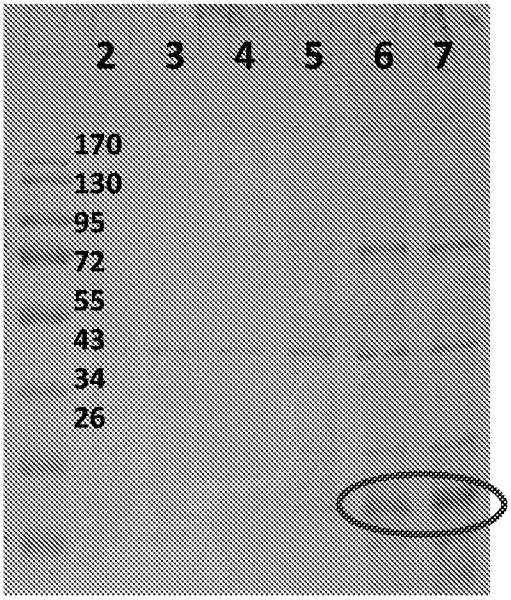

Figure 12