Innovation, Sciences et Développement économique Canada
Office de la Propriété Intellectuelle du Canada

Innovation, Science and Economic Development Canada
Canadian Intellectual Property Office

CA 2900745 C 2023/01/24

(11)(21) 2 900 745

(12) BREVET CANADIEN
CANADIAN PATENT

(13) C

(86) **Date de dépôt PCT/PCT Filing Date:** 2014/02/14
(87) **Date publication PCT/PCT Publication Date:** 2014/08/21
(45) **Date de délivrance/Issue Date:** 2023/01/24
(85) **Entrée phase nationale/National Entry:** 2015/08/10
(86) **N° demande PCT/PCT Application No.:** GB 2014/050435
(87) **N° publication PCT/PCT Publication No.:** 2014/125290
(30) **Priorité/Priority:** 2013/02/14 (GB1302597.8)

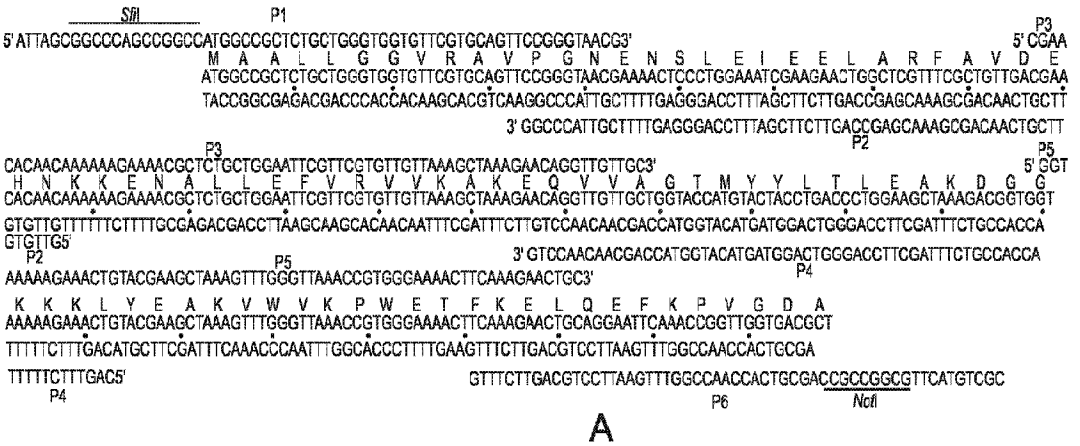
(51) **Cl.Int./Int.Cl.** ***C07K 14/415*** (2006.01),
A61K 38/16 (2006.01), ***C07K 19/00*** (2006.01),
C12N 15/29 (2006.01), ***C12N 15/62*** (2006.01),
C40B 30/04 (2006.01), ***C40B 40/02*** (2006.01),
C40B 40/10 (2006.01), ***C40B 50/06*** (2006.01),
G01N 33/53 (2006.01)

(72) **Inventeurs/Inventors:**
MCPHERSON, MICHAEL, GB;
TOMLINSON, DARREN, GB

(73) **Propriétaire/Owner:**
UNIVERSITY OF LEEDS, GB

(74) **Agent:** GOWLING WLG (CANADA) LLP

(54) **Titre : PROTEINES D'ECHAFAUDAGE DERIVEES DE CYSTATINES VEGETALES**
(54) **Title: SCAFFOLD PROTEINS DERIVED FROM PLANT CYSTATINS**



(57) **Abrégé/Abstract:**
The present invention relates to scaffold proteins derived from plant cystatins and to nucleic acids encoding them. The scaffolds are highly stable and have the ability to display peptides. The scaffolds are particularly well suited for constructing libraries, e.g. in phage display or related systems. The invention also related to various uses of the scaffolds, including in therapy, diagnosis, environmental and security monitoring, synthetic biology and research, and to cells and cell cultures expressing the scaffold proteins.

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
21 August 2014 (21.08.2014)

WIPO | PCT

(10) International Publication Number
WO 2014/125290 A1

(51) International Patent Classification:
A61K 38/16 (2006.01) *G01N 33/53* (2006.01)
C07K 14/415 (2006.01)

(21) International Application Number:
PCT/GB2014/050435

(22) International Filing Date:
14 February 2014 (14.02.2014)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
1302597.8 14 February 2013 (14.02.2013) GB

(71) Applicant: **UNIVERSITY OF LEEDS** [GB/GB]; Leeds LS2 9JT (GB).

(72) Inventors: **MCPHERSON, Michael**; C/o School of Molecular and Cellular Biology, Faculty of Biological Sciences, Astbury Building Level 7, University of Leeds, Leeds LS2 9JT (GB). **TOMLINSON, Darren**; 40 Westfield Lane, South Milford, Leeds LS25 5AP (GB).

(74) Agent: **HARRISON GODDARD FOOTE LLP**; Delta House, 50 West Nile Street, Glasgow, Glasgow GB G12NP (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, BN, 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, IR, IS, JP, KE, KG, 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, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, 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, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, 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, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report (Art. 21(3))*
— *with sequence listing part of description (Rule 5.2(a))*

WO 2014/125290 A1

(54) Title: SCAFFOLD PROTEINS DERIVED FROM PLANT CYSTATINS

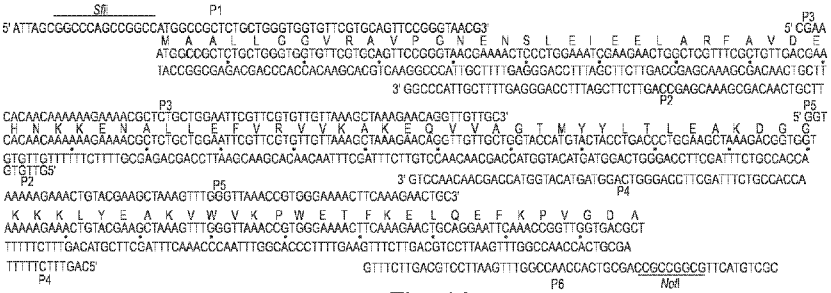


Fig. 1A

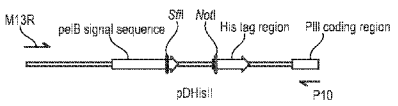


Fig. 1B

(57) Abstract: The present invention relates to scaffold proteins derived from plant cystatins and to nucleic acids encoding them. The scaffolds are highly stable and have the ability to display peptides. The scaffolds are particularly well suited for constructing libraries, e.g. in phage display or related systems. The invention also related to various uses of the scaffolds, including in therapy, diagnosis, environmental and security monitoring, synthetic biology and research, and to cells and cell cultures expressing the scaffold proteins.

SCAFFOLD PROTEINS DERIVED FROM PLANT CYSTATINS

The present invention relates to novel proteins, nucleic acids encoding them, and to methods of using them. In particular, the present invention relates to proteins with utility as scaffolds for displaying peptide sequences, and to use of these scaffolds in areas such as therapy, diagnosis, environmental and security monitoring, synthetic biology and research.

Background of the Invention

The use of 'so-called' protein scaffolds is gaining attention in biochemistry as a potential route to generating novel ligand binding proteins for use in research and medicine. Such scaffold proteins, used to display one or more peptide sequences, can potentially provide an alternative to antibodies or antibody fragments.

The term 'protein scaffold' is used to describe a type of polypeptide structure that is observed in differing contexts and with distinct biochemical functions. Because of their intrinsic conformational stability it has been reasoned that such scaffolds might be amenable to protein engineering.

It is conventionally thought that immunoglobulins (antibodies) owe their function to the composition of a conserved framework region and a spatially well-defined antigen-binding site made of hypervariable peptide segments, these segments are variable both in sequence and in conformation. After antibody engineering methods along with library techniques resulted in successes in the selection of functional antibody fragments, interest began to grow in using other protein architectures to synthesise useful binding proteins.

Descriptions of the desirable properties of a suitable protein scaffold taken include the following;

"Candidates for suitable protein scaffolds ought to exhibit a compact and structurally rigid core that is able to present surface loops of varying sequence and length or to otherwise tolerate side chain replacements in a contiguous surface region, including exposed hydrophobic residues, without significant changes in their folding properties."
Skerra A: Engineered protein scaffolds for molecular recognition. J Mol Recognit 2000, 13:167-187 and 'The term 'scaffold', as used in protein engineering, describes a single chain polypeptidic framework typically of reduced size (< 200 AA) and containing a highly structured core associated with variable portions of high conformational tolerance allowing

insertions, deletions, or other substitutions'. Wurch et al., Trends in Biotechnology, November 2012, 30, 575-582.

However, not all kinds of polypeptide fold which may appear attractive for the engineering of loop regions at a first glance will indeed permit the construction of independent ligand binding sites with high affinities and specificities.

Prior art scaffolds include inactivated staphylococcal nuclease, green fluorescent protein (GFP) and thioredoxin A (TrxA), the fibronectin type III domain ('Fn3'), lipocalin family proteins, bilin binding protein (BBP), as well as isolated protein folds such as the Z domain of staphylococcal protein A, "affibodies", anticafins, and ankyrin repeats, and others.

WO 2006/131749 describes several rational mutations made in Stefin A to improve it as a scaffold. The modified Stefin A scaffold comprises mutations at the following three sites Lys71-Leu73, V48D and G4W and is referred to as STM (Stefin A Triple Mutant).

WO2009/136182 describes further refinements of STM scaffolds.

There remains a need for improved scaffold proteins. In particular, it has been found that prior art scaffolds are often unable to sufficiently stabilise peptide aptamers, and the presence of the aptamers in the scaffold actually causes the scaffold to significantly deform. See for example Woodman *et al.*, 'Design and Validation of a Neutral Protein Scaffold for the Presentation of Peptide Aptamers', J. Mol. Biol. (2005) 352, 1118–1133

Ideally such improved scaffold proteins would provide one or more of the following benefits:

- Improved stability to provide a rigid framework that does not deform;
- Smaller size;
- Improved ability to support high quality and complexity libraries;
- Improved affinity of selected binding proteins for their target; and
- Simplicity of further manipulation due to accessible N-and C-terminal ends.

Statements of the Invention

According to the present invention there is provided a synthetic protein having a sequence derived from or related to a plant cystatin.

Suitably the synthetic protein comprises a consensus sequence of several (e.g. 10 or more, 20 or more, or 50 or more) plant cystatin proteins.

Suitably the synthetic protein is a scaffold protein comprising sites adapted or suitable for insertion of heterologous peptide sequences.

The scaffold protein of the present invention is thus preferably not based on a single naturally occurring protein sequence, but is a novel protein that does not exist in nature, being derived through careful design of consensus sequence from representatives of the class of plant cystatins.

In one embodiment the synthetic protein comprises the amino acid sequence NSLEIEELARFAVDEHNKKENALLEFVRVVKAKEQVVAGTMYYLTLEAKDGGKKKLYEAK VWVKPWENFKELQEFKPVGDA (SEQ ID NO 1), or a variant thereof. Preferably the variant has a sequence at least 50%, more preferably 70%, identical thereto.

It has been found that a synthetic protein comprising such a sequence provides a highly stable and useful scaffold protein.

In another embodiment the synthetic protein comprises the amino acid sequence GNENSLEIEELARFAVDEHNKKENALLEFVRVVKAKEQVVAGTMYYLTLEAKDGGKKKLYE AKVWVKPWENFKELQEFKPVGDA (SEQ ID NO 2), or a variant thereof. Preferably the variant has a sequence at least 50%, more preferably 70%, identical thereto.

In another embodiment the synthetic protein comprises the amino acid sequence ATGVRAVPGNENSLEIEELARFAVDEHNKKENALLEFVRVVKAKEQVVAGTMYYLTLEAKD GGKKKLYEAKVWVKPWENFKELQEFKPVGDA (SEQ ID NO 3), or a variant thereof. Preferably the variant has a sequence at least 50%, more preferably 70%, identical thereto.

More preferably the synthetic protein of the present invention comprises an amino acid sequence which is at least 75%, 80%, 85%, 90%, 95% or 99% identical to SEQ ID NOS 1, 2 or 3 above. Generally higher levels of identity are preferred.

In some cases the synthetic protein comprises or consists of an amino acid sequence which is identical, or substantially identical, to SEQ ID NO 1, 2 or 3.

In preferred embodiments of the present invention the synthetic protein set out above comprises at least one heterologous peptide sequence inserted therein. The synthetic proteins of the present invention have particular utility as scaffold proteins used to constrain and display peptide sequences. The present invention thus extends to both the 'empty' scaffold protein (i.e. without any heterologous peptide sequences present) and the scaffold protein with one or more heterologous sequences inserted therein.

Accordingly, a preferred embodiment of the present invention is a synthetic scaffold protein having a sequence as described herein, the scaffold protein displaying one or more heterologous peptides inserted at appropriate points in the scaffold. By 'displaying' it is meant that the peptide sequence is inserted into the scaffold protein at a location which allows the peptide to be exposed on the surface of the scaffold under suitable conditions, e.g. non-denaturing conditions, such as *in vivo* or in an *in vitro* assay. Such scaffold proteins displaying one or more heterologous peptides are often referred to as peptide aptamers or multiple peptide aptamers, though it should be noted that the term aptamers is used somewhat inconsistently in the art to refer to the peptides themselves or the complete protein.

Preferably the heterologous peptide sequence is from 3 to 30 amino acids in length, more preferably from 4 to 20 amino acids, more preferably from 4 to 16 amino acids. The synthetic proteins of the present invention have been found to be particularly suitable for displaying heterologous sequences from 5 to 13 amino acids in length.

Preferably the heterologous peptide sequence is inserted in a loop region of the synthetic protein or at the N-terminus of the protein. Loop regions can be defined as regions which are not involved with ordered secondary or tertiary structures of the protein when under conventional conditions (i.e. the protein is correctly folded and not under denaturing conditions), but may contribute to function and/or the correct spatial organisation of secondary structure elements in the protein. The position of loop regions within the scaffold protein can of course vary between different protein variants. Loop regions of the synthetic protein can be determined by examining the secondary and tertiary structure of the protein, and methods to achieve this are well known in the art, including X-ray crystallography, NMR, and also *in silico* methods.

In the present invention it is preferred that the heterologous peptides are inserted in at least one of the following positions in the protein:

- the loop between a first and second region of β-sheet (known as LOOP1); and

- the loop between a third and fourth region of β-sheet (known as LOOP2).

First, second, third and fourth is, of course, intended to be interpreted as relative to the protein sequence, i.e. from the N- to C-terminus of the protein.

Preferably heterologous peptides are inserted at both of these positions, i.e. at LOOP1 and LOOP2.

It is preferred that the loop length between adjacent regions of beta sheet is from 3 to 20 amino acids in length, more preferably from 5 to 13 amino acids in length, and it is believed that a loop length of approximately 9 amino acids in length is optimal.

Exemplary pairs of heterologous peptides for insertion into LOOP1 and LOOP2 are set out in Table 2, i.e. SEQ ID NOS 25 to 72. These examples form preferred embodiments of the present invention.

In addition, a further insertion point for a heterologous peptide is at or near (e.g. within 4 amino acids) the N-terminus of the protein. The peptide sequence at this point is believed to be less critical to binding a target than insertions at the other two positions mentioned above, but nonetheless it would appear to have a role in binding.

Thus in one embodiment of the present invention, the synthetic scaffold protein comprises three heterologous peptides, one in each of the locations discussed above.

In the case of the specific sequences set out above, preferred loop regions within the protein sequence are located as underlined:

NSLEIEELARFAVDEHNKKENALLEFVRVVKAKEQVVAGTMYYLTLEAKDGGKKKLYEAKV WVKPWENFKELQEFKPVGDA (SEQ ID NO 1)

GNENSLEIEELARFAVDEHNKKENALLEFVRVVKAKEQVVAGTMYYLTLEAKDGGKKKLYE AKVWVKPWENFKELQEFKPVGDA (SEQ ID NO 2)

MATGVRAVPGNENSLEIEELARFAVDEHNKKENALLEFVRVVKAKEQVVAGTMYYLTLEAK DGGKKKLYEAKVWVKPWENFKELQEFKPVGDA (SEQ ID NO 3)

The underlined regions could be completely or partially replaced by the heterologous peptide, or the heterologous peptide could be inserted within the underlined regions without removal of the loop regions. Additionally, a heterologous peptide can be added to the N terminus of SEQ ID NO 1 or 2.

However, it should be noted that the heterologous peptide could be inserted in other loop regions, for example the loop regions between α-helix and first β-sheet and/or the second and third region of β-sheet (which may be referred to as LOOP 3 and LOOP 4 respectively for convenience). When the scaffold protein is correctly folded, these loop regions are positioned at the opposite side of the scaffold from those mentioned above (LOOP 1 and LOOP 2). By inserting heterologous peptides loop regions on both sides of the scaffold it is possible to produce a divalent moiety, able to bind a first target on one side of the scaffold and a second target at the other (generally opposite) side, and this may be a desirable embodiment in certain situations.

Where the synthetic protein comprises an amino acid sequence as set out above, the sequence can be contiguous or non-contiguous. For example, the sequence will be non-contiguous where a heterologous peptide has been inserted within the synthetic protein, e.g. at one of the loop regions mentioned above, the heterologous peptide thus separating portions of the abovementioned sequences.

It will be noted that calculations of sequence identity should for the present invention, in general, be adapted so as to take account of the situation where a heterologous peptide is inserted into the synthetic protein. When such an insertion has occurred, the inserted peptide sequence should typically be disregarded when calculating sequence identity. This is because the synthetic proteins of the present invention function as scaffolds, in which case the inserted peptides are inserted at points in the synthetic protein where they will be displayed at the surface of the protein. The inserted peptide sequences are by their very nature and purpose highly variable. In such a case it is the sequence of the scaffold portion of the protein which is of concern, as it provides the stable framework for display of the peptides, rather than the intentionally highly variable inserted peptide sequences.

The heterologous peptide can be inserted into the synthetic protein with or without the removal of amino acids normally found in the synthetic protein. That is to say the heterologous peptide can be inserted at an end of the synthetic protein or between two amino acids within the synthetic protein, without any normal amino acids of the synthetic protein being removed. Alternatively, when a peptide is inserted into the synthetic protein

one or more amino acids normally present in the synthetic scaffold protein can be removed/replaced. e.g. the 'loop' amino acids VVAG, PWE and ATG (when present) can be removed in sequences 1, 2 and 3 above, or the corresponding loop sequences in a sequence variant.

Particularly preferred scaffold proteins according to the invention are set out in SEQ ID NOS 74 to 79 below, and the associated description of potential modifications to the scaffolds.

The term "identity" in respect of protein sequence refers to a degree of similarity between proteins in view of differences in amino acids, but which takes into account different amino acids which are functionally similar in view of size, lipophilicity, acidity, etc. A percentage identity can be calculated by optimal alignment of the sequences using a similarity-scoring matrix such as the Blosum62 matrix described in Henikoff S. and Henikoff J.G., P.N.A.S. USA 1992, 89: 10915-10919. Calculation of the percentage identity and optimal alignment of two sequences using the Blosum62 similarity matrix and the algorithm of Needleman and Wunsch (J. Mol. Biol. 1970, 48: 443-453) can be performed using the GAP program of the Genetics Computer Group (GCG, Madison, WI, USA) using the default parameters of the program. Specific parameters for calculating percentage identity for protein sequences and nucleic acid sequences in respect of the present invention are described below.

Variants of the proteins that also form part of the present invention may contain variations in the amino acid sequence due to deletions, substitutions, insertions, inversions or additions of one or more amino acids in said sequence or due to an alteration to a moiety chemically linked to the protein. For example, a protein variant may comprise a carbohydrate or PEG structure attached to a protein. The proteins of the invention may include one or more such protein modification. The proteins of the invention may include a deletion of 1 or more amino acids, e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10, amino acids from the N or C terminus, provided the variant retains the desired function.

Substitutional variants of proteins are those in which at least one residue in the amino acid sequence has been removed and a different residue inserted in its place. The proteins of the present invention can contain conservative or non-conservative substitutions.

The term "conservative substitution", relates to the substitution of one or more amino acid substitutions for amino acid residues having similar biochemical properties. Typically, conservative substitutions have little or no impact on the activity of a resulting protein. For example, a conservative substitution in a protein may be an amino acid substitution that

does not substantially affect the ability of the protein to fold correctly and otherwise perform its usual biological function. Screening of variants of the proteins of the present invention can be used to identify which amino acid residues can tolerate an amino acid substitution. In one example, the relevant melting temperature or the amount of α-helix and β-sheet of a modified protein is not reduced by more than 25%, preferably not more than 20%, especially not more than 10%, when one or more conservative amino acid substitutions are effected.

One or more conservative substitutions can be included in a protein of the present invention. In a preferred example, 10 or fewer conservative substitutions are included in the protein. A protein of the invention may therefore include 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more conservative substitutions. A polypeptide can be produced which contain one or more conservative substitutions by manipulating the nucleotide sequence that encodes that polypeptide using, for example, standard procedures such as site-directed mutagenesis or PCR. Alternatively, a polypeptide can be produced to contain one or more conservative substitutions by using peptide synthesis methods as known in the art.

Examples of amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative substitutions include: Ser for Ala; Lys for Arg; Gln or His for Asn; Glu for Asp; Asn for Gln; Asp for Glu; Pro for Gly; Asn or Gln for His; Leu or Val for Ile; Ile or Val for Leu; Arg or Gln for Lys; Leu or Ile for Met; Met, Leu or Tyr for Phe; Thr for Ser; Ser for Thr; Tyr for Trp; Trp or Phe for Tyr; and Ile or Leu for Val. In one embodiment, the substitutions are among Ala, Val Leu and Ile; among Ser and Thr; among Asp and Glu; among Asn and Gln; among Lys and Arg; and/or among Phe and Tyr. Further information about conservative substitutions can be found in, among other locations, Ben-Bassat *et al.*, (J. Bacteriol. 169:751-7, 1987), O'Regan *et al.*, (Gene 77:237-51, 1989), Sahin-Toth et al., (Protein Sci. 3:240-7, 1994), Hochuli *et al.*, (Bio/Technology 6:1321-5, 1988), WO 00/67796 (Curd et al.) and in standard textbooks of genetics and molecular biology.

Other variants can be, for example, functional variants such as salts, amides, esters, and specifically C-terminal esters, and N-acyl derivatives. Also included are peptides which are modified *in vivo* or *in vitro*, for example by glycosylation, amidation, carboxylation or phosphorylation.

Proteins according to the present invention can be modified by a variety of chemical techniques to produce derivatives having essentially the same activity as the unmodified peptides, and optionally having other desirable properties. For example, carboxylic acid

groups of the protein, whether carboxyl-terminal or side chain, may be provided in the form of a salt of a pharmaceutically-acceptable caution or esterified, for example to form a C1-C6 alkyl ester, or converted to an amide, for example of formula $CONR^1R^2$ wherein R^1 and R^2 are each independently H or C1-C6 alkyl, or combined to form a heterocyclic ring, such as a 5- or 6-membered ring. Amino groups of the peptide, whether amino-terminal or side chain, may be in the form of a pharmaceutically-acceptable acid addition salt, such as the HCl, HBr, acetic, benzoic, toluene sulfonic, maleic, tartaric and other organic salts, or may be modified to C1-C6 alkyl or dialkyl amino or further converted to an amide. Hydroxyl groups of the peptide side chains may be converted to alkoxy or ester groups, for example C1-C6 alkoxy or C1-C6 alkyl ester, using well-recognized techniques. Phenyl and phenolic rings of the peptide side chains may be substituted with one or more halogen atoms, such as F, Cl, Br or I, or with C1-C6 alkyl, C1-C6 alkoxy, carboxylic acids and esters thereof, or amides of such carboxylic acids. Methylene groups of the peptide side chains can be extended to homologous C2-C4 alkylenes. Thiols can be protected with any one of a number of well-recognized protecting groups, such as acetamide groups. Those skilled in the art will also recognize methods for introducing cyclic structures into the peptides of this disclosure to select and provide conformational constraints to the structure that result in enhanced stability.

In preferred embodiments of the present invention the synthetic protein has a melting temperature (Tm) of at least 90 °C, more preferably at least 95 °C, and most preferably at least 100 °C. Tm in the case of proteins is also known as the denaturation midpoint, and is defined as the temperature at which both the folded and unfolded states are equally populated. Tm is generally determined for the synthetic protein in which no heterologous peptide sequences have been inserted. Insertion of heterologous sequences can, and often does, reduce the Tm and therefore the most meaningful value is obtained when empty scaffold proteins are compared, and this is the basis of the Tm figures mentioned above. It is a surprising advantage of the present invention that such highly stable proteins have been obtained. Melting temperature is a particularly useful indicator of protein stability. The relative proportions of folded and unfolded proteins can be determined by many techniques known to the skilled person, including differential scanning calorimetry, UV difference spectroscopy, fluorescence, circular dichroism (CD), and NMR (see Pace, C. Nick, and J. Martin Scholtz. "Measuring the conformational stability of a protein" 299-321).

The extremely high temperature stability of the synthetic proteins of the present invention is also an indicator of the very high structural stability of the proteins when at normal temperatures (e.g. around 25 °C *in vitro* and 37 °C *in vivo*). This means that the synthetic

proteins of the present invention are very well suited to act as scaffolds for displaying heterologous peptides. Their structural stability means that such heterologous peptides will be displayed consistently and accurately, without disruption of the scaffold protein structure. Because the synthetic proteins of the present invention are so stable, it is an indication that they will perform better than known scaffolds which do not have the same level of stability.

It is surprising that such a small scaffold protein would have a thermal transition temperature of, for example, 101 °C, based on the source of the parental protein sequences of land plants that were used as a basis for its design. The land plant parents commonly grow at ambient external temperatures and so only a few crops such as rice might be expected to grow in tropical temperatures whilst he majority grow in moderate to cold northern climates.

As mentioned above, addition of heterologous peptides can, and typically does, reduce the stability of the synthetic scaffold proteins of the present invention. However, it has been found that the stability of the scaffolds of the present invention remains higher than prior art scaffolds. Accordingly, in a preferred embodiment the synthetic scaffold proteins having at least one heterologous peptide inserted therein have a Tm of 60 °C or higher, more preferably 70 °C or higher, and especially 80 °C or higher. This indicates that the scaffold proteins having at least one heterologous peptide inserted therein are highly stable.

Preferably the synthetic protein of the present invention comprises a linker or tag. The linker or tag can be an amino acid linker or tag, or another type of linker or tag. The tag can be any tag which provides a desired functionality to the synthetic protein, e.g. one which allows easy isolation or purification of the protein. An exemplary tag is a polyhistidine tag (also known as a His-tag), and other well-known tags include Myc-tag, SBP tag, S-tag, calmodulin tag, etc.

In a further embodiment of the present invention there is provided a synthetic protein as set out above connected to a substrate or moiety. The moiety can be a label, carrier, protein or the like. The synthetic protein can be connected directly to the substrate or moiety, or it can be connected via a linker. The synthetic protein can be connected covalently or non-covalently to the substrate or moiety. Non-covalent systems for linking proteins to a substrate or moiety include, but are not restricted to, the biotin-avidin system.

Where the moiety is a label, it can be a fluorescent label, radioactive label, enzymatic label or any other label known to the skilled person.

Examples of fluorescent labels include, but are not restricted to, organic dyes (e.g. cyanine, fluorescein, rhodamine, Alexa Fluors, Dylight fluors, ATTO Dyes, BODIPY Dyes, etc.), biological fluorophores (e.g. green fluorescent protein (GFP), R-Phycoerythrin, etc.), and quantum dots.

Examples of enzymatic labels include, but are not restricted to, horseradish peroxidase (HRP), alkaline phosphatase (AP), glucose oxidase and β-galactosidase.

Another well-known label is biotin. Biotin labels are typically composed of the biotinyl group, a spacer arm and a reactive group that is responsible for attachment to target functional groups on proteins. Biotin can be useful for attaching the labelled protein to other moieities which comprise an avidin moiety.

Various strategies for labelling proteins are well known to the skilled person, and they could be readily applied to the synthetic proteins of the present invention.

The substrate to which the protein is bound can be any suitable surface, e.g. the surface of a container such as a 96 well plate.

Where the moiety is a carrier, it can, for example, be a bead (e.g. a magnetic bead) or other particle.

Typically labels, linkers or tags are added on the C- and/or N-termini of the protein. However can also be added via any region of the protein which is sufficiently spaced from the loops in which heterologous peptides are inserted so as not to interfere with target binding, most preferably on the opposite side from the loops with insertions.

In an embodiment of the present invention there is provided a fusion protein comprising a synthetic protein as set out above connected to a second protein. The second protein may be a protein having a desired activity. In certain embodiments of the invention the second protein may be a phage coat protein or another protein useful in a surface display system, and uses for such fusion proteins in scanning methods are described in more detail below. The second protein could of course be another synthetic protein, thus providing a homo- or hetero-multimeric scaffold protein.

In a further aspect of the present invention there is provided a library comprising a population of synthetic proteins, as described above, wherein various members of the

population comprise a variety of heterologous peptides having different sequences. Suitable libraries of heterologous peptides can be created using combinatorial techniques known to the skilled person, and nucleic acid sequences encoding a suitable set of heterologous peptide sequences can be obtained through numerous commercial sources. Preferably such a library will have a complexity of 10^8 or higher, more preferably 10^9 or higher, and most preferably 10^{10} or higher.

Such a library has utility in selecting particular synthetic scaffold proteins which bind to a target entity. The target entity may be any entity to which it is desirable to have a protein bind specifically. Exemplary targets include, but are not limited to, proteins/peptides (e.g. receptors or ligands), small molecules (e.g. pharmaceutical molecules), nucleic acids (e.g. DNA or RNA) and inorganic compounds.

The library can comprise a population of the synthetic proteins of the present invention adapted for display in a suitable display system, e.g. a surface display system. The surface display system can suitably be a phage display. Alternatively, the surface display system can be a mRNA display, ribosome display, CIS-display or other non-covalent or covalent protein display system, bacterial display or yeast display. The yeast two-hybrid system or expression in bacterial or mammalian cell systems are commonly used *in vivo* systems for screening for aptamers which bind to target. Collectively these display systems are often referred to as biopanning systems.

For example, a phage display can be a filamentous phage (e.g. M13) display system. The synthetic protein can therefore be fused with pIII, pVIII, pVI, pVII or pIX proteins. Other phage display systems which can be used in the present invention include T4, T7 and λ phage display. It should be noted that the synthetic proteins of the present invention are versatile and can be used with any suitable display system, and various suitable systems are well known to the skilled person. Phage display is commonly used in the selection of antibodies, but the techniques are directly applicable to the proteins of the present invention. Summaries of phage display technologies can be found in, for example, Lowman H.B., Clackson T. (2004) Phage display: a practical approach. Oxford University Press. pp. 10-11.

In a particularly preferred embodiment of the present invention the synthetic protein of the present invention, especially when used in construction of a library, comprises the sequence:

NSLEIEELARFAVDEHNKKENALLEFVRVVKAKEQ(X_n)TMYYLTLEAKDG
GKKKLYEAKVWVK(X_n)NFKELQEFKPVGDA (SEQ ID NO 4)

wherein X is any amino acid and n is the number of amino acids in the sequence, and wherein n is preferably from 3 to 30, more preferably from 4 to 20, and yet more preferably from 4 to 15.

Preferably the synthetic protein comprises the sequence:
NSLEIEELARFAVDEHNKKENALLEFVRVVKAKEQ(X_{5-13})TMYYLTLEA
KDGGKKKLYEAKVWVK(X_{5-13})NFKELQEFKPVGDA (SEQ ID NO 5)

More preferably the synthetic protein comprises the sequence:
NSLEIEELARFAVDEHNKKENALLEFVRVVKAKEQ(X_9)TMYYLTLEA
KDGGKKKLYEAKVWVK(X_9) NFKELQEFKPVGDA (SEQ ID NO 6)

As discussed above, proteins comprising a sequence which is 50%, more preferably 70%, identical to SEQ ID NO 4, 5 or 6 are within the scope of the invention. More preferably the synthetic protein of the present invention comprises an amino acid sequence which is at least 75%, 80%, 85%, 90%, 95% or 99% identical to SEQ ID NO 4, 5 or 6. Generally higher levels of identity are preferred.

In any of the above examples, the protein may contain an additional amino acid sequence at the N-terminus. This can comprise some or all of the sequence MATGVRAVPGNE (SEQ ID NO 80). Alternatively or additionally it can comprise a further heterologous peptide sequence, e.g. of from 3 to 20 amino acids in length.

In embodiments of the present invention it may be preferred that the N terminal amino acid is a methionine. This can be achieved, for example, by adding on an additional methionine or by replacing the normal N-terminal amino acid with a methionine.

Particularly preferred scaffold proteins of the present invention are set out in SEQ ID NOS 74 to 79. Variants of these proteins according to the criteria set out above are also embodiments of the present invention.

In another aspect the present invention provides a polynucleotide which encodes a synthetic protein according to the present invention. The polynucleotide may be DNA or RNA. If the polynucleotide is DNA, it may be in single stranded or double stranded form. The single strand might be the coding strand or the non-coding (anti-sense) strand.

A polynucleotide according to the present invention may comprise a sequence encoding a protein having a sequence according to SEQ ID NO 1, 2, 3, 4, 5 or 6, or any one of SEQ ID NO 74 to 79, or a variant thereof, as discussed above.

In one embodiment the polynucleotide comprises the sequence:
aacgctctgctggaattcgttcgtgttgttaaagctaaagaacaggttgttgctggtaccatgtacta
cctgaccctggaagctaaagacggtggtaaaaagaaactgtacgaagctaaagtttgggttaaaccgt
gggaaaacttcaaagaactgcaggagttcaaaccggttggtgacgct
(SEQ ID NO 7), which encodes the amino acid set out in SEQ ID NO 1.

In another embodiment the polynucleotide comprises the sequence:
ggtaacgaaaactccctggaaatcgaagaactggctcgtttcgctgttgacgaacacaacaaaaaaga
aaacgctctgctggaattcgttcgtgttgttaaagctaaagaacaggttgttgctggtaccatgtact
acctgaccctggaagctaaagacggtggtaaaaagaaactgtacgaagctaaagtttgggttaaaccg
tgggaaaacttcaaagaactgcaggagttcaaaccggttggtgacgct
(SEQ ID NO 8), which encodes the amino acid set out in SEQ ID NO 2.

In another embodiment the polynucleotide comprises the sequence:
Gctaccggtgttcgtgcagttccgggtaacgaaaactccctggaaatcgaagaactggctcgtttcgc
tgttgacgaacacaacaaaaaagaaaacgctctgctggaattcgttcgtgttgttaaagctaaagaac
aggttgttgctggtaccatgtactacctgaccctggaagctaaagacggtggtaaaaagaaactgtac
gaagctaaagtttgggttaaaccgtgggaaaacttcaaagaactgcaggagttcaaaccggttggtga
cgct
(SEQ ID NO 9), which encodes the amino acid set out in SEQ ID NO 3.

Polynucleotides that have a nucleic acid sequence that is a variant of the identified nucleic acid sequences may be isolated by a method comprising the steps of: a) hybridizing a DNA comprising all or part of the identified sequence as reflected in any one of SEQ ID NOs 7, 8 or 9, under stringent conditions against nucleic acids of interest; and b) isolating said nucleic acids by methods known to a person skilled in the art. The hybridization conditions are preferably highly stringent.

According to the present invention the term 'stringent' means washing conditions of 1 x SSC, 0.1% SDS at a temperature of 65°C; 'highly stringent' conditions refer to a reduction in SSC towards 0.3 x SSC, more preferably to 0.1 x SSC. Preferably the first two washings are subsequently carried out twice each during 15-30 minutes. If there is a need to wash under highly stringent conditions an additional wash with 0.1x SSC is performed once during 15

minutes. Hybridization can be performed overnight in 0.5 M phosphate buffer pH 7.5 with 7% SDS at 65°C. Such hybridization methods are disclosed in any standard textbook on molecular cloning, for example: Molecular Cloning: a laboratory manual, 3rd ed.; editors: Sambrook *et al.*, CSHL press, 2001.

Variants of the sequences depicted in SEQ ID NOs 7, 8 or 9 may also be identified by comparing the sequence *in silico* to other sequences that may be comprised in a computer database. Sequences may be compared with sequences in databases using a BLAST program (BLASTF 2.1. 2 [Oct.-19-2000]) (Altschul, SF, TL Madden, AA Schaffer, J Zhang, Z Zhang, W Miller, and DJ Lipman, "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", Nucleic Acids Res. 1997,25: 3389-3402).

Preferred embodiments of the invention are polynucleotides encoding proteins having at least 50%, more preferably 70%, more preferably 80%, more preferably 95% identity with SEQ ID NOS 1-6, and yet more preferred are those polynucleotides encoding proteins having at least 97% identity with SEQ ID NOS 1, 2, 3, 4, 5 or 6, or any one of SEQ ID NO 74 to 79,, with those encoding proteins having at least 98 or 99% identity being more preferred. Most preferred are polynucleotides encoding the protein of SEQ ID NO 1, 2, 3, 4, 5 or 6, or any one of SEQ ID NO 74 to 79.

Due to the degeneracy of the genetic code, polynucleotides encoding an identical or substantially identical amino acid sequence may utilize different specific codons. All polynucleotides encoding the proteins as defined above are considered to be part of the invention.

In this regard it should be noted that the sequences discussed above have been optimised for expression in the bacterial host *Escherichia coli*, as this is a convenient system and is typically used for the techniques such as phage display and for subsequent production of selected binding proteins. However, anyone skilled in the art could choose to express the proteins in an alternative hosts system, and in that case a change in the nucleotide sequence that encodes the selected protein (i.e. codon optimisation) may be beneficial, and are thus part of the present invention. Optimising codon use for any particular host is routine for the skilled person.

In particular preferred embodiments the polynucleotides according to the invention are isolated polynucleotides comprising a sequence having at least 80% identity to the nucleic

acid sequence of SEQ ID NOs 7, 8, or 9 (excluding sequences encoding heterologous peptides, where relevant)..

Polynucleotides according to the present invention can, of course, contain additional sequences, such as sequences encoding heterologous peptides, expression control sequences or sequences encoding other proteins or protein subunits.

More preferred are those polynucleotides comprising a sequence having at least 90% identity, and yet more preferred at least 95% identity, preferably 99% identity, most preferred 100% identity to the entire sequence of SEQ ID NOs 7, 8 or 9 (excluding sequences encoding heterologous peptides, where relevant).

Also included within the definition of polynucleotides are modified RNAs or DNAs. Modifications in the bases of the nucleic acid may be made, and bases such as inosine may be incorporated. Other modifications may involve, for example, modifications of the backbone.

"% identity" defines the relation between two or more polynucleotides or polypeptides on the basis of a comparison between their aligned sequences.

Identity can be calculated by known methods. Identity, or homology, percentages as mentioned herein in respect of the present invention are those that can be calculated with the GAP program, running under GCG (Genetics Computer Group Inc., Madison, WI, USA).

For polypeptide sequence comparison:

- Alignment algorithm: Needleman and Wunsch, J. Mol. Biol. 1970, 48: 443-453.
- As a comparison matrix for amino acid similarity the Blosum62 matrix is used (Henikoff and Henikoff, supra).
- The following gap scoring parameters are used:
    - Gap penalty: 12
    - Gap length penalty: 2
    - No penalty for end gaps.

For nucleotide sequence comparison:

- Alignment algorithm: Needleman and Wunsch (supra).
- Comparison matrix:
    - match = +10, mismatch = 0.

- Gap penalty: 50.
- Gap length penalty: 3.

Approaches to developing consensus stabilised proteins that include the incorporation of other parameters, such as phylogenetically unbiased methods as described by Jaeckel, Bloom et al. 2010, can also be used.

However, as discussed above, it will be noted that the various sequence identity calculations should be amended to take account of the situation where one or more heterologous peptides are optionally inserted into the synthetic protein. When such an insertion has occurred, the inserted sequence which encodes the peptide(s) should generally be disregarded when calculating sequence identity for the reasons set out above.

Nucleic acids, especially DNA, according to the invention will be useful for *in vivo* or *in vitro* expression of the encoded protein. When the polynucleotides according to the invention are used for expression of the encoded proteins, the polynucleotides may include, in addition to the coding sequence for the protein, other coding or non-coding sequences, for example, leader sequences or fusion portions, linker sequences, marker sequences, promoters, enhancers and the like.

The polynucleotides according to the invention may be used in the production of recombinant proteins according to the invention. This can be achieved through expression of the proteins in a suitable host cell or multicellular organism. When used for expression of an encoded polypeptide, the polynucleotides may advantageously include, in addition to the coding sequence for the polypeptide, other coding sequences, for example, signal sequences, leader sequences, targeting sequences, fusion portions, marker sequences, sequences to assist in purification and the like.

A wide variety of host cell and cloning vehicle combinations can be used for cloning and expression. Polynucleotides of the present invention may be cloned into any appropriate expression system. Suitable expression systems include bacterial expression system (e. g. *Escherichia coli* DH5α and BL21™(DE3)), a viral expression system (e.g. Baculovirus), a yeast system (e.g. *Saccharomyces cerevisiae*) or eukaryotic cells (e.g. COS-7, CHO,BHK, HeLa, HD11, DT40, CEF, or HEK-293T cells) A wide range of suitable expression systems are available commercially. Typically the polynucleotide is cloned into an appropriate vector under control of a suitable constitutive or inducible promoter and then introduced into the host cell for expression.

In another aspect the present invention therefore provides a recombinant vector comprising a polynucleotide according to the invention. Suitable vectors include bacterial or yeast plasmids, cosmids, phagemids, fosmids, wide host range plasmids and vectors derived from combinations of plasmid and phage or virus DNA. An origin of replication and/or a dominant selection marker can suitably be present in the vector.

The vectors according to the invention are suitable for transforming a host cell. Examples of suitable cloning vectors are plasmid vectors such as pBR322, the various pUC, pET, pEMBL and Bluescript plasmids, or viral vectors such as retroviruses, lentiviruses, adenoviruses, adeno-associated viruses. pcDNA3.1 is a particularly preferred vector for expression in animal cells.

Particularly useful for the cloning and expression of the libraries of scaffold proteins are phage or phagemid vectors such as those derived from filamentous phage such as M13 or fd or caspid phage such as bacteriophage T7.

When used in the expression of the proteins of the present invention, a vector according to the present invention typically comprises an expression control sequence operably linked to the nucleic acid sequence coding for the protein to control expression of the relevant polynucleotide. Such expression control sequences generally comprise a promoter sequence and additional sequences which regulate transcription and translation and/or enhance expression levels. Suitable expression control sequences are well known in the art and include eukaryotic, prokaryotic, or viral promoter or poly-A signal. Expression control and other sequences will, of course, vary depending on the host cell selected or can be made inducible. Examples of useful promoters are the tac promoter (Deboer, Comstock et al. 1983), T7 promoter (Studier and Moffatt 1986), SV-40 promoter (Science 1983, 222: 524-527), the metallothionein promoter (Nature 1982,296: 39-42), the heat shock promoter (Voellmy et al., P.N.A.S. USA 1985,82: 4949-4953), the PRV gX promoter (Mettenleiter and Rauh, J. Virol. Methods 1990, 30: 55-66), the human CMV IE promoter (US 5,168,062), the Rous Sarcoma virus LTR promoter (Gorman et al., P.N.A.S. USA 1982, 79: 6777-6781), or human elongation factor 1 alpha or ubiquitin promoter. Prokaryotic control sequences include, for example, lac, tac T7, ara, and other known promoters. Prokaryotic promoters in general which can be used in the present invention are discussed in Goldstein MA, Doi RH, 'Prokaryotic promoters in biotechnology'. Biotechnol Annu Rev. 1995;1:105-28. Many other suitable control sequences are known in the art, and it would be routine for the skilled person to select suitable sequences for the expression system being used.

After the polynucleotide has been cloned into an appropriate vector, the construct may be transferred into the cell (e.g. bacterium or yeast) by means of an appropriate method, such as electroporation, $CaCl_2$ transfection or lipofectins. When a baculovirus expression system is used, the transfer vector containing the polynucleotide may be transfected together with a complete baculo genome.

These techniques are well known in the art and the manufacturers of molecular biological materials (such as Clontech, Agilent, Promega, and/or Invitrogen Life Technologies) provide suitable reagents and instructions on how to use them. Furthermore, there are a number of standard reference text books providing further information on this, e.g. Rodriguez, R. L. and D. T. Denhardt, ed., "Vectors: A survey of molecular cloning vectors and their uses", Butterworths, 1988; Current protocols in Molecular Biology, eds.: F. M. Ausubel *et al.*, Wiley N. Y. , 1995; Molecular Cloning: a laboratory manual, supra; and DNA Cloning, Vol. 1-4, 2nd edition 1995, eds.: Glover and Hames, Oxford University Press).

In a further aspect the present invention also provides a cell capable of expressing a recombinant protein, characterised in that the cell comprises a polynucleotide according to the invention encoding the recombinant protein to be expressed. Suitably the cell is a host cell transformed with a polynucleotide or vector as described above. The polynucleotide or vector according to the invention can be stably integrated into the genomic material of the cell or can be part of an autonomously replicating vector. "Recombinant" in this context refers to a protein that is not expressed in the cell in nature.

Host cells can be cultured in conventional nutrient media which can be modified, e.g. for appropriate selection, amplification or induction of transcription and thus expression of the recombinant protein. The host cells can be prokaryotic or eukaryotic. Mammalian, e.g. human, cell lines may be preferred, especially where the expressed proteins are intended for in vivo use in human or mammalian subjects. Suitable exemplary cell lines are mentioned above, and appropriate culture conditions for the various suitable cell types are well known to the person skilled in the art.

In a further aspect the present invention provides a cell culture comprising cells according to the invention.

In a further aspect the present invention provides a method of screening for a synthetic protein that binds to a target, the method comprising:

a) Providing a library comprising a population of synthetic scaffold proteins comprising one or more heterologous peptide sequences as described above;
b) Exposing said library to a target; and
c) Selecting synthetic scaffold proteins which bind to said target.

The target may suitably be a protein/peptide, a nucleic acid, or a small molecule or any other target of interest. The target can suitably be immobilised on a substrate, e.g. to the wells of a microtiter plate or on magnetic beads or other appropriate particles.

The method of screening may suitably be a display system, e.g. a surface display system. For example, a suitable surface display system could be a phage display, mRNA display, ribosome display, CIS (Odegrip, Coomber et al. 2004) or other non-covalent or covalent protein display (FitzGerald 2000), bacterial display (Lofblom 2011) and yeast display (Traxlmayr and Obinger 2012) or display on another eukaryotic cell (Makela and Oker-Blom 2008)l (Ho and Pastan 2009).

In certain embodiments of the present invention the proteins used in the screening method are thus fusions of the scaffold protein with a bacteriophage coat protein, so that the scaffold proteins are displayed on the surface of the viral particle. The protein displayed on the viral particle thus corresponds to the genetic sequence within the phage particle. Of course, where the surface display system is other than a phage display system the other suitable fusion proteins can be provided.

Where the target is immobilised on a substrate the surface displayed library is exposed to the substrate with the target immobilised thereon and, after a suitable period to allow the phage time to bind to the target, the substrate is rinsed. Proteins that bind to the target remain attached to the substrate, while others are washed away.

Suitably proteins which bind to the target are thereafter eluted or the associated phage cleaved from the target and, in the case of phage display, used to create more phage by infection of suitable bacterial hosts. This results in a new library which comprises an enriched mixture, containing considerably less irrelevant phage (i.e. phage which contains non-binding scaffold proteins) than were present in the initial library.

The steps of exposing the library to the target, rinsing, elution and infection are optionally repeated one or more times, further enriching the phage library in binding proteins. This is

generally referred to as panning. Thus the method may involve performing multiple panning steps.

Once a desired number of panning steps have been performed the nucleic acid linked to the surface displayed protein can be sequenced to identify the scaffold proteins which bind to the target.

It has been found that scaffold proteins according to the present invention can be obtained which bind with high affinity and high specificity to targets. For example scaffold proteins which bind with an affinity of 1×10^{-9} M have been obtained. Unexpectedly the libraries formed from scaffold proteins according to the present invention have revealed scaffold proteins that bind to proteins against which it has proved impossible to raise antibodies or antibodies with the desired specificity. For example, artificial binding proteins based on the scaffold have been selected proteins that specifically bind to Human Papilloma Virus (HPV) 16 E5, an understudied protein due to the repeated failure of ability to select lack of suitable antibodies against this target. HPV is a causative agent of cervical carcinoma. Artificial binding proteins have also been selected that discriminate between the two forms of the human Small Ubiquitin-like Modifier (SUMO), hSUMO1 and hSUMO2 whereas previously selected antibodies proved incapable of such discrimination. It is reasonable to expect that other useful binding proteins could be raised against other targets where antibody-based techniques have failed to provide a suitable solution.

In a further aspect the present invention provides a synthetic scaffold protein obtained by a screening method described above.

In a further aspect the present invention provides the use of synthetic protein as set out above, or a nucleic acid encoding such a synthetic protein, in research. For example, such proteins can be used in any area of research where antibodies or antibody fragments are typically used. Such methods may include interfering with protein/protein interactions, labelling targets (e.g. fluorescently), etc.

In a further aspect the present invention provides the use of synthetic protein as set out above, or a nucleic acid encoding such a synthetic protein, in environmental and security monitoring, or in synthetic biology.

In a further aspect the present invention provides the use of synthetic protein as set out above, or a nucleic acid encoding such a synthetic protein, in target detection.

In a further aspect the present invention provides the use of a synthetic scaffold protein as set out above, or a nucleic acid encoding such a synthetic protein, in therapy or diagnosis. For example, such proteins can be used in any area of diagnosis or therapy where antibodies or antibody fragments are typically used.

In a further aspect the present invention provides a pharmaceutical preparation comprising a synthetic protein as set out above and a pharmaceutically acceptable carrier or excipient.

In a further aspect the present invention provides a method for validating drug targets and identifying druggable domains on target proteins using synthetic protein as set out above.

In a further aspect the present invention provides a suitable method for selecting synthetic proteins that bind to small molecules including organic compounds and including drug compounds.

Suitably the synthetic proteins of the present invention can be linked to a suitable therapeutic agent. In such a situation the synthetic protein can provide a targeting function to deliver the therapeutic 'payload' to a desired target.

Preparation of pharmaceutical preparations according to the present invention is carried out by means conventional for the skilled person. Methods for preparing administrable compositions, whether for intravenous or subcutaneous administration or otherwise, will be known or apparent to those skilled in the art and are described in more detail in such publications as Remington: The Science and Practice of Pharmacy, Lippincott Williams and Wilkins; 21st Revised edition (1 May 2005).

'Synthetic protein' refers to a protein which is not found in nature and which has been formed through recombinant techniques.

The term 'protein' is generally used interchangeably with 'peptide' or 'polypeptide', and means at least two covalently attached amino acids linked by a peptidyl bond. The term protein encompasses purified natural products, or products which may be produced partially or wholly using recombinant or synthetic techniques. The term protein may refer to an aggregate of a protein, such as a dimer or other multimer, a fusion protein, a protein variant, or derivative thereof. The term also includes modified proteins, for example, a protein

modified by glycosylation, acetylation, phosphorylation, pegylation, ubiquitination, and so forth.

A protein may comprise amino acids not encoded by a nucleic acid codon. Unnatural amino acids can be introduced by post-translational modification, for example through the introduction of a cysteine or other appropriate residue at a specific position and then the conversion of this following protein purification to a dehydroalanine and then chemical reaction with a suitably modified unnatural amino acid. Alternatively, an unnatural amino acid could be incorporated into the sequence during translation either in vitro or in vivo through the suppression of a stop codon, normally UAG via an appropriate heterologous tRNA/tRNA synthetase couple that charges a suppressing tRNA with an unnatural amino acid that can then be incorporated into the protein at the position of the UAG codon.

The protein of the present invention can be a single subunit protein or a multi-subunit protein wherein at least one subunit comprises the said sequence.

In certain embodiments the protein can comprise two or more synthetic proteins as set out above, e.g. in the form of a fusion protein. In such a case each of the synthetic proteins can bind to the same target or they can bind to different targets. Binding the same can be increase avidity and apparent binding affinity, whilst those binding distinct targets can be useful for the development of cross-linking or dual recognition modules.

'Heterologous peptide' or 'Heterologous peptide sequence' in the context of the present invention refers to a peptide sequence which is not normally found in the relevant location, e.g. in the loop region of the scaffold protein. Typically such peptides are relatively short, i.e. containing fewer than 30 amino acids, typically 20 or fewer. Such a peptide can have essentially any sequence. Indeed it is desirable that libraries containing a vast number of heterologous peptide sequences will be displayed in scaffold proteins.

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. 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. 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.

Specific Embodiments of the Invention

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

- Figure 1 shows a consensus phytocystatin PHYTC57 derived from 57 phytocystatin amino acid sequences (A) PHYTC57 synthetic gene shown as a double strand sequence together with the overlapping oligonucleotides (P1 to P6) used to generate the gene by recursive PCR. The coding sequence is also shown as single letter amino acid code. The positions of the two restriction sites *SfiI* and *NotI* are shown. (B) Schematic representation of the cloning region of pDHisII which is based on pHEN1. The positions of the relevant regions encoding the pelB signal sequence, hexhistidine tag and N-terminal section of the M13 phage PIII protein are shown. The positions of the standard M13 primer binding sites for M13R and P10 are indicated as are the unique *Sfi*I and *Not*I restriction sites.
- Figure 2 shows graph of the results of papain inhibition assays for a modified oryzacystatin lacking residue Asp86 (OSA-IΔD86) and PHYTC57 at varying concentrations of phytocystatins, showing the enhanced efficacy of PHYTC57.
- Figure 3 shows surface plasmon resonance measurement of the interaction between cystatins immobilized on an NTA sensorchip, and papain. The experiments were performed at several concentrations of papain and data were analysed using the Biaevaluation3 software package (BIACORE) with global fitting of the data to the Langmuir 1:1 binding model. OSA-IΔD86, a modified rice cystatin I; CPA, papaya cystain; CUN, orange cystatin; CEWC, chicken egg white cystatin; PHYTC57, consensus cystatin.
- Figure 4 shows (A) Thermal stability of OSA-IΔD86 (grey square) and PHYTC57 (black triangle) shown as residual enzyme activity assayed at 25 °C following incubation at 100 °C for the times shown. PHYTC57 displays greater thermal stability than does OSA-IΔD86), B) Effect of simulated gastric fluid treatment on the inhibitory properties of OSA-

IΔD86 (grey square) and PHYTC57 (black triangle) showing that PHYTC57 retains activity over a longer period of time than does OSA-IΔD86. (C) SDS-PAGE analysis of the stability of PHYTC57 and OSA-IΔD86 to incubation with simulated gastric fluid. The time of incubation is shown in seconds and the positions of marker proteins (M) are indicated on the left in kDa with the position of pepsin from the assay and the cystatin indicated. (D) Western blot analysis of PHYTC57 following simulated gastric fluid treatment for the times indicated (seconds) using an anti-His tag antibody. The positions of marker proteins (M) are indicated in kDa.

- Figure 5 PHYTC57 and OSA-IΔD86 comparison showing amino acid differences between the proteins PHYTC57 and OSA-IΔD86. (A) Alignment of protein sequences. (B) Representation of the position of the amino acid changes from (A) on the 3D structure of OSAI with residues changes labeled. The binding region is also shown with the N-terminal, QVVAG and PW loops labeled.
- Figure 6 shows a crystal structure of an Adhiron isolated from the library and sequence of an Adhiron scaffold derived from PHYTC57. (A) The Adhiron scaffold is shown and the inserted loops are indicated. (B) Codon optimised nucleic acid sequence and amino acid sequence for the Adhiron 81 amino acid scaffold. The alpha helix, beta sheets and the insertion regions for loop1 and loop2 are highlighted.
- Figure 7 shows biochemical characterisation of Adhiron scaffold and sequencing of the library. (A) Differential scanning calorimetry was performed to determine the melting temperature of the Adhiron scaffold (Tm 101°C). (B) Circular dichroism was used to examine the structure of the Adhiron scaffold and of three selected Adhiron proteins containing loop insertions and all show very high β structure. (C) The Adhiron phage library was used to infected *E. coli* ER2738 cells. 96 random clones were isolated and sequenced. The graph represents the percentage of each amino acid within the loop regions. An ideal library would contain 5.26% of each amino acid; cysteine was not included in the library.
- Figure 8 shows a comparison of the stability of (A) the Adhiron scaffold compared with (B) a representative small soluble well characterised protein, lysozyme, by differential scanning calorimetry. (C) shows that for an Adhiron selected to bind to a myc antibody, the addition of the loops into the scaffold reduces the Tm to 85 °C but this still represents a higher melting temperature than most scaffold proteins. This Adhiron protein can undergo repeated cycles of denaturation and renaturation as shown by the series of scans.
- Figure 9 shows phage ELISA results for yeast SUMO (ySUMO). (A) Phage ELISA using 24 clones isolated from the third pan round. Phage produced by each clone were

incubated in wells containing ySUMO or control. The image was recorded three minutes after addition of 3,3',5,5'-Tetramethylbenzidine (TMB) substrate. (B) Graph showing the absorbance at 560 nm of the phage ELISA for ySUMO (hatched) and control (white) wells.

- Figure 10 shows purification and characterisation of an Adhiron specific for yeast SUMO (Ad-ySUMO). (A) Ad-ySUMO was expressed in BL21 cells and cell lysates were heated to 50 °C, 60 °C, 70 °C, 80 °C, 90 °C and 100 °C for 20 minutes and the precipitate was removed by centrifugation at 15,000 xg. Aliquots of 5 µl of cleared lysates for each temperature were separated on a 15% SDS-PAGE gel, and stained with coomassie to visualise the proteins. (B) Lysates were incubated with Ni-NTA beads for 1 hr. Post incubated lysate (5 µl) and purified Ad-ySUMO (10 µl) were run on a 15% SDS-PAGE gel and stained with coomassie for visualisation. (C) Biotinylated Ad-ySUMOs were used to detect ySUMO (hatched bars) and did not detect human SUMO (white bars) by ELISA. (D) Western blots using biotinylated Ad-ySUMO clones 10, 15, 20, and 22 against 0.5 µg of yeast SUMO (upper panel) and mixed with 20 µg of HEK293 cell lysate (lower panel).
- Figure 11 shows phage ELISA results for Adhirons identified in screens against a series of targets, i.e. growth factor protein FGF1 (A), a cell surface receptor CD31 (B), and a peptide (C). Graphs represent the absorbance readings of each well after the addition of TMB. The wells containing the target are shown as hatched bars whilst the control wells are shown as white bars.
- Figure 12 shows immunofluorescence images of HPV16 E5 GFP (target) and HPV16 E5 GFP without epitope for Adhiron (control). Adhiron E5 was conjugated to quantum dots and used to detect E5 protein in the mammalian cells. Cells were stained with DAPI (DNA stain), GFP, and E5 (with the Adhiron-Quantum dots).
- Figure 13 shows an Adhiron targeting hSUMO2. (A) An ABP raised against hSUMO2 was tested to determine specificity for hSUMO1 and hSUMO2 by ELISA. The hSUMO2 ABP specifically bound to hSUMO2, which was reflected in the binding affinity. (B) Blot showing the ability of the hSUMO2 ABP to inhibit RNF4's SUMO-targeted ubiquitin ligase activity. (C) A control vector or hSUMO2 binder was expressed in cells and analysed using an anti-FLAG antibody and cultured with arsenic to induce nuclear bodies, PML (green). RNF4 promotes degradation of PML. Blocking the interaction between hSUMO2 and RNF4 alters PML degradation. This is the first description of a hSUMO2 binding protein that specifically binds to and inhibits hSUMO2 without interacting with hSUMO1. The hSUMO2 Adhiron also specifically blocks the domain that

it's interacting with and does not affect other functions of the hSUMO2, as demonstrated in (C).

- Figure 14 shows a graph that represents a blood clot formation and lysis turbidity assay. The solid line represents the normal formation and lysis. The dashed lines represent the effect of five different Adhirons on this process. The five different Adhirons contain different epitopes and are having different effects on clot formation and lysis. Some are prolonging clotting time, some are prolonging lysis time and some are prolonging clotting and lysis time. One of the Adhirons is completing inhibiting clot formation. This demonstrates that the different Adhirons are binding to and inhibiting different regions of fibrinogen and therefore represent a really novel way of studying protein function.
- Figure 15 shows a confocal image of FITC labelled fibrinogen after clot formation with a fibrinogen binding Adhiron and a control Adhiron that does not bind fibrinogen. This demonstrates the ability of the Adhiron to modify the normal clot response.
- Figure 16 shows fluoroscopy images that demonstrate the expression of functional Adhirons in mammalian cells. It can be seen that the human SUMO2 binding Adhiron alters the degradation of the nuclear phosphoprotein PML leading to an increase in these PML nuclear bodies.
- Figure 17 shows the co-crystal structure of FcγRIIIa and bound Adhiron. (A) The Adhiron binds to an allosteric site that affects receptor binding to IgG. This site has also been identified as a target for small molecule drug design providing evidence that Adhirons can provide important information about druggable sites. (B) Adhirons have also been found to target the direct binding site of IgG on FcγRIIIa. The Adhiron contained two loops and a 4 amino acid unstructured N-terminal sequence. The N-terminus peptides are also contributing to the interaction between Adhiron and FcγRIIIa. The information gained by understanding the binding interaction between the Adhiron and FcγRIIIa will help identify small molecules via in silico screening. This provides an intriguing novel approach for future drug discovery methodologies. In addition the crystal structures demonstrate the ability of the scaffold to present loops that can either extend the beta sheets of the core scaffold or more flexible loops that create alternative conformational interaction surfaces. This is facilitated by the inherent core stability of the scaffold, which potentially makes this scaffold unique.
- Figure 18. NMR spectra. (A) Overlay of 1H-15N HSQC fingerprint spectra for the ABP scaffold (light grey) and Yeast Sumo Adhiron 15 (dark grey). (B) Overlay of 1H-15N HSQC fingerprint spectrum for Yeast SUMO Adhiron 15 (dark grey) and 1H-15N TROSY HSQC spectrum for the Yeast SUMO protein and Yeast SUMO- Adhiron 15 complex (light grey).

- Figure 19 shows a graph of the change of impedance versus concentration of SUMO binding Adhiron. This provides an example of the potential use of Adhirons in impedance based biosensor devices and shows that Adhirons bound to a surface have a larger dynamic range compared to antibodies.
- Figures 20 to 24 show sequence alignments of the LOOP 1 and LOOP2 regions of Adhirons selected against a range of targets. This analysis allows identification of the range of binders against a given target and facilitates the development of potential consensus binding regions in some cases.
- Figure 20 shows a sequence alignment of LOOP1 and LOOP2 regions of several Adhirons that bind to Lectin-like oxidized LDL receptor-1 (LOX1).
- Figure 21 shows a sequence alignment of LOOP1 and LOOP2 regions of several Adhirons that bind to Human Growth Hormone (HGH).
- Figure 22 shows a sequence alignment of LOOP1 and LOOP2 regions of several Adhirons that bind to yeast small ubiquitin-like modifier (SUMO).
- Figure 23 shows a sequence alignment of LOOP1 and LOOP2 regions of several Adhirons that bind to penicillin binding protein 2a (PBP2a).
- Figure 24 shows a sequence alignment of LOOP1 and LOOP2 regions of several Adhirons that bind to a peptide target.
- A Figure 25 shows phage ELISA results for Adhirons identified in screens against an organic compound, posaconazole. Graphs represent the absorbance readings of each well after the addition of TMB. The wells containing the target are shown as hatched bars whilst the control wells are shown as white bars.
- Figure 26. A graph of the change of impedance based detection of various concentrations of fibrinogen from micromolar to attomolar by a fibrinogen binding Adhiron. This provides an example of the potential use of Adhirons in impedance based biosensor devices and shows that Adhirons bound to a surface can detect low concentrations of target protein within a 15 minute incubation and display a large linear dynamic range. The two data points at 3×10^{-12} micromolar were measured immediately after adding the analyte (lower data point) and after a 15 minute incubation (upper data point).
- Figure 27 shows a sequence alignment of LOOP1 and LOOP2 regions of several Adhirons that bind to Growth factor receptor-bound protein 2 (Grb2) Src homology 2 domain.
- Figure 28 shows a sequence alignment of LOOP1 and LOOP2 regions of several Adhirons that bind to Signal Transducer and Activator of Transcription 3 (STAT3) Src homology 2 domain.

It has been stated that "To prompt wider interest in a particular protein scaffold, it is necessary to demonstrate that specificities for different kinds of relevant ligands can be generated, that the derived binding proteins are practically useful, and that they offer at least some benefits over conventional antibody fragments. These criteria are not met by many of the protein scaffolds proposed so far and for most of them merely initial engineering efforts have been described." (Skerra 2007). The present invention is therefore of clear value in providing a useful, versatile and attractive protein scaffold, as demonstrated below.

The high level of success in terms of identifying important bioactive binding proteins attributable to the novel Adhiron scaffold and derived library is due in large part to the very high stability of the core scaffold which provides a highly rigid framework upon which variable loop regions can be displayed. These loop regions have sufficient flexibility to adopt a range of conformations and can thus interact with a wide range of conformational features on target molecules allowing selection of binding reagents against a wide range of molecular targets, including notable small organic molecules. These structural aspects of this unique scaffold lead to functional outcomes that have not been achieved with other scaffold proteins.

There may also be benefits in the use of a plant-based protein for many applications, particularly those involving humans since the protein is not derived from a human protein and therefore will not be involved in natural interactions with human proteins. The only potential interactions would be against cysteine proteases but the active binding regions that could interact with such proteins are typically replaced or removed in the Adhiron scaffolds. However, humans come into contact with and tolerate plant-derived proteins constantly through for example, food, cosmetic products, and medicinal compositions.

CREATION OF AN EXEMPLARY SCAFFOLD PROTEIN (termed PHYTC57)

Introduction

A consensus approach to protein design starts with a multiple sequence alignment of members of a protein family to derive a single consensus sequence in which each position is normally occupied by the residue that occurs most frequently. Residues that define the structure, folding pathway, and stability of the folded protein will tend to be conserved, while those required for a common biological function such as catalytic residues in an enzyme, or residues that interact with a conserved target protein, are also likely to be conserved. A

natural protein will not usually contain all the conserved consensus residues because proteins only evolve to be sufficiently stable to perform their biological role *in vivo*, and in many cases this may include a degree of instability to facilitate turnover and regulation of biological processes (Steipe, Schiller et al. 1994).

We were interested to explore whether a consensus approach could be applied to enhance the inhibitory properties and stability of phytocystatins. Phytocystatins are small (~100aa) protein inhibitors of cysteine proteases (Kondo et al. 1991). A detailed phylogenomic analysis of the cystatin superfamily reveals the relationships between the different classes of cystatins (Kordis and Turk 2009). Phytocystatins comprise three regions that are involved in binding to cysteine proteases, the N-terminal region, a QVVAG loop and a PW loop (Margis et al. 1998). Previously, based on a structural model and sequence alignments, we generated a variant of rice cystatin, oryzacystatin I (OSA-I or OC-I), that lacks residue Asp 86, close to the conserved PW loop and named OSA-IΔD86. This variant displayed a 13-fold improvement in K_i against both papain and the *Caenorhabditis elegans* gut-specific cysteine protease GCP-1 (Urwin et al. 1995; McPherson et al. 1997; Urwin et al. 1997; Urwin et al. 1998; Urwin et al. 2000; Urwin et al. 2001; Urwin et al. 2003; Lilley et al. 2004). We then demonstrated good levels of transgenic resistance, against plant nematodes, conferred by OSA-IΔD86 and other cystatins when expressed in the specialised feeding cells that develop within a nematode parasitized plant (Urwin et al. 1995; McPherson et al. 1997; Urwin et al. 1997; Urwin et al. 1998; Urwin et al. 2000; Urwin et al. 2001; Urwin et al. 2003; Lilley et al. 2004).

Here we describe the design, construction and characterisation of a consensus phytocystatin based on multiple sequence alignment of the amino acid sequences of 57 phytocystatins. We show that this consensus phytocystatin displays efficient inhibition of the cysteine protease papain and enhanced stability in thermostability and digestibility assays.

Materials and methods

Design of the consensus phytocystatin hereinbelow referred to as PHYTC57

Following a tBLAST search of the GenBank data base a multiple alignment of phytocystatin sequences was performed using CLUSTALW (http://clustalw.genome.ad.jp/; BLOSUM62; Gap opening penalty = 12; Gap extension penalty = 2) with further manual alignment using the program CINEMA (http://www.bioinf.manchester.ac.uk/dbbrowser/CINEMA2.1/) (Parry-Smith et al. 1998; Lord et al. 2002). The most commonly used amino acid at each position was determined and the variable N- and C-terminal ends were truncated to give a

consensus protein of 95 amino acids in length. A synthetic coding region was designed to encode the consensus phytocystatin (phytc57) with codon usage optimised for expression in *E. coli* by using the appropriate codon frequency table (http://www.kazusa.or.jp/codon/). The coding region was constructed from six oligonucleotides (P1 – P6; see Figure 1) each of approximately 70 nt in length that included regions of at least 10 nt overlap between adjacent oligonucleotides. The flanking oligonucleotides P1 and P6 also contained an *Sfi*I site and a *Not*I site, respectively, for cloning into the vector pDHisII where the coding sequence was fused to the vector–derived *pelB* signal sequence coding region. Retention of a signal peptidase cleavage site was checked by submitting the N-terminal sequence of the PELB/PHYTC57 protein to the automatic signal sequence prediction programme Signal P (Bendtsen et al. 2004); http://www.cbs.dtu.dk/services/SignalP/). Pairs of oligonucleotides (P1 + P2), (P3 + P4) and (P5 + P6) were annealed and converted to double stranded fragments by self-priming. The reactions (10 µl) contained 25 pmole of each primer, 0.2 mM of each dNTP, 1x *Pwo* Buffer (10 mM Tris-HCl pH 8.85, 25 mM KCl, 5 mM $(NH_4)_2SO_4$, 2 mM $MgSO_4$) and 5 units *Pwo* DNA polymerase (Boehringher) and were subjected to one cycle of 95 °C for 2 min, 25 °C for 4 min and then 72 °C for 5 min. The subsequent joining of the products (P1/P2) to (P3/P4) and then (P1/P2/P3/P4) to (P5/P6) was performed in an identical manner, to generate full-length product. A 5 µL aliquot of the final reaction was used as template in a 50 µL PCR together with 50 pmole each of the flanking primers P1 and P6, 0.2 mM each dNTP, 1x *Pwo* buffer and 5 units *Pwo*. The reaction was subjected to 95 °C for 1 min, then 30 cycles of 95 °C, 30 sec, 55 °C, 30 sec and 72 °C, 30 sec. The product was digested with *Sfi*I and *Not*I, recovered from an agarose gel and cloned into *Sfi*I and *Not*I restricted pDHisII. The newly constructed gene region was sequenced to confirm correct primer assembly and the expected DNA sequence. The sequence of the consensus coding region is shown in Figure 1A.

Synthesis of other phytocystatin genes

Synthetic DNA sequences optimised for *E. coli* expression were similarly designed, constructed and cloned for phytocystatins from rice (*Oryza sativa*/; *osa-1ΔD86* a modified form of *oc-I* GenBank accession number U54702 lacking the codon for Asp 86 (Urwin *et al.* 1995), satsuma orange (*Citrus unshiu*; *cun*, GenBank accession number C95263) and papaya (*Carcia papaya*; *cpa*, GenBank accession number X71124 (Song *et al.* 1995). The chicken egg white cystatin gene (*cewc*) sequence was amplified from a pQE30-derived recombinant plasmid using the gene specific primers:

cewcF 5'ATTAGCGGCCCAGCCGGCCATGGCCAGCGAGGACCGCTCCCGGC3' (SEQ ID NO 10)

cewcR 5'CGCTGTACTTGCGGCCGCCCTGGCACTTGCTTTCCAGC3' (SEQ ID NO 11)

that introduced an *Sfi*I site and *Not*I site (underscored) respectively.

PCR reactions were carried out with 50 pmole of each primer, a final concentration of 0.2 mM of each dNTP (Promega) and approximately 10 ng template in a volume of 50 µl containing 1x SuperTaq buffer (10 mM Tris-HCl pH 9.0, 1.5 mM $MgCl_2$, 50 mM KCl, 0.1 % (v/v) Triton™ X-100). To ensure high fidelity amplification during the PCR a 10:1 mix of SuperTaq (HT Biotechnology):*Pfu* (Boehringer Mannheim) DNA polymerases were used with 1 unit of polymerase mix per reaction. The reactions were subjected to 1 cycle of 95 °C for 1 min then 20 cycles of 95 °C for 30 sec, 55 °C for 30 sec and 72 °C for 30 sec. A final step of 72 °C for 30 sec was used to ensure that all products were full length.

pDHisII construction

Phytocystatin coding regions were initially cloned into a modified version of the phagemid vector pHEN1 (Hoogenboom et al. 1991) as gene III fusions. The pHEN vector was modified by addition of a hexa-histidine region encoded by complementary oligonucleotides that were phosphorylated and annealed to create a linker with appropriate single strand ends for ligation with the *Not*I cleaved vector. The oligonucleotide sequences were;

HisF: 5'- GGCCGCAGAGGATCGCATCACCATCACCATCACGG -3' (SEQ ID NO 12)

HisR: 5'- GGCCCCGTGATGGTGATGGTGATGCGATCCTCTGC -3' (SEQ ID NO 13)
and the *Not*I complementary ends are underscored.

The pHEN1 vector was digested with *Not*I and dephosphorylated using 5 units of shrimp alkaline phosphatase (NEB) in a 50 µl reaction containing 100 mM NaCl, 50 mM Tris-HCl (pH 7.9), 10 mM $MgCl_2$ and 1 mM DTT for 30 minutes at 37 °C before heating to 65 °C for 15 min. The distal *Not*I site was destroyed by PCR mutagenesis using the primers

XhoF: 5'ATCACGCTCGAGCAGAACAAAAACTCATCTCAG3' (SEQ ID NO 14)
XhoR: 5'TGTTCTGCTCGAGCGTGATGGTGATGGTGATGGCG3' (SEQ ID NO 15)
BamHIR: 5'TGGCCTTGATATTCACAAACG3' (SEQ ID NO 16)
M13R: 5'AGCGGATAACAATTTCACACAGGA3' (SEQ ID NO 17)

The XhoF primer was used together with BamHIR primer and the XhoR together with the M13R primer to generate two fragments that were annealed and amplified in a second PCR

that included primers M13R and BamHIR. The introduced *Xho*I restriction site is underscored. The resulting product was cloned as an *Sfi*I/*Bam*HI fragment into pHEN1 vector similarly digested. The presence of the *Xho*I site was screened for by restriction analysis of isolated phagemid DNA and the insert in a positive clone, pDHisII (Figure 1B), was confirmed by DNA sequence analysis.

Expression of cystatins in pDHisII

Initially, expression of cystatins were performed using constructs in the vector pDHisII which adds C-terminal $RGS(H)_6$ and myc tags. Expression studies in the *E. coli* host strain HB2151 allow suppression of the amber codon within the cystatin-gene III fusion resulting in accumulation of cystatin in the periplasm. Cultures (1L) were grown in 2xTY media with 0.1 % (v/v) glucose and 100 µg/mL ampicillin then induced by IPTG to 1 mM and grown at 30 °C, 16 hours. Cell pellets were resuspended in 20 mL 50 mM Tris pH 8, 20 % (w/v) sucrose at 4 °C and periplasmic preparations performed. To the periplasmic fraction 1 mL Ni-NTA resin (GE Healthcare) was added and incubated with mixing for 16 hours at 4 °C. The resin was washed 8 times with 1 mL aliquots of wash buffer (50 mM $NaHPO_4$, 500 mM NaCl, pH 6), then 3 times with wash buffer containing 40 mM imadiazole. Cystatin was eluted with 200 µL wash buffer containing 250 mM imadiazole at 4 °C for 1 hour, before dialysis against PBS. Protein was aliquoted and frozen in liquid nitrogen. Samples were analysed by SDS-PAGE and electrospray mass spectrometry.

Expression of cystatins in pET101

For more efficient expression of soluble protein, the phytocystatin genes were sub-cloned together with the C-terminal 6-His tag from pDHis II into pET101 by directional TOPO™ cloning (Invitrogen). The primers

cystaSDFWD 5'CACCATGAAATCACTATTGCTTACG3' (SEQ ID NO 18)
cystaSDREV 5'CTACTAGTGATGGTGATGGTGATGCG3' (SEQ ID NO 19)

were used to PCR amplify the phytocystatins genes from the original cloning vector pDHisII using the conditions described above. Positive clones were confirmed by DNA sequence analysis and were introduced into BL21(DE3) Star cells (Invitrogen). Cultures were grown at 30°C and expression of phytocystatins was induced by addition of IPTG (to 1 mM) for 16 h. The cells were harvested by centrifugation (4,000 x *g*, 10 min), sonicated on ice (3 x 1 min), the cell debris pelleted by centrifugation (10,000 x *g*, 10 min). The supernatant was loaded onto a metal chelating column (Pharmacia) charged with 0.1 mM $NiCl_2$. After extensive

washing the phytocystatins were eluted with 100 mM imidazole and dialysed extensively into HBS buffer (10 mM HEPES pH 7.4, 150 mM NaCl, 0.005% P20) and stored at -70 °C.

Papain assay

The inhibition of papain (Sigma) was assayed using pGlu-Phe-Leu-p-NA (Sigma), a synthetic substrate for cysteine proteinases (Filippova et al. 1984). 100ng of papain, in 50 µl of incubation buffer (0.15 M MES/OH pH 5.8, 4 mM NaEDTA, 4 mM DTT), was added to 50µl of sample, either with no cystatin or containing a known concentration of cystatin and pre-incubated for 30 minutes. 50 µl of 1 mM pGlu-Phe-Leu-p-NA (in incubation buffer and 30% DMSO) was added and papain activity monitored by a linear increase in OD at 415 nm, over 15 min at 25 °C.

Surface Plasmon resonance (SPR) experiments

Papain was obtained from Sigma as a lyophised powder. For the SPR experiments a stock solution of 1 mg/ml was prepared by dissolving in HBS buffer containing 1 mM DTT. This stock was diluted to the required concentrations with the same buffer. All SPR experiments were performed on a BIACORE™ 3000 instrument using an NTA sensorchip. The running buffer was HBS-EP (10 mM HEPES, 0.15 M NaCl, 0.005 % P20, pH 7.4) and all experiments were carried out at 25 °C. Phytocystatins were immobilised (approximately 400 response units (RU) on flowcells 2 to 4 with flowcell 1 left blank. To monitor binding of papain to the phytocystatins, 240 µl of HBS buffer was injected over all flow cells at a flow rate of 80 µl/min (association phase) followed by normal buffer flow for 3 min (dissociation phase) to provide a subtraction blank. This injection was repeated with papain. Proteins were stripped from the sensorchip using EDTA and the surface re-charged with nickel before repeating the experiment with a different concentration of papain. At least 5 cycles were performed for each concentration to allow kinetic analysis. The data from these binding experiments were analysed using the Biaevaluation3 software package (BIACORE) and by global fitting the data to the Langmuir 1:1 binding model.

Thermal stability of phytocystatins

Aliquots of PHYTC57 and OSA-IΔD86 were prepared at 0.5 µg protein/mL in 50 mM phosphate buffer pH 7.4. The samples were immediately placed in a boiling water bath. Samples were removed at various times and plunged into liquid nitrogen, before storage at -70 °C. Residual inhibitory activity, to determine the remaining level of functional cystatin, was determined by the papain inhibition assay.

Digestibility assay

Simulated gastric fluid was made up as described (Astwood et al. 1996) immediately before use and contained 0.32% (w/v) pepsin (Sigma) in 0.03 M NaCl adjusted to pH 1.2 with concentrated HCl. Four replicates of PHYTC57 and of OSA-IΔD86 at 0.2 mg/ml final concentration, were incubated in SGF at 37 °C. At intervals 200 µl aliquots were removed and immediately mixed with 75 µl 0.2M Na_2CO_3 to terminate digestion. Each sample was divided into three sub-samples and, as previously described (Atkinson et al. 2004), were separated on 15% SDS-PAGE gels. Two gels were stained for total protein with Coomassie blue, one to show the effect of simulated gastric fluid on PHYTC57 and the second to compare digestibility of PHYTC57 with OSA-IΔD86. The third gel was subjected to western blot analysis using a mouse antibody to detect the 6His-tag (Qiagen) of PHYTC57 with visualisation using alkaline phosphatase activity. A final sample from each time point was analysed for residual inhibitory activity against papain by the enzyme assay as outlined above.

Results

Synthesis of phytocystatin genes

For the design of the consensus phytocystatin coding region a tBLASTN search of the Genbank database was undertaken using OSA-I (*Oryza sativa*; U54702), ZMA2 (*Zea mays*; D38130) and HAN1 (*Helianthus annuus*; Q10993) protein sequences as search probes. The list of sequences used to derive the consensus sequence is shown in Table 1. Sequences were identified from databases by homology searching. The table shows a systematic name for each cystatin together with the organism name and common name of the plant and the Genebank accession number. Coding sequences were translated and aligned using the program CLUSTALW and the alignment was displayed using the program CINEMA (Parry-Smith et al. 1998; Lord et al. 2002) to allow improvement by manual alignment. A consensus sequence was then derived by identifying the most common amino acid at each position (Figure 1A). The length of the consensus protein was set at 95 amino acids with the N-terminus positioned four residues before the conserved N-terminal glycine residue, and thus before the first β-strand (β1). The C-terminus was set 15 residues after the conserved PW motif and thus after the last β-strand (β5). These criteria were based on the X-ray structures of CEWC (Bode et al. 1988) and human stefin B (Stubbs et al. 1990) and the NMR structure of OSA-I (Nagata et al. 2000).

Table 1. Phytocystatin sequences used to derive the consensus (PHYTC57) sequence.

Phytocystatin	Organism name	Common name	Accession number
AAR	*Ambrosia artemisiifolia*	Short ragweed	L16624
ACE	*Allium cepa*	Onion	AA508918
ATH1	*Arabidopsis thaliana*	Arabadopsis	Z17618
ATH2	*Arabidopsis thaliana*	Arabadopsis	Z97341
ATH3	*Arabidopsis thaliana*	Arabadopsis	Z17675
ATH4	*Arabidopsis thaliana*	Arabadopsis	ATAJ110
ATH6	*Arabidopsis thaliana*	Arabadopsis	AC002409
ATH8	*Arabidopsis thaliana*	Arabadopsis	Z37263
AVU	*Artemisia vulgaris*	Mugwort	AF143677
BCA1	*Brassica campestris*	Chinese cabbage	L41355
BCA2	*Brassica campestris*	Chinese cabbage	L48182
BCA3	*Brassica campestris*	Chinese cabbage	U51119
CPA	*Carcia papaya*	Papaya	X71124
CSA	*Cucumis sativus*	Cucumber	AB014760
CSAT	*Castanea sativa*	Chestnut	AJ224331
CUN	*Citrus unshiu*	Satsuma orange	C95263
DCA	*Daucus carota*	Carrot	D85623
DCAR	*Dianthus caryophyllus*	Carnation	AF064734
GHI2	*Gossypium hirsutum*	Cotton	AI728662
GHI3	*Gossypium hirsutum*	Cotton	AI726250
GMA1	*Glycine max*	Soybean	D64115
GMA2	*Glycine max*	Soybean	U51583
GMA3	*Glycine max*	Soybean	U51855
GMA4	*Glycine max*	Soybean	U51854
GMA5	*Glycine max*	Soybean	AI495568
GMA7	*Glycine max*	Soybean	AI938438
HAN1	*Helianthus annuus*	Sunflower	Q10993
IBA	*Ipomoea batatas*	Sweet potato	AF117334
MCR1	*Mesembryanthemum crystallinum*	Ice plant	AA856241
MCR2	*Mesembryanthemum crystallinum*	Ice plant	AA887617
MDO	*Malus domestica*	Apple tree	AT000283
OSA1	*Oryza sativa*	Rice	U54702
OSA2	*Oryza sativa*	Rice	X57658
OSA5	*Oryza sativa*	Rice	C25431
PAM	*Persea americana*	Avacado	JH0269
PBA	*Populus balsamifera*	Poplar	AI167046
PCO	*Pyrus comunis*	Pear	U82220
PTA	*Pinus taeda*	Pine	AI812403
PTR	*Populus tremula*	Poplar	AI162398
RCO1	*Ricinus communis*	Castor bean	Z49697
RCO2	*Ricinus communis*	Castor bean	T23262

SBI	*Sorghum bicolor*	Sorghum	X87168
SLA	*Silene latifolia*	White campion	Z93053
SLY1	*Lycopersicon esculentum*	Tomato	AF083253
SLY2	*Lycopersicon esculentum*	Tomato	X73986
SLY5	*Lycopersicon esculentum*	Tomato	AI1781497
STU1	*Solanum tuberosum*	Potato	L16450
STU2	*Solanum tuberosum*	Potato	L16450
STU3	*Solanum tuberosum*	Potato	L16450
STU4	*Solanum tuberosum*	Potato	L16450
STU5	*Solanum tuberosum*	Potato	L16450
STU6	*Solanum tuberosum*	Potato	L16450
STU7	*Solanum tuberosum*	Potato	L16450
STU8	*Solanum tuberosum*	Potato	L16450
STU10	*Solanum tuberosum*	Potato	X74985
VUN	*Vigna unguiculata*	Cowpea	Z21954
ZMA1	*Zea mays*	Maize	D10622
ZMA2	*Zea mays*	Maize	D38130
ZMA4	*Zea mays*	Maize	AI001246
ZMA5	*Zea mays*	Maize	AI740162

The synthetic gene encoding PHYTC57 was generated from six oligonucleotides (P1 – P6) each approximately 70 nt, designed to include regions of at least 10 nt overlaps between adjacent oligonucleotides as described in Materials and Methods. For comparative study of naturally occurring phytocystatins we adopted a similar synthetic gene approach using oligonucleotides for the *osa-IΔD86*, *cun* (*Citrus unshiu*; Satsuma) and *cpa* (*Carcia papaya*; Papaya) phytocystatin coding regions designed from the unique sequence of the appropriate protein with codon changes to reflect *E. coli* codon usage. The chicken egg white cystatin (*cewc*) coding region was PCR amplified from a pQE-derived plasmid. The genes were initially cloned into the phage display vector pDHisII by exploiting the *Sfi*I and *Not*I sites (Figure 1B).

Cystatin expression

Cystatins were initially expressed from pDHisII constructs and purified by Ni-NTA affinity chromatography. Analysis of the purified cystatins by electrospray ionisation mass spectrometry indicated that, unexpectedly, in each case there was a C-terminal truncation of the expressed protein. Table 2 shows the expected masses of the full-length cystatins and those with a 16 amino acid C-terminal truncation calculated using the Protein Calculator program (www.scripps.edu/~cdputnam/protcalc.html) together with the determined molecular masses. The sequence of the C-terminal end of the proteins is shown with the major truncation site indicated by an arrow. With the exception of CUN, the determined molecular

mass is in excellent agreement with forms of the fusion proteins that had lost the C-terminal 16 amino acids, but which retain the His tag (Table 2). In the case of CUN there is truncation of fewer than 16 residues, but this was not characterised further. To ensure expression of defined protein sequences the cystatin coding regions plus His-tag were therefore sub-cloned into pET101 by PCR and TOPO-facilitated cloning and expressed in BL21 (DE3) Star cells (Stratagene). Protein expression was induced by addition of IPTG to 1 mM for 16 hours and the cystatins purified under native conditions by Ni-NTA affinity chromatography by virtue of their C-terminal hexahisitdine tag. The cystatins were separated by SDS-PAGE to examine their purity which was estimated to be >98 %, their masses were in excellent agreement with the expected values by ESI-MS (data not shown) and so these proteins were used for further analyses.

Table 2. Electrospray mass spectrometry results for cystatins expressed from HB2151.

Phytocystatin	Predicted mass (Da)		Experimental mass (Da)
	Full length	Truncated	
OSA1ΔD86	14329	12603	12603.3± 0.8
CPA	14321	12596	12593.5± 6.8
PHYTC57	13855	12129	12128.3± 0.8
CEWC	16399	14608	14607.4± 1.5
CUN	14222	12496	12728.0± 1.1
↓ ……..**RGSHHHHHHARAEQKLISEEDLNGAA** (SEQ ID NO 20)			

Cystatin binding activity

Enzyme assays with papain using the artificial substrate pGlu-Phe-Leu-p-nitroanilide confirmed that PHYTC57 was an active cysteine protease inhibitor. *Ki* values were not determined, but from the IC50 values for PHTYC57 and OSA-IΔD86 (4.6×10^{-8} M and 1.8×10^{-7} M respectively) it is clear that PHYTC57 is a more potent inhibitor than OSA-IΔD86. To directly measure the interaction between the cystatins and protease we measured the binding kinetics using BIAcore surface plasmon resonance analysis. The cystatins were immobilised onto nickel coated sensor chips by the C-terminal His-tags. Papain was then allowed to bind to the immobilised cystatin and measurements were made at several papain concentrations. Sensorgrams for the binding of OSA-IΔD86, CPA, CUN, CEWC and PHYTC57 to papain are shown in Figure 3. The data at each concentration were fitted to the Langmuir 1:1 binding model and the kinetic constants were determined. The data showed a good fit to the model consistent with the known 1:1 stoichiometry of cystatin inhibition of

cysteine proteases. A summary of the kinetic constants for these cystatins is shown in Table 3. PHYTC57 displays higher association and lower dissociation kinetics compared with the naturally occurring cystatins tested, with an equilibrium constant K_D of 6.3 x 10^{-12} M, indicating a tight binding complex with papain. This value is two orders of magnitude lower than the K_D value measured for chicken egg white cystatin (3.9 x 10^{-10} M), three orders of magnitude lower than the improved phytocystatin OSA-IΔD86 (4.7 x 10^{-9} M) and four orders of magnitude lower than the phytocystatins CUN (1.4 x 10^{-8} M) and CPA (2 x 10^{-8} M).

Table 3. Kinetic parameters determined by surface plasmon resonance. Association and dissociation rate constants (K_a and K_d) and equilibrium constants (K_A and K_D) are shown.

	K_a (1/M.s)	K_d (1/s)	Chi^2	K_A (1/M)	K_D (M)
OSAIΔD86	2.6 x 10^5	1.2 x 10^3	0.98	2.1 x 10^8	4.7 x 10^{-9}
CUN	2.6 x 10^5	3.5 x 10^3	2.3	7.3 x 10^7	1.4 x 10^{-8}
CPA	2.6 x 10^5	5.2 x 10^3	3.2	5.0 x 10^7	2.0 x 20^{-8}
CEWC	1.3 x 10^6	4.9 x 10^4	5.3	2.6 x 10^9	3.9 x 10^{-10}
PHYTC57	3.5 x 10^5	2.2. x 10^6	13.9	1.6 x 10^{11}	6.3 x 10^{-12}

Phytocystatin stability

We were interested to explore whether PHYTC57 displayed greater stability when compared with a well-characterised parental phytocystain, OSA-IΔD86. Samples of these phytocystatins were incubated for various times in a boiling water bath, chilled and then tested for residual inhibitory activity in the papain assay (Figure 4A). The consensus protein PHYTC57 displays greater thermostability (t½ = 17 min) than OSA-IΔD86 (t½ = 6 min) while inhibitory activity can still be detected at 80 min with PHYTC57 compared with only 58 min for OSA-IΔD86.

We also tested the stability of OSA-IΔD86 and PHYTC57 using a simulated gastric fluid (SGF) digestibility assay. The proteins were incubated for various times in freshly prepared SGF before neutralising the reaction. Samples were then analysed in two ways, by enzyme assays to determine the residual inhibitory activity (Figure 4B) and by SDS-PAGE (Figure 4C) to determine whether the protein was digested and. Enhanced stability of PHYTC57 was observed in the digestion studies with the PHYTC57 and OSA-IΔD86 displaying t½ values of 260 sec and 30 sec respectively in enzyme assays. These assay data indicate that some

99% of OSA-IΔD86 is destroyed in the simulated gastric fluid between 30 and 60 seconds after incubation. For PHYTC57 this level of inactivation does not occur until between 2 and 5 minutes demonstrating that PHYTC57 is more resistant to the digestion conditions. The Coomassie stained SDS-PAGE results (Figure 4C), which identify full-length protein, support these data with only trace OSA-IΔD86 present at 30 sec whereas for PHYTC57 some protein is still present at 120 sec. To confirm these results for PHYTC57 we analysed the SGF digestion products by western blot analysis using an anti-6His tag antibody. As shown in Figure 4D this reveals that the majority of the protein has indeed been destroyed by 300 sec SGF treatment, although, trace amounts of intact PHYTC57 remain at 300 and even 420 sec treatment. If PHYTC57 was to be used for transgenic plant expression, the fact that the majority of full-length PHYTC57 protein and all inhibitory activity has been lost following a 10 min incubation in SGF would mean that PHYTC57 should be readily destroyed during the digestive process following any inadvertent host digestion.

Discussion

The consensus approach to protein design provides a method to generate a sequence that does not exist in nature. Such a sequence should optimise the conserved functional sequence parameters of a family of homologous proteins. In particular critical residues that are involved in the structure and folding of the family are likely to be highly conserved (Lehmann and Wyss 2001; Main et al. 2003a; Main et al. 2003b; Main et al. 2005) (Forrer et al. 2004) (Steipe 2004). Depending upon the biological roles of individual members of the family, functional residues may or may not be conserved. In the case of the phytocystatins which show functional conservation in the form of cysteine protease binding and inhibition, there is a reinforcement of the highly conserved sequence motifs that are involved in interaction with the cysteine protease active site. The consensus approach should therefore give rise to an optimised protein in which the biological function, cysteine protease binding, as well as general stability are enhanced.

Plants contain a class of cystatins that have distinctive characteristics from animal cystatins and stefins, and plants do not contain stefins. In particular for the animal cystatins and stefins the N-terminal Gly sequence, is important for inhibition of cysteine proteases whilst this is not the case for plant cystatins [Abe, K. et al. (1988) J. Biol. Chem 263, 7655-7659]. In addition the plant cystatins contain a typical sequence [LVI][AGT][RKE][FY][AS][VI]X[EDQV][HYFQ]N (SEQ ID NO 81) that is not found in other cystatins or in stefins.

SPR analysis of immobilised cystatins with papain revealed that PHYTC57 is more effective at forming and maintaining a protein:protein interaction complex with papain than are the 3 parental phytocystatins tested here. In particular the dissociation rate constant is reduced indicating that once bound, the PHYTC57:papain complex is more stable than those formed by the other cystatins tested. It has previously been reported that the animal-derived cystatins, which contain disulphide bonds are more efficient cystatins than characterised phytocystatins whose efficacy ranges from around 10 nM for OSA-I (Urwin et al. 1995) to pM for soybean cystatins (Koiwa et al. 2001) against papain. In our studies we observe that PHYTC57 is more effective at binding papain than the animal cystatin chicken egg white cystatin. In addition PHYTC57 displays enhanced thermal stability as well as greater stability in a simulated digestibility assay.

The fact that PHYTC57 displays enhanced properties does not mean that it could not be enhanced further. The reinforcement of the conserved motifs that comprise the binding region may represent the biologically optimal binding sequence, but perhaps not the most effective. For example, Koiwa demonstrated that alteration of the third loop region in soybean cystatins enhanced efficacy of the resulting variants (Koiwa et al. 2001). There has been a report that novel sequences within the QVVAG region confer enhanced inhibitory activity (Melo et al. 2003). In terms of potential use in transgenic plant systems it is unlikely that further enhancements in stability would be beneficial due to the need to ensure that transgenic products do not accumulate in the environment.

We have focussed here on the relative binding efficiency of PHYTC57 compared with OSA-IΔD86. This rice cystatin variant was the first enhanced phytocystatin generated and it has been subjected to a wide range of studies leading to transgenic plant expression and plant nematode resistance trials (Urwin et al. 1995; McPherson et al. 1997; Urwin et al. 1997; Urwin et al. 1998; Urwin et al. 2000; Urwin et al. 2001; Urwin et al. 2003; Lilley et al. 2004). PHYTC57 displays 34 amino acid differences from OSA-IΔD86. A comparison of the positions of these amino acids differences between OSA-IΔD86 and PHYTC57 is shown in Figure 5 in which the substitutions are colour coded and represented upon the structural model for OSA-I (pdb code 1EQK) (Nagata et al. 2000). The differences are dispersed throughout the structural scaffold and include both conservative changes and non-conservative changes. It is interesting that 6 of the changes involve the introduction of negatively charged amino acids. Detailed comparative biochemical, sequence analysis and mutagenesis studies of individual parental cystatin molecules with the consensus protein could be undertaken to define the most critical "consensus" substitutions leading to

enhanced properties, thus enhancing our understanding of protein structure-function relationships.

The surprisingly significant enhanced stability of PHTC57 led us to consider the potential for this small consensus protein as a new scaffold. Protein scaffolds for the selection of new binding functions are proving to be useful in a wide range of applications as replacements for antibodies (Skerra 2007) including in medical applications (Wurch, Pierre et al. 2012). An example of the development of a cystatin as a scaffold for peptide aptamer selection based on human stefin A has been reported (Woodman, Yeh et al. 2005). There has been a recent report of the development of an Fn3-like consensus protein from 15 fibronectin or tenascin Fn3-like domain sequences which is proposed as a potential scaffold (Jacobs, Diem et al. 2012). The naturally occurring Fn3 domain 10 is a well-studied scaffold developed by Koide and colleagues (Koide, Bailey et al. 1998; Karatan, Merguerian et al. 2004; Koide, Gilbreth et al. 2007). Due to its enhanced thermostability, small size and lack of cysteines we anticipate that the consensus PHYTC57 will prove useful as a binding protein scaffold for displaying variable peptide loop libraries for screening against a range of target molecules to identify novel artificial binding proteins.

PHYTC57 therefore offers potential benefits for transgenic plant defence schemes as an improved cysteine protease inhibitor targeted at pathogens such as plant nematodes, and for development as a scaffold protein for selecting new binding functionalities.

CREATION OF SCAFFOLD PROTEIN LIBRARY, SCREENING AND TESTING FUNCTION OF 'ADHIRONS'

Introduction

The present inventors designed a novel artificial binding (scaffold) protein based on the consensus sequence of 57 plant-derived phytocystatins described above (termed PHYTC57 above, but referred to below as 'Adhiron'). This artificial protein meets all the requirements (small, monomeric, high solubility and high stability and the lack of disulphide bonds and glycosylation sites) to be a good scaffold for peptide presentation. We chose the VVAG and the PWE regions of the Adhiron scaffold for peptide presentation with nine randomized positions in each loop.

Based on high yields of Adhiron scaffold expressed in *E. coli* we hypothesised that protein alterations within the two loops have a tolerable effect on the protein expression level and

stability of the scaffold. Therefore our scaffold seems to be amenable for use in generating combinatorial libraries for screening with the phage display technology (Smith 1985). The success of phage displays system relies on the quality of the initial DNA library, which is mainly derived by its diversity. Improved library diversity can be achieved by using trinucleotide (trimer)-synthesized oligos (Kayushin, Korosteleva et al. 1996) which provide theoretically equal levels of introduction of the different amino acids as well as avoidance of stop codons and cysteine (Virnekas, Ge et al. 1994; Krumpe, Schumacher et al. 2007). Furthermore, trimer insertions or deletions will not lead to a shift in reading frame mutation thereby still producing potentially functional proteins. Therefore we have chosen a trimer mixture encoding the 19 naturally occurring amino acids excluding cysteine for the loop-randomised oligos.

Our work demonstrates that Adhirons have a high potential to play a key role in generating research reagents, diagnostics as well as therapeutics (drug discovery).

The Adhiron scaffold shows remarkably high thermal stability (Tm ca. 101 °C) above that reported for any other non-repeat scaffold protein, and can be expressed at high levels in prokaryotic expression systems to produce recombinant protein reproducibly. We have constructed a phage-display library based on the insertion of randomised amino acid sequences to replace residues at two loop regions within the Adhiron. The library has a complexity of approximately 3×10^{10}, with greater than 86% full length clones after phage production, indicating the very high quality of the library. As a demonstration of the efficacy of the library, the yeast Small Ubiquitin-like Modifier protein (SUMO) was screened to identify artificial binding protein (Adhiron) reagents capable of binding to this target protein. More than 20 individual Adhirons were identified that bind to yeast SUMO (ySUMO) as assessed by a phage enzyme-linked immunosorbent assay (ELISA). DNA sequencing indicated that the majority show partial sequence homology within one of the two loop regions to the known SUMO interactive motif (Val/Ile-X-Val/Ile-Val/Ile; where X is any amino acid). Four Adhiron coding regions were sub-cloned into the vector pET11 and recombinant protein was expressed, purified and tested in ELISA and Western blot analyses. The four Adhirons had low nanomolar affinities for ySUMO and showed high specificity to ySUMO with low level binding to human SUMO1 protein. Furthermore, we screened the Adhiron library against a number of other targets, namely fibroblast growth factor (FGF1), platelet endothelial cell adhesion molecule (PECAM-1), also known as cluster of differentiation 31 (CD31), and a 10 amino acid peptide with a cysteine on the N-terminus for thiol linkage to biotin (Cys-Thr-His-Asp-Leu-Tyr-Met-Ile-Met-Arg-Glu) and also identified Adhirons against these targets as confirmed by phage ELISA. We have, therefore, developed a versatile, highly stable and well

expressed scaffold protein, termed Adhiron, that is capable of displaying randomised peptide loops and we have demonstrated the ability to select highly specific, high affinity binding reagents from an Adhiron library against a range of targets for use in multiple applications.

Materials and Methods

Construction of Adhiron library

A consensus sequence derived from alignment of 57 phyocystatin sequences was identified as described above and a codon-modified gene designed for expression in *E. coli* was synthesised (GenScript). The Adhiron scaffold coding region and Adhiron library coding regions were cloned between NheI and NotI restriction sites to create a fusion coding region with the 3' half of the gene III of bacteriophage M13 in a phagemid vector pBSTG1, a derivative of pDHisII which is derived from pHEN1 (Hoogenboom, Griffiths et al. 1991) and which also contains a DsbA signal peptide (pBSTG1- DsbA-Adhiron). The library was constructed by splice overlap extension (SOE) of two PCR products (Horton, Cai et al. 1990) and all primers were synthesised by Ella Biotech.

The first PCR product extended from the DsbA coding sequence to the first inserted loop and was generated by the primers:
Forward primer 5' – TCTGGCGTTTTCTGCGTC – 3' (SEQ ID NO 21),
Reverse primer 5' – CTGTTCTTTCGCTTTAACAAC – 3' (SEQ ID NO 22).

The second PCR product introduced two nine amino acid loop regions into the scaffold protein at loop 1 and loop 2 by using the following primers. The PstI site used for cloning is underscored:
Forward loop
5'GTTGTTAAAGCGAAAGAACAGNNNNNNNNNNNNNNNNNNNNNNNNNNNACCATGTAC
CACTTGACCCTG – 3' (SEQ ID NO 23),
Reverse loop
5'CTGCGGAACTCCTGCAGTTCTTTGAAGTTNNNNNNNNNNNNNNNNNNNNNNNNNNC
TTAACCCAAACTTTCGCTTCG – 3' (SEQ ID NO 24).

The degenerate positions (NNN) were introduced as trimers representing a single codon for each of the 19 amino acids excluding cysteine and there were no termination codons. The primers were also designed to introduce NheI (forward) and PstI (reverse) restriction sites to facilitate cloning into the pBSTG1 phagemid vector that contains an in-frame amber stop codon to allow translational read through to create an Adhiron-truncated pIII fusion protein.

PCR was performed using Phusion™ High Fidelity Polymerase (NEB) at 98 °C for 5 minutes followed by 20 cycles of 98 °C, 10 sec; 56 °C, 15 sec; 72 °C, 15 sec followed by 72 °C for 5 minutes. PCR products were purified by gel extraction (Qiagen), and used for SOEing with 10 cycles using the protocol above. The PCR product was digested with NheI and PstI and was gel extracted then cloned into the pBSTG1- Adhiron phagemid that had also been digested with NheI and PstI to leave the DsbA signal sequence and C-terminal coding region of the Adhiron, to generate the DNA based Adhiron library. Electroporation was used to introduce the ligated library products into *E. coli* ER2738 electrocompetent cells (Lucigen). In total 20 mL of ER2738 cells were electroporated with 50ng of library DNA per 50ul of ER2738 cells. Cells were allowed to recover for 1 hr in 2TY medium and were then grown at 37 °C, 225 rpm to an OD_{600} of 0.6 in 2 litres of 2TY medium. 1 µl M13KO7 helper phage (NEB) (10^{14} /ml) were added and allowed to infect the cells with shaking at 90 rpm for 1 hr, and then the culture was allowed to produce phage particles overnight at 25 °C in the presence of kanamycin (50 µg/ml). The phage were precipitated with 6% polyethylene glycol 8000 and 0.3 M NaCl, and suspended in 50% glycerol for storage. Library size was determined to be ~$3x10^{10}$ with a minimal vector only background.

Target Preparation and phage display

The following protocols are described for yeast SUMO but an identical protocol was used for the screening or other targets. Yeast SUMO (ySUMO) protein was expressed in BL21 (DE3) cells using IPTG induction and purified by Ni-NTA resin (Qiagen) affinity chromatography according to the manufacturer's instructions. Purity was confirmed by SDS-PAGE. Yeast SUMO was biotinylated using EZ-link™ NHS-SS-biotin (Pierce), according to the manufacturer's instructions. Biotinylation was confirmed using streptavidin conjugated to horse radish peroxidase (HRP) to detect the biotin on ySUMO absorbed onto Immuno 96 Microwell™ Nunc MaxiSorp™ (Nunc) plates. Phage display library biopanning was performed as follows:

5µl of the phagemid library, containing 10^{12} phagemid particles, was mixed with 95µl Phosphate buffer saline, 0.1% Tween™-20 (PBST) and pre-panned three times in high binding capacity streptavidin coated wells (Pierce) for a total of 1 hour. 100µl of 100nM biotinylated ySUMO was added to the panning well for 1 hour with shaking on a Heidolph VIBRAMAX 100 at speed setting 3 prior to adding the pre-panned phage for 2.5 hours also on the vibrating platform. Panning wells were washed 10 times in 300 µl PBST using a plate washer (Tecan Hydroflex), and eluted with 100 µl of 50 mM glycine-HCl (pH 2.2) for 10 minutes, neutralised with 1 M Tris-HCL (pH 9.1), and further eluted with 100 µl of

triethlyamine 100 mM for 6 minutes and neutralised with 50 µl of 1M Tris-HCl (pH 7). Eluted phage were incubated with exponentially growing ER2738 cells (OD_{600} = 0.6) for 1 hr at 37°C and 90 rpm. Cells were plated onto Lysogeny Broth agar plates supplemented with 100 µg/ml carbenicillin and grown at room temperature overnight. The next day, the colonies were scraped into 5 ml of 2TY medium and inoculated into 25 ml of 2TY medium supplemented with carbenicillin (100 µg/ml) to reach an OD_{600} of 0.2, incubated at 37 °C, 225 rpm for 1 hr and infected with ca.1 x 10^9 M13K07 helper phage. After 1 hr incubation at 90 rpm, kanamycin was added to 25 µg/ml, cells were incubated overnight at 25 °C at 170 rpm, and phage were precipitated with 6 % polyethylene glycol 8000, 0.3M NaCl and resuspended in 1 ml of 10 mM Tris, pH 8.0, 1 mM EDTA (TE buffer). 2 µl of this phage suspension was used for the second round of selection. This time phage display was performed using streptavidin magnetic beads (Invitrogen). Phage were pre-panned with 10 µl of washed beads for 1 hr on a Stuart SB2 fixed speed rotator (20 rpm), and 10 µl of beads were labelled with 100µl of 100nM biotinylated ySUMO for four hours on the same rotator. Yeast SUMO labelled beads were washed three times in PBST prior to adding the pre-panned phage for 1 hr. Beads were washed 5 times in PBST using a magnet to separate beads from solution after each washe, then eluted and amplified as above. The final pan was performed using neutravidin high binding capacity plates (Pierce), as previously described for the first panning round, but this time the phage were eluted using 100 µl 100 mM dithiothreitol (DTT) on a vibrating platform for 20 min prior to infection of ER2738 cells. Phage were recovered from wells containing target protein and controls wells to determine the level of amplification in target wells.

Phage ELISA

Individual ER2738 colonies from the final pan were picked and grown in 100 µl of 2TY with 100 µg/ml of carbenicillin in a 96 deep well plate at 37 °C (900 rpm) for 6 hr. 25 µl of the culture was added to 200 µl of 2TY containing carbenicillin and grown at 37°C (900 rpm) for 1 hr. M13K07 helper phage (10 µl of 10^{11}/ml) were added, followed by kanamycin to 25 µg/ml and the bacteria were grown overnight at 25 °C (450 rpm). Streptavidin coated plates (Pierce) were blocked with 2 x casein blocking buffer (Sigma) overnight at 37 °C. The following day the plates were labelled with 0.4 nM of biotinylated yeast SUMO for 1 hr, the bacteria were collected by centrifugation at 3000 rpm for 5 min and 45 µl of growth medium containing the phage was added to wells containing biotinylated yeast SUMO or a well containing the biotinylated linker and incubated for 1 hr. Wells were washed using a Tecan HydroFlex™ plate washer 3 times in 300 µl PBST, and a 1:1000 dilution of HRP-conjugated anti-phage antibody (Seramun) in 100 µl PBST was added for 1hr. Wells were washed 10

times in 300 µl PBST and binding was visualised with 100 µl 3,3',5,5'-Tetramethylbenzidine (TMB) liquid substrate (Seramun) and measured at 560 nm.

Adhiron protein production

The DNA coding sequences of Adhirons that bound to yeast SUMO were amplified by PCR, the product was restriction digested with NheI and PstI and cloned into pET11a containing the Adhiron scaffold and digested with the same restriction sites. Colonies were picked and grown overnight in 5 ml LB medium at 37 °C, 225 rpm and plasmid DNA was purified as minipreps (Qiagen) and sequenced to confirm the presence of the correct insert. Plasmids were transformed into BL21 (DE3) cells by heat shock and colonies were grown overnight at 37 °C. The following day the culture was added to 400 ml of LB medium, grown to an OD_{600} of 0.6 at 250 rpm at 37 °C and isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to 1 mM final concentration. Cells were grown for a further 6 hr, harvested by centrifugation at 3000g and re-suspended in 25 ml of 1x Bugbuster (Novagen). Benzonase was added according to the manufacturer's instructions and the suspension was mixed at room temperature for 20 minutes, heated to 50 °C for 20 minutes and centrifuged for 20 minutes at 9400 xg. The cleared supernatant was mixed with Ni-NTA resin 500 µl of slurry for 1hr, washed 3 times in 30 ml wash buffer (50mM PBS, 500 mM NaCl, 20 mM imidazole, pH 7.4) and eluted in 1 ml of elution buffer (50mM PBS, 500 mM NaCl, 300 mM imidazole, pH 7.4). 100 µg of the SUMO binding Adhirons (Ad-ySUMO) were biotinylated using NHS SS-biotin (Pierce) according the manufacturer's instructions for use in ELISAs and Western blotting.

ELISA analysis

5 ng (unless otherwise indicated) of target protein in PBS was absorbed on to Immuno 96 Microwell™ Nunc MaxiSorp™ plate wells overnight at 4 °C. The next day 200 µl of 3 x blocking buffer was added to the wells and incubated at 37 °C for 4 hours with no shaking. Biotinylated yeast SUMO binding Adhirons at 100 µg/ml were diluted 1:1000 in PBST containing 2 x blocking buffer and 50 µl aliquots were incubated in target wells for 1hr with shaking. Wells were washed 3 x in 300 µl PBST, and streptavidin conjugated to horse radish peroxidase (HRP) (Invitrogen) diluted 1:1000 in 50 µl PBST was added to the wells for 1 hr. Wells were washed 6 x in 300 µl PBST and binding was visualised with 50 µl TMB liquid substrate and the absorbance measured at 560 nm.

Western Blot analysis

Target protein or target protein mixed with HEK293 cell lysate (20 µg) was mixed with loading buffer (Laemmli, 60 mM Tris-Cl pH 6.8, 2% SDS, 10% glycerol, 5% β-

mercaptoethanol, 0.01% bromophenol blue), boiled for 3 min, centrifuged for 1 min at 15,000 xg and then resolved in a 15% SDS-polyacrylamide gel. Proteins were transferred to PVDF membranes for 45 minutes at 4 Watts (Amersham Biosciences) and incubated for 1 hr in blocking buffer (5% BSA in PBS 0.1% Tween) followed by incubation for 1 hr with Ad-ySUMO (100 µg/ml diluted 1:1000 PBST). Bound Ad-ySUMOs were detected using streptavidin conjugated HRP and chemiluminescence (ECL Plus kit, Amersham).

Protein-protein interaction affinity measurement

The BLitz™ (ForteBio) dip and read streptavidin biosensors were used to estimate affinity of binding of the biotinylated Ad-ySUMO binders, according to the manufacturer's instructions. In brief, at least 4 readings at different ySUMO concentrations (0.25 mM – 1 mM), were used to measure the affinity of each Ad-ySUMO. A global fit was used to calculate the affinity of each Ad-ySUMO. These readings were comparable to affinities measures made using a Biacore surface plasmon resonance instrument.

Results

Adhiron design and phage display

The Adhiron gene was originally designed to create a more potent protease inhibitor, however, due to its potential as a scaffold protein for presenting constrained peptide regions for molecular recognition (Figure 6A) we decided to investigate its use as such a scaffold. The gene sequence was codon optimised to enhance expression in an *E. coli* expression system (Figure 6B). Restriction sites were introduced to facilitate cloning of the gene into the pBSTG1 phagemid vector to allow in-frame translational read through of an amber stop codon to allow an Adhiron-truncated pIII fusion protein to be produced when expressed in non-suppressor cells such as ER2738 cells but to allow production of the Adhiron only in suppressor cells such as JM83. The Adhiron pIII fusion protein was expressed from the phagemid vector pBSTG1, while the other components to allow replication and packaging of the phagemid DNA into M13 phage particles were introduced using M13KO7 helper phage. Expression of the Adhiron-pIII fusion protein was confirmed by Western blot analysis using an anti-pIII antibody. The thermal stability of the Adhiron scaffold was tested by differential scanning calorimetry, which showed a melting temperature of 101 °C (Figure 7A). The structural integrity of the consensus sequence was examined using circular dichroism, which demonstrated a high ratio of beta sheet to alpha helix and random coil (Figure 7B).

We then compared the thermal stability by differential scanning calorimetry of the Adhiron scaffold (Figure 8A) with a representative small soluble well characterised protein, lysozyme

(Figure 8B) which shows that lysozyme is significantly less stable (Tm ca. 65 °C) than the Adhiron. We then tested an Adhiron selected to bind to a myc antibody, the addition of the loops into the scaffold reduces the Tm to 85 °C but this still represents a higher melting temperature than most scaffold proteins. This Adhiron protein can undergo repeated cycles of denaturation and renaturation as shown by the series of scans (Figure 8C).

Library Design

The introduction of peptide encoding sequences suitable for molecular recognition was guided by the predicted loop positions within the structure of the Adhiron (Figure 6). Loop1 was positioned between the first and second beta strands and loop2 was positioned between the third and fourth beta strands. Sequences comprising nine random amino acids (excluding cysteine) were introduced at both loop positions replacing four and three amino acids in loop1 and loop2, respectively. To determine if extension of these loop regions disrupts the structure of the Adhiron three individual Adhirons with loop insertions were isolated, expressed and examined by circular dichroism (Figure 7B). All three clones maintained a high proportion of beta structure, with one clone displaying an increase in beta structure content likely indicating extension of the beta strands into the new loop regions. This demonstrates that loop insertion does not affect the structure of the scaffold. We generated a phage display library of complexity approximately 3×10^{10}. To check the amino acid composition, 96 clones were isolated from ER2738 cells infected with the library. We examined the sequence of phage clones to determine any bias in amino acid composition or other undesirable consequences introduced during phage production (Figure 7C). No bias in amino acid distribution was observed. 86.5 % clones were full-length variants while 3.1 % of clones were the Adhiron scaffold with no inserts, and 10.4 % of the clones showed frame shifts and so were likely of no value in the library. This very high proportion of full-length coding sequences at the level of the phage genome demonstrates the high quality of the library generated.

Library Screening

Library screening was performed initially using yeast SUMO as the target. Yeast SUMO was biotinylated to allow immobilisation of the protein via avidin binding proteins and to ensure that the target was not adsorbed directly onto plastic or particle surfaces which can sometimes lead to denaturation of the target protein. This ensures that the target protein maintains its three dimensional structure allowing for the selection of binding proteins that recognise either linear, or conformational epitopes. Over 1000-fold amplification in colony recovery was observed compared to control samples by panning round three. Twenty four clones were isolated and their ability to bind to the SUMO target was confirmed by phage

ELISA (Figure 9). All clones tested showed strong binding to yeast SUMO with little or no binding to the control wells demonstrating the specificity of the Adhirons. The clones were sequenced, which identified 22 distinct Adhirons and the sequences of the loop regions in these clones is shown in Table 4.

Table 4. Showing the two loop sequences for the 24 Ad-ySUMO binders identified from the screen.

Ad-ySUMO	Loop1 (SEQ ID No)	Loop2 (SEQ ID No)
1	WDLTGNVDT (25)	WDDWGERFW(49)
2	IDLTNSFAS(26)	DINQYWHSM(50)
3	INLMMVSPM(27)	GIQQNPSHA(51)
4	IDLTHSLNY(28)	GLTNEIQKM(52)
5	IDLTHSLNY(29)	GLTNEIQKM(53)
6	IDLTEWQDR(30)	PEPIHSHHS(54)
7	WVDMDYYWR(31)	MDEIWAEYA(55)
8	IDLTQTEIV(32)	EPGIIPIVH(56)
9	IDLTDVWID(33)	GLMTQTNSM(57)
10	IIIHENDAD(34)	GIMDGLNKY(58)
11	WILNNTQFI(35)	VLEGPDRWTV(59)
12	WYERSENWD(36)	RDYGFTLVP(60)
13	WDLTTPINI(37)	YEDYQTPMY(61)
14	WFDDEYDWI(38)	DYAATDLYW(62)
15	IDLTQPHDS(39)	YEEDEYWRM(63)
16	IDLTQSFDM(40)	PIDSNFTGT(64)
17	WYLLDVMDD(41)	HDRRYKQAE(65)
18	WIDRGQYWD(42)	IHNGYTIMD(66)
19	WSEADNDWH(43)	LDLETWQHF(67)
20	IDLTGQWLF(44)	PLWQYDAQY(68)
21	IDLTQSFDM(45)	PSHHNYQTM(69)
22	IDLTQSFDM(46)	PIDSNFTGT(70)
23	IDLTQPHDS(47)	PHDELNWNM(71)
24	WEDFQTHWE(48)	DVGQLLSGI(72)

Clones 4 and 5 are identical and clones 16 and 22 are also identical. Interestingly clones 15 and 23, as well as 21 and 22 contain the same amino acid sequence in loop1 but different sequences in loop2. This sequence variation further supports the complex nature of the library. Analysis of the sequences identified a commonly occurring sequence of IDLT in positions 1 to 4 of loop1 in 12 of the clones indicating that this may be an important motif in binding to at least one epitope on the ySUMO. Also either a P or G in position occurs at position 1 of the second loop in 9 distinct clones and a P or G occurs in a position within residues 2 to 5 in another 6 clones potentially indicating that some structural feature may be important in binding. Interestingly the IDLT motif is similar to the human SUMO1 binding site of the MEF2 E3 ligase PIASx (VDVIDLT – SEQ ID NO 73) (Song, Durrin et al. 2004; Song, Zhang et al. 2005). Four clones were selected for further characterisation; clones 15 and 22

as this loop1 sequence occurred more than once, clone 20 as it also contained the IDLT motif and clone 10 as it contained a distinct motif in loop1.

Characterisation of the Adhiron-ySUMO (Ad-YSUMO) proteins

Due to the high thermal stability of the Adhiron scaffold (Figure 7A) we predicted that to aid purification it should be possible to heat denature and precipitate the majority if *E. coli* proteins without affecting the integrity of the expressed Adhirons. To test this we heated lysates for 20 minutes at 50 °C, 60 °C, 70 °C, 80 °C, 90 °C and 100 °C centrifuged to pellet the denatured protein and analysed the supernatants by SDS-PAGE (Figure 10A). Heating the lysate dramatically decreased the quantity of bacterial protein in the supernatant but did not significantly reduce Adhiron levels. A temperature of 50 °C was suitable to precipitate the majority of bacterial proteins and so was adopted in future studies. Figure 9B demonstrates that the purified Ad-ySUMOs show high purity using a batch metal affinity purification method and that in some samples such as clone 10, the majority, of the protein was not isolated, potentially due to the limiting amount of the resin used during this purification. The estimated level of protein expressed was approximately 100 mg/L. The affinities of the Adhirons were estimated by using BioLayer Interferometry with BLitz™ (ForteBio) dip and read biosensors. The affinities were 11.5, 2.4, 14.2, and 9.0 nM for Ad-ySUMO10, 15, 20 and 22, respectively. These values are in line with affinities normally seen for good antibodies.

To further evaluate the use of the Adhirons as research reagents the Ad-ySUMOs were biotinylated and used in ELISA (Figure 10C) and Western blot analysis (Figure 10D). The Ad-ySUMOs bind to yeast SUMO but not to human SUMO1 (data not shown for SUMO1 Western blots) (n=3). To determine the specificity of the reagents, yeast SUMO was mixed with HEK293 cell lysates. Interestingly, Ad-ySUMO10 and 15 show specific binding to yeast SUMO with no binding to other proteins but Ad-ySUMO20 and 22 bind to many proteins in the lysates by Western blotting (n=3).

Further example screens

To further evaluate the ability of the Adhiron library to identify specific reagents capable of binding to a range of targets we screened against a growth factor (FGF1), a receptor (CD31), and a peptide sequence. All screens were performed over three panning rounds. Phage ELISA was used to examine the ability of Adhirons to bind to the corresponding target (Figure 11). Interestingly, the majority of the clones tested for FGF1 and CD31 showed specific binding, whereas only three clones from the peptide screen showed specific binding. Further panning rounds against the peptide increased the ratio of hits to background so that

80 % of the clones picked showed binding to target. This result is not unexpected due to the small size and limited likelihood of appropriate epitope presentation of the peptide compared to the larger and therefore potentially multiple epitope sites of the proteins.

To confirm that expressed Adhirons bind to their targets we have used the Blitz™ to analyse three distinct recombinant Adhirons for both CD31 and the peptide target. The Adhirons were expressed and purified as soluble proteins. The K_D values for CD31 Adhirons ranged from 8.5×10^{-8} to 6.8×10^{-9} M while those for the peptide ranged from 3.3×10^{-8} to 3.5×10^{-8}.

Additionally, as shown in Figure 25 we have identified Adhirons that bind to an organic molecule, posaconazole,

Adhiron crystal structure

Figure 17 shows the crystal structure of an Adhiron complexed with a FcgRIIIa receptor domain. This reveals the 3D structure of the Adhiron showing the key structural elements of 4 beta strands and an alpha helix. The structure is unexpected in terms of a more compact nature and an apparently less twisted structure than is seen for X-ray structures of other cystatins, including for example stefin A. It is interesting that the beta structure extends into the loop regions to varying degrees.

The importance of displaying two randomised loops for selection of binding molecules for at least some targets is highlighted in this structure by the intimate interaction of both loops with the receptor protein. These loops correspond to LOOP1 and LOOP2 described in more detail below.

Discussion

We have developed a new scaffold protein based on a consensus design of plant cystatin proteins, termed Adhiron, and which displays an extremely high thermal stability with a Tm of 101 °C. This scaffold has been used to produce libraries by the introduction of two 9 amino acid variable regions. These variable sequences were encoded by oligonucleotides in which the variable positions are a subset of trimers comprising a single codon for each amino acid with the exception of cysteine. This resulted in a very similar distribution of each amino acid within the library. The library was of very high quality with 86.5% of clones representing full length variant clones.

The library was configured in a filamentous phage display format as a truncated gp3 fusion and has been screened against various target proteins. Analysis of Adhirons identified by screening against yeast SUMO revealed a number of proteins with distinct sequences in their variable regions. In some cases there are similarities which implies binding to the same site on the SUMO, whereas other clones do not show sequence conservation. All clones bind to ySUMO and not to a range of control proteins, indicating specificity. We have also identified Adhirons that bind specifically to a growth factor, human FGF1, a human receptor protein domain from CD31, a peptide and an organic compound. The ability to select binders against organic compounds as well as wide range of proteins is an important finding as most scaffolds have structural features that favour particular classes of target molecules.

The Adhirons can be conveniently purified by the inclusion of a temperature step at 50 °C which denatures many of the endogenous *E. coli* host proteins, thus enhancing the efficiency of affinity purification of the Adhiron. The X-ray structure of a complex of an Adhiron that binds to the human FcgRIII receptor provides useful information not only on the complex but also on the Adhiron protein and reveals a more compact and largely beta structure that supports the CD data. The compact nature of the scaffold, which is more pronounced than seen in other structures of stefins or cystatins seems likely to contribute to its high thermal stability. The interaction interface revealed by the X-ray structure indicates that both loop regions are involved in the interaction with the receptor domain.

The demonstration of successful and high quality libraries based on this highly stable, small and easily purified scaffold coupled with our robust and effective strategy for screening against a range of proteins makes our Adhiron library a valuable and novel resource for the development of reagents for a wide range of scientific, medical and commercial applications.

VARIANTS OF ADHIRONS

Variants of the Adhiron scaffold have been produced. The sequence shown in SEQ ID No1 and Figure 6 is for the shortest version Adhiron examined, and comprises 81 residues in length before addition of further functional sequences such as a linker and His-tag or other sequences. However, two longer scaffold proteins have also been produced (SEQ ID No 2 and 3), and each may be preferable in certain circumstances. It is possible that further deletion from the scaffold may be possible, but it is believed that SEQ ID No1, or variants thereof, are near to the optimum minimal length without stability of the scaffold protein being unduly compromised.

Full length 'Adhiron 92' (92 aa) has the following sequence (which consists of the scaffold sequence SEQ ID No 3 plus a linker and His-tag, which are underlined):
ATGVRAVPGN ENSLEIEELA RFAVDEHNKK ENALLEFVRV VKAKEQVVAG TMYYLTLEAK DGGKKKLYEA KVWVKPWENF KELQEFKPVG DA AAAHHHHHH (SEQ ID NO 74)

Short 'Adhiron 84' (84 aa) has the following sequence (which consists of the scaffold sequence SEQ ID No 2 plus a linker and His-tag, which are underlined):
GNENSLEIEE LARFAVDEHN KKENALLEFV RVVKAKEQVV AGTMYYLTLE AKDGGKKKLY EAKVWVKPWE NFKELQEFKP VGDA AAAHHHHHH (SEQ ID NO 75)

Shortest 'Adhiron 81' (81 aa), which is shown in Figure 6, has the following sequence (which consists of the scaffold sequence SEQ ID No 1 plus a linker and His-tag, which are underlined):
NSLEIEELAR FAVDEHNKKE NALLEFVRVV KAKEQVVAGT MYYLTLEAKD GGKKKLYEAK VWVKPWENFK ELQEFKPVGD A AAAHHHHHH (SEQ ID NO 76)

The underlined sequence comprises an additional 3 Ala linker and 6 His detection/purification tag. This tag is not part of the scaffold *per se*, but is a useful addition to the protein for obvious reasons.

SPECIFIC EXAMPLES OF ADHIRONS FOR LIBRARIES

Exemplary Adhiron sequences which are useful for preparing scaffold protein libraries, i.e. libraries in which a variety of peptides have been inserted into the scaffold, are as follows:

Adhiron 92 (sequence shown includes and additional Met, linker and tag)
M***ATG***VRAVPGNENSLEIEELARFAVDEHNKKENALLEFVRVVKAKEQ***VVAG***TMYYLTLEAK DGGKKKLYEAKVWVK***PWE***NFKELQEFKPVGDA AAAHHHHHH (SEQ ID NO 77)

One or more of the following modifications can be/have been made:

- An additional methionine residue (in bold) has been added at the N-terminus to facilitate translation.
- An N-terminal portion is located at amino acid residues 2-4 (in bold and italics), and suitably these 3 amino acids are replaced by an insert of typically from 3 up to about 20 amino acids.

- LOOP1 is located at amino acid residues 47-50 (numbered to exclude the N-terminal met) (in bold and italics), and suitably these 4 amino acids can be replaced by an insert of typically from 4 up to about 20 amino acids. A loop length of from 5 to 13 amino acids is preferred, and it is believed that a loop length of 9 amino acids is optimal.
- LOOP2 is located at amino acid residues 76-78 (numbered to exclude the N-terminal met) (in bold and italics), and suitably these 3 amino acids can be replaced by an insert of an insert of typically from 3 up to about 20 amino acids. A loop length of from 5 to 13 amino acids is preferred, and it is believed that a loop length of 9 amino acids is optimal.
- A C-terminal linker and His-tag is present. The length and composition of the linker can be varied, and the tag could of course be adapted to any suitable purification system.

Adhiron 84 (sequence shown includes and additional Met, linker and tag)

MGNENSLEIEELARFAVDEHNKKENALLEFVRVVKAKEQ***VVAG***TMYYLTLEAKDGGKKKLY EAKVWVK***PWE***NFKELQEFKPVGDA AAAHHHHHH (SEQ ID NO 77)

One or more of the following modifications can/have be made:

- An additional methionine residue (in bold) has been added at the N-terminus to facilitate translation.
- An N-terminal peptide sequence can be added to the N-terminus of the Adhiron (i.e. between the methionine residue and the first glycine as shown above), and this addition can be typically from 3 up to about 20 amino acids.
- LOOP1 is located at amino acid residues 39-42 (numbered to exclude the N-terminal met) (shown in bold and italics), and suitably these 4 amino acids can be replaced by an insert of typically from 4 up to about 20 amino acids. A loop length of from 5 to 13 amino acids is preferred, and it is believed that a loop length of 9 amino acids is optimal.
- LOOP2 is located at amino acid residues 68-70 (numbered to exclude the N-terminal met) (shown in bold and italics), and suitably these 3 amino acids can be replaced by an insert of typically from 3 up to about 20 amino acids. A loop length of from 5 to 13 amino acids is preferred, and it is believed that a loop length of 9 amino acids is optimal.
- A C-terminal linker and His-tag is present. The length and composition of the linker can be varied, and the tag could of course be adapted to any suitable purification system.

Adhiron 81 (excludes the Met)

MNSLEIEELARFAVDEHNKKENALLEFVRVVKAKEQ***VVAG***TMYYLTLEAKDGGKKKLYEAK VWVK***PWE***NFKELQEFKPVGDA AAAHHHHHH (SEQ ID NO 78)

One or more of the following modifications can/have be made:

- An additional methionine residue (in bold) has been added at the N-terminus to facilitate translation.
- An N-terminal loop can be added to the N-terminus of the Adhiron, and this addition can be typically from 3 up to about 20 amino acids.
- LOOP1 is located at amino acid residues 36-39 (numbered to exclude the N-terminal met) (shown in bold and italics), and suitably these 4 amino acids can be replaced by an insert of typically from 4 up to about 20 amino acids. A loop length of from 5 to 13 amino acids is preferred, and it is believed that a loop length of 9 amino acids is optimal.
- LOOP2 is located at amino acid residues 65-67 (numbered to exclude the N-terminal met) (shown in bold and italics), and suitably these 3 amino acids can be replaced by an insert of typically from 3 up to about 20 amino acids. A loop length of from 5 to 13 amino acids is preferred, and it is believed that a loop length of 9 amino acids is optimal.
- A C-terminal linker and His-tag is present. The length and composition of the linker can be varied, and the tag could of course be adapted to any suitable purification system.

Thus, taking Adhiron 92 as an example, a particularly suitable scaffold protein for use in a display system may take the form:

N-TERM PEPTIDE LOOP 1

M**XXXXXX**VRAVPGNENSLEIEELARFAVDEHNKKENALLEFVRVVKAKEQ**XXXXXXXXX**TM

LOOP2 linker and tag

YYLTLEAKDGGKKKLYEAKVWVK**XXXXXXXXX**NFKELQEFKPVGDA AAAHHHHHH

where X is any amino acid. (SEQ ID NO 79)

In general LOOP1 and LOOP2 are believed to be of primary importance in target binding, as supported by the crystal structures (Figure 17). In certain embodiments of the invention only one of LOOP1 and LOOP2 can be replaced with a peptide sequence, but in general it is preferred that both are replaced. Whilst the N-TERM is not envisaged as being quite as important as LOOPs 1 and 2, in many circumstances inserting a suitable peptide at the N-TERM may result in improved binding affinity and specificity.

Figures 20 to 24 show sequence alignments of the LOOP1 and LOOP2 regions of several Adhirons which bind to LOX1, HGH, yeast SUMO, PBP2 and a peptide respectively.

Additionally, Figures 27 and 28 show LOOP1 and LOOP2 regions of several Adhirons which bind to Grb2 and STAT3.

It should be noted that peptides can optionally be inserted into the loop regions without removal of the existing amino acids, but this is typically less preferred.

ADDITIONAL EXAMPLES OF THE WIDE UTILITIES OF ADHIRONS

Immunofluorescence:

Example – Reagents to detect viral proteins

Despite repeated attempts over many years it has not proven possible to raise an antibody that identifies the Human Papilloma Virus E5 protein. (Quantitative measurement of human papillomavirus type 16 E5 oncoprotein levels in epithelial cell lines by mass spectrometry. Sahab et al., J Virol. 2012 Sep;86(17):9465-73; Wetherill, LF, Ross, R and Macdonald, A (2012). HPV E5: An enigmatic oncoprotein. Small DNA Tumour Viruses. Ed. K. Gaston. Caister Academic Press. pg55-70)

The utility of the Adhiron system is thus demonstrated by the example that Adhirons were raised against HPV16 E5 viral protein using as a target a peptide with identity to a region of the E5 protein. The E5 Adhiron was biotinylated and used in immunofluorescence to detect E5 protein in overexpressed cells. The Adhiron does not cross react with other HPV serotypes. In addition the E5 Adhirons have been conjugated to Quantum Dots and used to detect E5 protein in human samples. Conjugation to Quantum Dots increased the sensitivity of the reagents. See Figure 12. HPV16 E5 GFP (target) and HPV16 E5 GFP without epitope for Adhiron (control) were expressed in mammalian cells. Adhiron E5 was conjugated to quantum dots and used to detect E5 protein in the mammalian cells. Cells were stained with DAPI (DNA stain), GFP, and E5 (with the Adhiron-Quantum dots). The Adhiron only binds to the target showing specificity to the E5 protein.

Inhibiting and modifying protein - protein interactions:

Example 1. Reagents that inhibit binding to SUMO.

There have been no antibodies raised that are able to specifically and differentially bind to human SUMO 2 (hSUMO2). Adhirons were raised against hSUMO2 and multiple Adhirons that specifically bind to SUMO2 rather than human SUMO1 were identified. Figure 13

demonstrates that the hSUMO2 Adhirons have a functional effect on a protein-protein interaction by binding to hSUMO2 and preventing RNF4, a polySUMO specific E3 ubiquitin ligase from binding with hSUMO2. The hSUMO2 Adhiron has this effect without affecting ubiquitination of other proteins. In the presence of ATP hSUMO2 normally binds to RNF4 causing ubiquitination of the target proteins (black smear at the top of the gel in lane 2). In the presence of increasing concentrations of the Adhirons (lanes 3 to 9) the level of ubiquitination decreases.

Example 2. Reagents that alter fibrin clot and lysis.

Fibrinogen was screened to identify Adhirons that could alter clot formation and lysis. Numerous Adhirons have been identified that alter this process in plasma samples. The graph shown in Figure 14 represents the clot formation and lysis turbidity assay. The black line represents the normal time course of clot formation and lysis. The grey lines represent the effects of five different Adhirons on this process. Control non-fibrinogen binding Adhirons have no effect on this assay. The effects of the Adhirons include, reduced clot formation, increased lysis time, and increased clotting time. This demonstrates an ability of the Adhirons to modulate protein function by inhibiting protein-protein interactions. Figure 15 shows a confocal image of FITC fluorescently labelled fibrinogen after clot formation in the presence of a fibrinogen binding Adhiron and a control Adhiron.

Expression of Adhirons in mammalian cells:

Adhirons were raised against human SUMO2 as described in Example 1 and expressed in mammalian HEK293 cells using the pcDNA3.1 mammalian expression vector. The Adhirons were fused with a FLAG tag and a nuclear localisation signal. Control cells (no Adhiron expressed) and cells expressing a human SUMO2 specific Adhiron were treated with arsenic (As) for 2 hrs then washed and allowed to recover for 12 and 24 hr. Arsenic causes an increase in promyelocytic leukaemia (PML) protein bodies but SUMO regulates the degradation of these bodies. Cells were stained using an anti-FLAG antibody to identify the Adhiron, and for PML (organiser of nuclear bodies). The human SUMO2 Adhiron alters the degradation of PML leading to an increase in these bodies. This demonstrates ability to express functional Adhirons in mammalian cells. The results are shown in Figure 16.

Co-crystallisation and other structural biology methods to identify druggable sites on proteins:

Figure 17 shows the co-crystal structure of FcgRIIIa (grey) and bound Adhiron (white). Adhirons were identified that bind to FcgRIIIa then used in a range of assays to show that the inhibit IgG binding, including cell- and SPR-based assays. The Adhirons were then co-

crystallised and the structure solved (diagram above). This identified druggable sites, including an allosteric site, on FcgRIIIa. The Adhiron is also suitable for NMR studies and, as an example, Figure 18 shows 1H-15N HSQC spectra of an anti-yeast SUMO Adhiron and the Adhiron-yeast SUMO complex. The ability to rapidly collect structural data on Adhirons will have many applications including identifying potential drug binding sites.

Incorporation into electronic devices for developing point of care devices:

Site directed mutagenesis of the coding sequence of the human SUMO2 Adhiron (described in Example 1) allowed the introduction of a cysteine at the C-terminal end of the oligohistidine tag. This allowed directional immobilisation of the Adhiron to an electronic device surface such that the molecular recognition loops are accessible to analyte. Upon binding of the target protein to Adhiron on the device a change in impedance can be measured. The change is concentration dependent (as shown in Figure 19) demonstrating the effective presentation of the Adhiron and the ability of Adhirons to be productively incorporated into electronic devices as a platform for biosensor applications.

This protocol was repeated for another Adhiron, in this case a human fibrinogen Adhiron. The results of this work are shown in Fig 26. Once again a concentration dependent change was observed, and this was demonstrated over a range from attomolar to micromolar concentrations.

Adhirons have been identified that bind to a range of targets:

The Adhiron library has been used to screen against a wide range of target molecules by using display methodology described in this document. Table 5 lists examples of targets against which Adhirons have been raised; these include proteins and small molecules. The Adhirons raised against these targets display high affinity and specificity. This demonstrates that that Adhirons provides a versatile scaffold molecule and that the libraries built using this scaffold are effective in identifying artificial binding proteins capable of binding to a broad range of target molecules. Examples of some of the sequences that have been isolated from screens against some of the targets are shown in Figures 20-24. We have recently also screened against magnetic particles produced by magnetotropic bacteria and have identified Adhirons that bind to epitopes on these multicomponent bioinorganic complexes.

Table 5 – Targets against which Adhirons have been successfully raised.

FcγRIIIa – protein	Interleukin 8 – protein
P7 – peptide	Human serum albumin - protein
E5 – peptide	C-reactive protein - protein

3D – protein	Beta 2-microglobulin - protein
HSV-1 gB – protein	Serum amyloid P - protein
HSV-1gD- protein	Vascular endothelial growth factor receptor 2 – protein
HSV-2 gD – protein	Oxidized low-density lipoprotein receptor 1 – protein
M2 – protein	Allograft inflammatory factor 1 - protein
HE4 – protein	CD30 – protein
S100 calcium binding protein B - protein	CD31 – protein
Yeast SUMO – protein	Beta secretase1 - protein
Human SUMO1 – protein	Proprotein convertase subtilisin/kexin type 9 – protein
Human SUMO2 – protein	Myosin e1 - protein
GST – protein	Enhance green fluorescent protein - protein
Growth Hormone – protein	Fyn – protein
Fibroblast growth factor 1 – protein	Lck – protein
Fibroblast growth factor receptor 1 - protein	ZAP70 – protein
Fibroblast growth factor receptor 3 - protein	TUBA8 – peptide
Phospho Fibroblast growth factor receptor 3	MRP1- peptide
Epidermal growth factor receptor 1 - protein	penicillin-binding protein 2a - protein
HER2 – protein	Dog IgE – protein
Epiregulin – protein	Horse IgG - protein
Amphiregulin – protein	Dog IgG – protein
Fibrincgen – protein	Dog CRP – protein
Complement C3 – protein	3 small molecules
Myoglobin – protein	Posaconazole – small molecule
CK19 – protein	GP-73 - protein
HE4 – protein	Thioredoxin - protein
CD27 – protein	Signal transducer and activator of transcription 1 - protein
Signal transducer and activator of transcription 3 - protein	Signal transducer and activator of transcription 4 - protein
Signal transducer and activator of transcription 5 - protein	Phosphoinositide 3-kinase p85 alpha - protein
Phosphoinositide 3-kinase p85 beta - protein	Phosphoinositide 3-kinase p55 - protein
myeloid leukemia cell differentiation protein – protein	B-cell lymphoma-extra large - protein
Nucleoside transporter protein C – membrane protein	Factor XIII - protein
Breakpoint cluster region protein - protein	Casein kinase II A1 - protein
Casein kinase II A2 - protein	Protein kinase C zeta - protein
Protein kinase cGMP dependent type II – protein	vaccinia related kinase 1 - protein
Hydrophobin protein 1 - protein	FusB - protein
Interleukin 17A - protein	Interleukin 6 - protein
Osteocalcin - protein	Osteopontin - protein
Parathyroid hormone - protein	Bacterial spores - protein
matrix metalloproteinase-3 - protein	Aminotransferase - protein
Cytokeratin 8 -protein	S100 – A3 - protein
S100 A6 – protein	Transacetylase – protein

serine protease inhibitors A1 - protein	serine protease inhibitors A3 - protein
GAPDH – protein	p53 - protein
Resistin – protein	Lipocalin 2 - protein
Procalcitonin - protein	

REFERENCES

Astwood, J. D., J. N. Leach, et al. (1996). "Stability of food allergens to digestion in vitro." Nature Biotechnology **14**(10): 1269-1273.

Atkinson, H. J., K. A. Johnston, et al. (2004). "Prima facie evidence that a phytocystatin for transgenic plant resistance to nematodes is not a toxic risk in the human diet." Journal of Nutrition **134**(2): 431-434.

Bendtsen, J. D., H. Nielsen, et al. (2004). "Improved prediction of signal peptides: SignalP 3.0." Journal of Molecular Biology **340**(4): 783-795.

Binz, H. K., M. T. Stumpp, et al. (2003). "Designing Repeat Proteins: Well-expressed, Soluble and Stable Proteins from Combinatorial Libraries of Consensus Ankyrin Repeat Proteins." Journal of Molecular Biology **332**(2): 489-503.

Bode, W., R. Engh, et al. (1988). "The 2.0 a X-Ray Crystal-Structure of Chicken Egg-White Cystatin and Its Possible Mode of Interaction With Cysteine Proteinases." Embo Journal **7**(8): 2593-2599.

Carter, P. J. (2011). "Introduction to current and future protein therapeutics: A protein engineering perspective." Experimental Cell Research **317**(9): 1261-1269.

Dai, M. H., H. E. Fisher, et al. (2007). "The creation of a novel fluorescent protein by guided consensus engineering." Protein Engineering Design & Selection **20**(2): 69-79.

Deboer, H. A., L. J. Comstock, et al. (1983). "THE TAC PROMOTER - A FUNCTIONAL HYBRID DERIVED FROM THE TRP AND LAC PROMOTERS." Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences **80**(1): 21-25.

Filippova, I. Y., E. N. Lysogorskaya, et al. (1984). "L-Pyroglutamyl-L-Phenylalanyl-L-Leucine-Para-Nitroanilide - a Chromogenic Substrate for Thiol Proteinase Assay." Analytical Biochemistry **143**(2): 293-297.

FitzGerald, K. (2000). "In vitro display technologies - new tools for drug discovery." Drug Discovery Today **5**(6): 253-258.

Forrer, P., H. K. Binz, et al. (2004). "Consensus design of repeat proteins." ChemBioChem **5**: 183-189.

Forrer, P., H. K. Binz, et al. (2004). "Consensus Design of Repeat Proteins." ChemBioChem **5**(2): 183-189.

Gebauer, M. and A. Skerra (2009). "Engineered protein scaffolds as next-generation antibody therapeutics." Current Opinion in Chemical Biology **13**(3): 245-255.

Grebien, F., O. Hantschel, et al. (2011). "Targeting the SH2-Kinase Interface in Bcr-Abl Inhibits Leukemogenesis." Cell **147**(2): 306-319.

Ho, M. and I. Pastan (2009). Mammalian Cell Display for Antibody Engineering. Methods in Molecular Biology. A. S. Dimitrov. **525:** 337-352.

Hoffmann, T., L. K. Stadler, et al. (2010). "Structure-function studies of an engineered scaffold protein derived from stefin A. I: Development of the SQM variant." Protein Eng Des Sel **23**(5): 403-413.

Hoogenboom, H. R., A. D. Griffiths, et al. (1991). "Multisubunit Proteins on the Surface of Filamentous Phage - Methodologies for Displaying Antibody (Fab) Heavy and Light-Chains." Nucleic Acids Research **19**(15): 4133-4137.

Horton, R., Z. Cai, et al. (1990). "Gene splicing by overlap extension: tailor-made genes using the polymerase chain reaction." Biotechniques **8**: 528-535.

Hutchison, C. A., S. Phillips, et al. (1978). "Mutagenesis at a specific position in a DNA sequence." Journal of Biological Chemistry **253**(18): 6551-6560.

Jacobs, S. A., M. D. Diem, et al. (2012). "Design of novel FN3 domains with high stability by a consensus sequence approach." Protein Engineering Design & Selection **25**(3): 107-117.

Jacobs, S. A., M. D. Diem, et al. (2012). "Design of novel FN3 domains with high stability by a consensus sequence approach." Protein Eng Des Sel **25**(3): 107-117.

Jaeckel, C., J. D. Bloom, et al. (2010). "Consensus Protein Design without Phylogenetic Bias." Journal of Molecular Biology **399**(4): 541-546.

Karatan, E., M. Merguerian, et al. (2004). "Molecular recognition properties of FN3 monobodies that bind the Src SH3 domain." Chemistry & Biology **11**(6): 835-844.

Kayushin, A., M. Korosteleva, et al. (1996). "A convenient approach to the synthesis of trinucleotide phosphoramidites-synthons for the generation of oligonucleotide/peptide libraries." Nucleic Acids Res **24**: 3748 - 3755.

Knappik, A., L. Ge, et al. (2000). "Fully synthetic human combinatorial antibody libraries (HuCAL) based on modular consensus frameworks and CDRs randomized with trinucleotides." Journal of Molecular Biology **296**(1): 57-86.

Kohler, G. and C. Milstein (1975). "Continuous cultures of fused cells secreting antibody of predefined specificity." Nature **256**(5517): 495-497.

Koide, A., C. W. Bailey, et al. (1998). "The fibronectin type III domain as a scaffold for novel binding proteins." Journal of Molecular Biology **284**(4): 1141-1151.

Koide, A., C. W. Bailey, et al. (1998). "The fibronectin type III domain as a scaffold for novel binding proteins." Journal of Molecular Biology **284**(4): 1141-1151.

Koide, A., R. N. Gilbreth, et al. (2007). "High-affinity single-domain binding proteins with a binary-code interface." Proceedings of the National Academy of Sciences of the United States of America **104**(16): 6632-6637.

Koiwa, H., M. P. D'Urzo, et al. (2001). "Phage display selection of hairpin loop soyacystatin variants that mediate high affinity inhibition of a cysteine proteinase." Plant Journal **27**(5): 383-391.

Komor, R. S., P. A. Romero, et al. (2012). "Highly thermostable fungal cellobiohydrolase I (Cel7A) engineered using predictive methods." Protein engineering, design & selection : PEDS **25**(12): 827-833.

Komor, R. S., P. A. Romero, et al. (2012). "Highly thermostable fungal cellobiohydrolase I (Cel7A) engineered using predictive methods." Protein Engineering Design and Selection **25**(12): 827-833.

Kondo, H., K. Abe, et al. (1991). "Gene Organization of Oryzacystatin-Ii, a New Cystatin Superfamily Member of Plant-Origin, Is Closely Related to That of Oryzacystatin-I But Different From Those of Animal Cystatins." Febs Letters **278**(1): 87-90.

Kordis, D. and V. Turk (2009). "Phylogenomic analysis of the cystatin superfamily in eukaryotes and prokaryotes." Bmc Evolutionary Biology **9**.

Krumpe, L., K. Schumacher, et al. (2007). "Trinucleotide cassettes increase diversity of T7 phage-displayed peptide library." BMC Biotechnology **7**(1): 65.

Lee, S.-C., K. Park, et al. (2012). "Design of a binding scaffold based on variable lymphocyte receptors of jawless vertebrates by module engineering." Proceedings of the National Academy of Sciences **109**(9): 3299-3304.

Lehmann, M., C. Loch, et al. (2002). "The consensus concept for thermostability engineering of proteins: further proof of concept." Protein Engineering **15**(5): 403-411.

Lehmann, M., R. Lopez-Ulibarri, et al. (2000). "Exchanging the active site between phytases for altering the functional properties of the enzyme." **9**(10): 1866-1872.

Lehmann, M. and M. Wyss (2001). "Engineering proteins for thermostability: the use of sequence alignments versus rational design and directed evolution." Current Opinion in Biotechnology **12**(4): 371-375.

Lilley, C. J., P. E. Urwin, et al. (2004). "Preferential expression of a plant cystatin at nematode feeding sites confers resistance to Meloidogyne incognita and Globodera pallida." Plant Biotechnology Journal **2**(1): 3-12.

Lofblom, J. (2011). "Bacterial display in combinatorial protein engineering." Biotechnology Journal **6**(9): 1115-1129.

Lord, P. W., J. N. Selley, et al. (2002). "CINEMA-MX: a modular multiple alignment editor." Bioinformatics **18**(10): 1402-1403.

Main, E. R. G., S. E. Jackson, et al. (2003). "The folding and design of repeat proteins: reaching a consensus." Current Opinion in Structural Biology **13**(4): 482-489.

Main, E. R. G., A. R. Lowe, et al. (2005). "A recurring theme in protein engineering: the design, stability and folding of repeat proteins." Current Opinion in Structural Biology **15**(4): 464-471.

Main, E. R. G., Y. Xiong, et al. (2003). "Design of stable alpha-helical arrays from an idealized TPR motif." Structure **11**(5): 497-508.

Makela, A. R. and C. Oker-Blom (2008). "The baculovirus display technology - An evolving instrument for molecular screening and drug delivery." Combinatorial Chemistry & High Throughput Screening **11**(2): 86-98.

Margis, R., E. M. Reis, et al. (1998). "Structural and phylogenetic relationships among plant and animal cystatins." Archives of Biochemistry and Biophysics **359**(1): 24-30.

McPherson, M. J., P. E. Urwin, et al. (1997). Engineering plant nematode resistance by an anti-feedant approach. Cellular and Molecular Basis for Plant-Nematode Interactions. C. Fenoll, S. Ohl and F. Grundler. The Netherlands, Kluwer: 237-249.

Melo, F. R., M. O. Mello, et al. (2003). "Use of phage display to select novel cystatins specific for Acanthoscelides obtectus cysteine proteinases." Biochimica Et Biophysica Acta-Proteins and Proteomics **1651**(1-2): 146-152.

Mosavi, L. K., T. J. Cammett, et al. (2004). "The ankyrin repeat as molecular architecture for protein recognition." Protein Science **13**(6): 1435-1448.

Mosavi, L. K., D. L. Minor, et al. (2002). "Consensus-derived structural determinants of the ankyrin repeat motif." Proceedings of the National Academy of Sciences **99**(25): 16029-16034.

Mullis, K., F. Faloona, et al. (1986). "Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction." Cold Spring Harb Symp Quant Biol **51 Pt 1**: 263-273.

Nagata, K., N. Kudo, et al. (2000). "Three-dimensional solution structure of oryzacystatin-I, a cysteine proteinase inhibitor of the rice, Oryza sativa L. japonica." Biochemistry **39**(48): 14753-14760.

Nixon AE, W. C. (2006). "Engineered protein inhibitors of proteases." Curr Opin Drug Discov Devel **9**(2): 261-268.

Nord, K., J. Nilsson, et al. (1995). "A combinatorial library of an α-helical bacterial receptor domain." Protein Engineering **8**(6): 601-608.

Odegrip, R., D. Coomber, et al. (2004). "CIS display: In vitro selection of peptides from libraries of protein-DNA complexes." Proceedings of the National Academy of Sciences of the United States of America **101**(9): 2806-2810.

Parizek, P., L. Kummer, et al. (2012). "Designed Ankyrin Repeat Proteins (DARPins) as Novel Isoform-Specific Intracellular Inhibitors of c-Jun N-Terminal Kinases." ACS Chemical Biology **7**(8): 1356-1366.

Parmeggiani, F., R. Pellarin, et al. (2008). "Designed armadillo repeat proteins as general peptide-binding scaffolds: Consensus design and computational optimization of the hydrophobic core." Journal of Molecular Biology **376**(5): 1282-1304.

Parry-Smith, D. J., A. W. R. Payne, et al. (1998). "CINEMA - a novel colour interactive editor for multiple alignments (Reprinted from Gene, vol 221, pg GC57-GC63, 1998)." Gene **221**(1): GC57-GC63.

Polizzi, K. M., J. F. Chaparro-Riggers, et al. (2006). "Structure-guided consensus approach to create a more thermostable penicillin G acylase." Biotechnology Journal **1**(5): 531-536.

Reichert, J. M. (2010). "Antibodies to watch in 2010." MAbs. **2**(1): 84-100.

Saiki RK, S. S., Faloona F, Mullis KB, Horn GT, Erlich HA, Arnheim N. (1985). "Enzymatic amplification of beta-globin genomic sequences and

restriction site analysis for diagnosis of sickle cell anemia." Science **230**(4732): 1350-1354.
Schlehuber, S. and A. Skerra (2005). "Anticalins as an alternative to antibody technology." Expert Opinion on Biological Therapy **5**(11): 1453-1462.
Skerra, A. (2007). "Alternative non-antibody scaffolds for molecular recognition." Current Opinion in Biotechnology **18**: 295-304.
Smith, G. (1985). "Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface." Science **228**(4705): 1315-1317.
Song, I., M. Taylor, et al. (1995). "Inhibition of Cysteine Proteinases By Carica-Papaya Cystatin Produced in Escherichia-Coli." Gene **162**(2): 221-224.
Song, J., L. K. Durrin, et al. (2004). "Identification of a SUMO-binding motif that recognizes SUMO-modified proteins." Proc Natl Acad Sci U S A **101**(40): 14373-14378.
Song, J., Z. Zhang, et al. (2005). "Small ubiquitin-like modifier (SUMO) recognition of a SUMO binding motif: a reversal of the bound orientation." J Biol Chem **280**(48): 40122-40129.
Stadler, L. K., T. Hoffmann, et al. (2011). "Structure-function studies of an engineered scaffold protein derived from Stefin A. II: Development and applications of the SQT variant." Protein Eng Des Sel **24**(9): 751-763.
Steipe, B. (2004). "Consensus-based engineering of protein stability: From intrabodies to thermostable enzymes." Protein Engineering **388**: 176-186.
Steipe, B., B. Schiller, et al. (1994). "Sequence Statistics Reliably Predict Stabilizing Mutations in a Protein Domain." Journal of Molecular Biology **240**(3): 188-192.
Steipe, B., B. Schiller, et al. (1994). "Sequence Statistics Reliably Predict Stabilizing Mutations in a Protein Domain." Journal of Molecular Biology **240**(3): 188-192.
Stubbs, M. T., B. Laber, et al. (1990). "The Refined 2.4a X-Ray Crystal-Structure of Recombinant Human Stefin- B in Complex With the Cysteine Proteinase Papain - a Novel Type of Proteinase-Inhibitor Interaction." Embo Journal **9**(6): 1939-1947.
Studier, F. W. and B. A. Moffatt (1986). "Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes." J Mol Biol **189**(1): 113-130.
Theurillat, J.-P., B. Dreier, et al. (2010). "Designed ankyrin repeat proteins: a novel tool for testing epidermal growth factor receptor 2 expression in breast cancer." Mod Pathol **23**(9): 1289-1297.
Traxlmayr, M. W. and C. Obinger (2012). "Directed evolution of proteins for increased stability and expression using yeast display." Archives of Biochemistry and Biophysics **526**(2): 174-180.
Urwin, P. E., H. J. Atkinson, et al. (1995). "Engineered Oryzacystatin-I Expressed in Transgenic Hairy Roots Confers Resistance to Globodera-Pallida." Plant Journal **8**(1): 121-131.
Urwin, P. E., J. Green, et al. (2003). "Expression of a plant cystatin confers partial resistance to Globodera, full resistance is achieved by pyramiding a cystatin with natural resistance." Molecular Breeding **12**(3): 263-269.
Urwin, P. E., A. Levesley, et al. (2000). "Transgenic resistance to the nematode Rotylenchulus reniformis conferred by Arabidopsis thaliana plants expressing proteinase inhibitors." Molecular Breeding **6**(3): 257-264.
Urwin, P. E., C. J. Lilley, et al. (1997). "Resistance to both cyst and root-knot nematodes conferred by transgenic Arabidopsis expressing a modified plant cystatin." Plant Journal **12**(2): 455-461.
Urwin, P. E., M. J. McPherson, et al. (1998). "Enhanced transgenic plant resistance to nematodes by dual proteinase inhibitor constructs." Planta **204**(4): 472-479.
Urwin, P. E., K. M. Troth, et al. (2001). "Effective transgenic resistance to Globodera pallida in potato field trials." Molecular Breeding **8**(1): 95-101.
Virnekas, B., L. Ge, et al. (1994). "Trinucleotide phosphoramidites: ideal reagents for the synthesis of mixed oligonucleotides for random mutagenesis." Nucleic Acids Res **22**: 5600 - 5607.

Von Behring, E., Kitasato, S., (1890). "Über das Zustandekommen der Diphterie-Immunität und der Tetanus-Immunität bei Thieren." Deutsche Medizinische Wochenzeitschrift **16**: 1113-1114.

Wojcik, J., O. Hantschel, et al. (2010). "A potent and highly specific FN3 monobody inhibitor of the Abl SH2 domain." Nat Struct Mol Biol **17**(4): 519-527.

Woodman, R., J. T. H. Yeh, et al. (2005). "Design and validation of a neutral protein scaffold for the presentation of peptide aptamers." Journal of Molecular Biology **352**(5): 1118-1133.

Wurch, T., A. Pierre, et al. (2012). "Novel protein scaffolds as emerging therapeutic proteins: from discovery to clinical proof-of-concept." Trends in Biotechnology **30**(11): 575-582.

Claims

1. A synthetic scaffold protein comprising a sequence derived from one or more plant cystatins, wherein the synthetic scaffold protein comprises an amino acid sequence which is at least 85% identical to NSLEIEELARFAVDEHNKKENALLEFVRVVKAKEQVVAGTMYYLTLEAKDGGKKKLYEAKVWVKPWENFKELQEFK PVGDA (SEQ ID NO 1), wherein the synthetic scaffold protein further comprises at least one heterologous peptide sequence, wherein the heterologous peptide is from 3 to 30 amino acids in length, wherein the heterologous peptide is inserted: (a) adjacent to any of the amino acid residues VVAG (SEQ ID NO:82); (b) in replacement of any of the amino acid residues VVAG(SEQ ID NO:82); (c) adjacent to any of the amino acid residues PWE; (d) in replacement of any of the amino acid residues PWE; or (e) any combination thereof, wherein the at least one heterologous peptide is heterologous to the amino acid sequence of SEQ ID NO:1, and wherein the at least one heterologous peptide is disregarded when calculating sequence identity.

2. The synthetic protein of claim 1 wherein the heterologous peptide sequence is from 3 to 20 amino acids in length.

3. The synthetic protein of claim 1 wherein the heterologous peptide sequence is from 5 to 13 amino acids in length.

4. The synthetic protein of any one of claims 1 to 3 comprising a further heterologous peptide inserted at or within 4 amino acids of the N-terminus of the protein.

5. The synthetic protein according to claim 1 which comprises the sequence:

NSLEIEELARFAVDEHNKKENALLEFVRVVKAKEQ(Xn)TMYYLTLEAKDG
GKKKLYEAKVWVK(Xn)NFKELQEFKPVGDA (SEQ ID NO 4)

wherein Xn is any combination of 3 to 30 amino acids.

Date Reçue/Date Received 2022-01-14

6. The synthetic protein according to claim 5 wherein Xn is any combination of 4 to 20 amino acids.

7. The synthetic protein according to claim 5 wherein Xn is any combination of 4 to 15 amino acids.

8. The synthetic protein of claim 5 which comprises the sequence:

NSLEIEELARFAVDEHNKKENALLEFVRVVKAKEQ(X5-13)TMYYLTLEA KDGGKKKLYEAKVWVK(X5-13)NFKELQEFKPVGDA (SEQ ID NO 5) wherein X5-13 is any combination of 5-13 amino acids.

9. The synthetic protein of claim 6 which comprises the sequence:

NSLEIEELARFAVDEHNKKENALLEFVRVVKAKEQ(X9)TMYYLTLEA KDGGKKKLYEAKVWVK(X9) NFKELQEFKPVGDA (SEQ ID NO 6) wherein X9 is any combination of 9 amino acids.

10. The synthetic protein of any one of claims 5 to 9 which comprises an additional amino acid sequence at or within 4 amino acids of the N-terminus.

11. The synthetic protein of claim 10 wherein the additional amino acid sequence is at the N-terminus and comprises the sequence MATGVRAVPGNE (SEQ ID NO 80).

12. The synthetic protein of claim 10 wherein the additional amino acid sequence is at the N-terminus and comprises a further heterologous peptide sequence.

13. The synthetic protein of any one of claims 1-12 which has a melting temperature (Tm) of 60 °C or higher.

Date Reçue/Date Received 2022-01-14

14. The synthetic protein of any one of claims 1-12, which has a melting temperature (Tm) of 70 °C or higher.

15. The synthetic protein of any one of claims 1-12, which has a melting temperature (Tm) of 80 °C or higher.

16. The synthetic protein of any one of claims 1-15 further comprising a linker or tag, wherein the linker or tag is connected to the C- or N-terminus of the protein.

17. The synthetic protein of any one of claims 1-16 connected to a substrate or moiety, wherein the moiety is a label, carrier, or a further protein.

18. The synthetic protein of claim 17 wherein the substrate or moiety is attached to one of the C- or N-terminus of the protein.

19. A library comprising a population of synthetic proteins, wherein each synthetic protein is defined according to any one of claims 1 to 18 and wherein members of the population comprise a variety of heterologous peptides having different sequences.

20. The library of claim 19 having a complexity of 10^8 or higher.

21. The library of claim 19 having a complexity of 10^9 or higher.

22. The library of claim 19 having a complexity of 10^{10} or higher.

23. The library of claim 19 to 22, wherein the population of the synthetic proteins is adapted for display in a display system suitable for biopanning.

Date Reçue/Date Received 2022-01-14

24. A polynucleotide comprising a coding sequence which encodes the synthetic protein according to any one of claims 1 to 15.

25. The polynucleotide according to claim 24 further comprising, in addition to the coding sequence which encodes the synthetic protein, an additional coding or non-coding sequence.

26. The polynucleotide of claim 25 wherein the additional coding or non-coding sequence is a sequence encoding a heterologous peptide, a sequence encoding a leader sequence, a sequence encoding a linker, a sequence encoding a marker, a sequence encoding a purification tag, a promoter sequence or an enhancer sequence.

27. A recombinant vector comprising the polynucleotide according to any one of claims 24 to 26.

28. The vector according to claim 27 further comprising an expression control sequence operably linked to the polynucleotide which encodes the synthetic protein to control expression therefrom.

29. A cell comprising the polynucleotide according to any one of claims 24 to 26 or the vector according to any one of claims 27 to 28.

31. A cell culture comprising one or more cells according to claim 29 and a culture medium.

32. A method of screening for synthetic protein candidates that specifically bind to a target, the method comprising:

a) providing the library according to any one of claims 19 to 23;

b) exposing said library to the target; and

Date Reçue/Date Received 2022-01-14

c) selecting for synthetic scaffold proteins which specifically bind to the target.

33. The method of claim 32 wherein the target is a protein, peptide, a nucleic acid or a small molecule.

34. The method of claim 32 or 33 which uses a display system.

35. The method of any one of claims 32 to 34 which is a method of phage display.

36. The method of claim 35 wherein each of the selected synthetic scaffold proteins comprises a fusion of the synthetic protein with a bacteriophage coat protein, so that the synthetic protein is displayed on the surface of a viral particle.

37. A composition comprising the synthetic protein according to any one of claims 1 to 18 and a pharmaceutically acceptable carrier or excipient.

Date Reçue/Date Received 2022-01-14

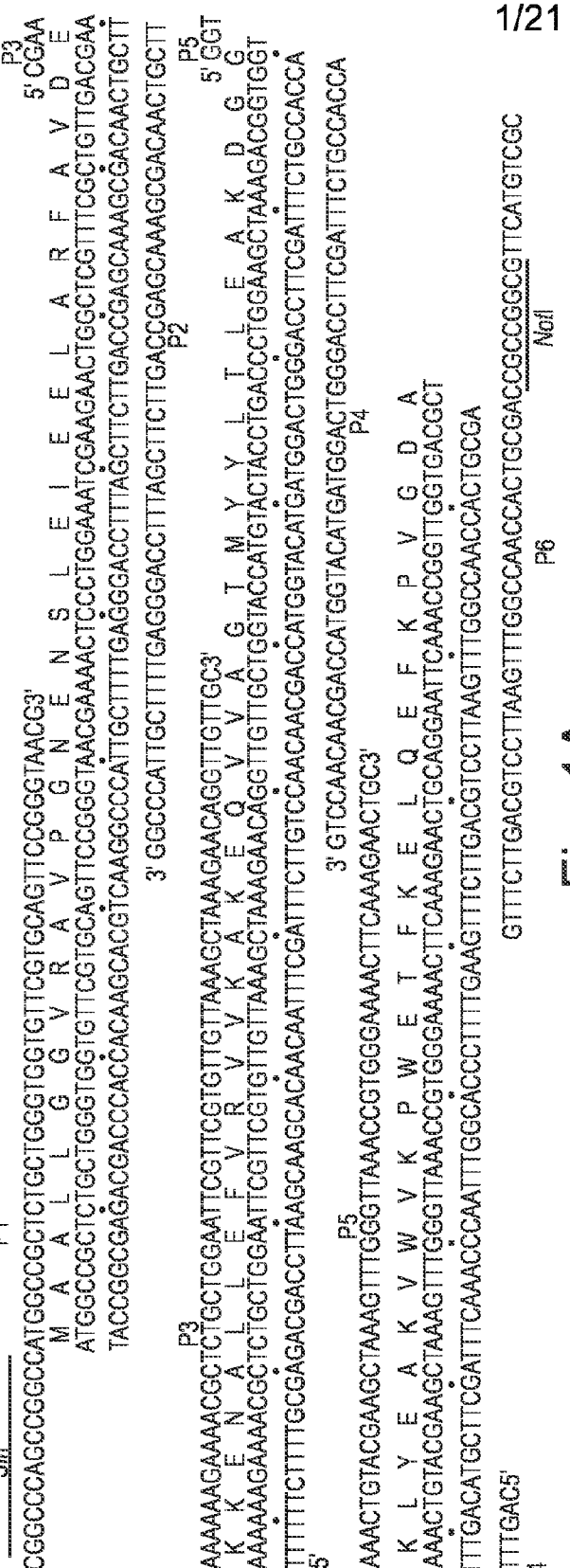


Fig. 1A

Fig. 1B

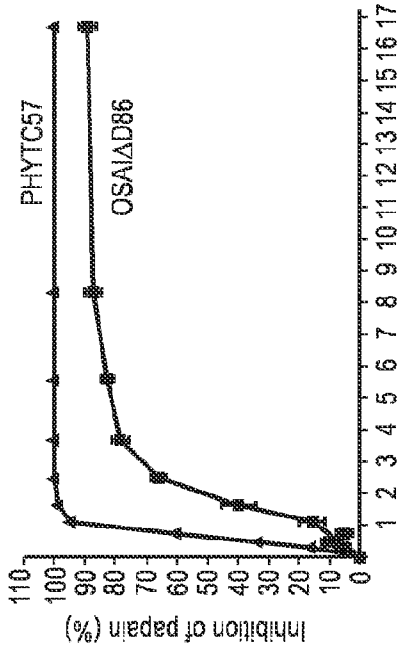


Fig. 2

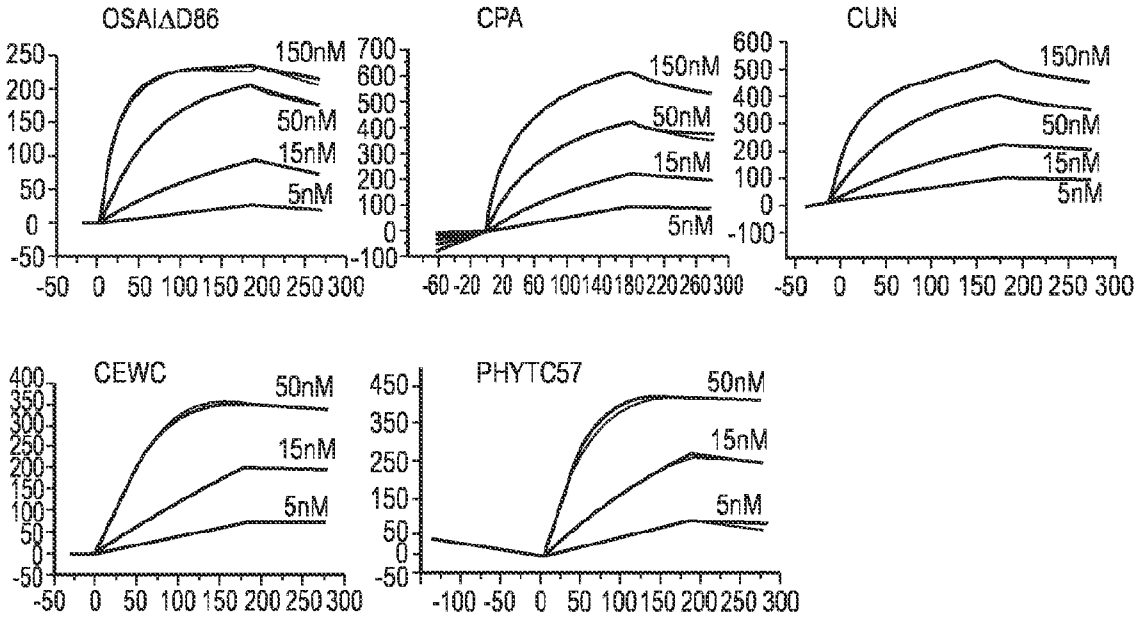

OSAIΔD86
150nM
50nM
15nM
5nM
CPA
150nM
50nM
15nM
5nM
CUN
150nM
50nM
15nM
5nM
CEWC
50nM
15nM
5nM
PHYTC57
50nM
15nM
5nM

Fig. 3

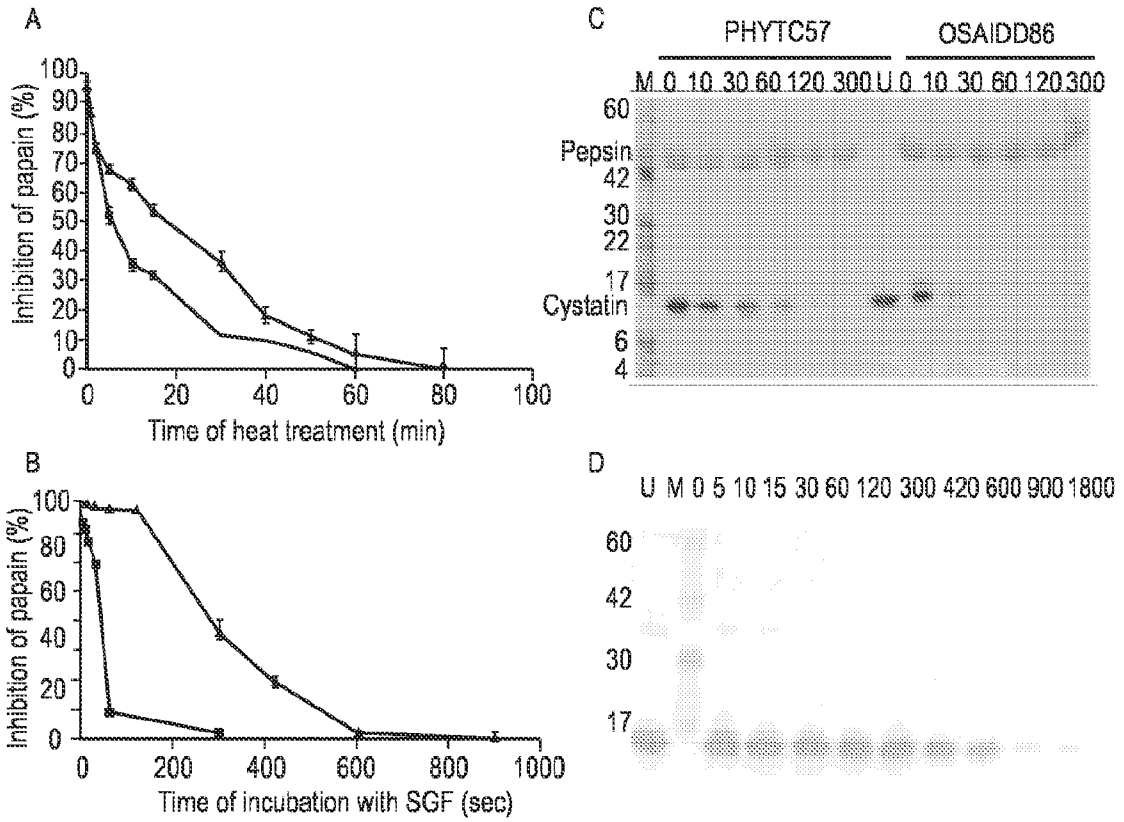

A
Inhibition of papain (%)
Time of heat treatment (min)
B
Inhibition of papain (%)
Time of incubation with SGF (sec)
C
PHYTC57
OSAIDD86
M 0 10 30 60 120 300 U 0 10 30 60 120 300
Pepsin
Cystatin
D
U M 0 5 10 15 30 60 120 300 420 600 900 1800

Fig. 4

A

```
OSAIDD86 MSSDGGPVLG GVEPV.GNEN DLHLVDLARF AVTEHNKKAN SLLEFEKLVS
PHYTC57  ....MAALLG GVRAVPGNEN SLEIEELARF AVDEHNKKEN ALLEFVRVVK

OSAIDD86 VKQQVVAGTL YYFTIEVKEG DAKKLYEAKV WEKPWM.FKE LQEFKPVDASANA
PHYTC57  AKEQVVAGTM YYLTLEAKDG GKKKLYEAKV WVKPWETFKE LQEFKPVGDA
```

B

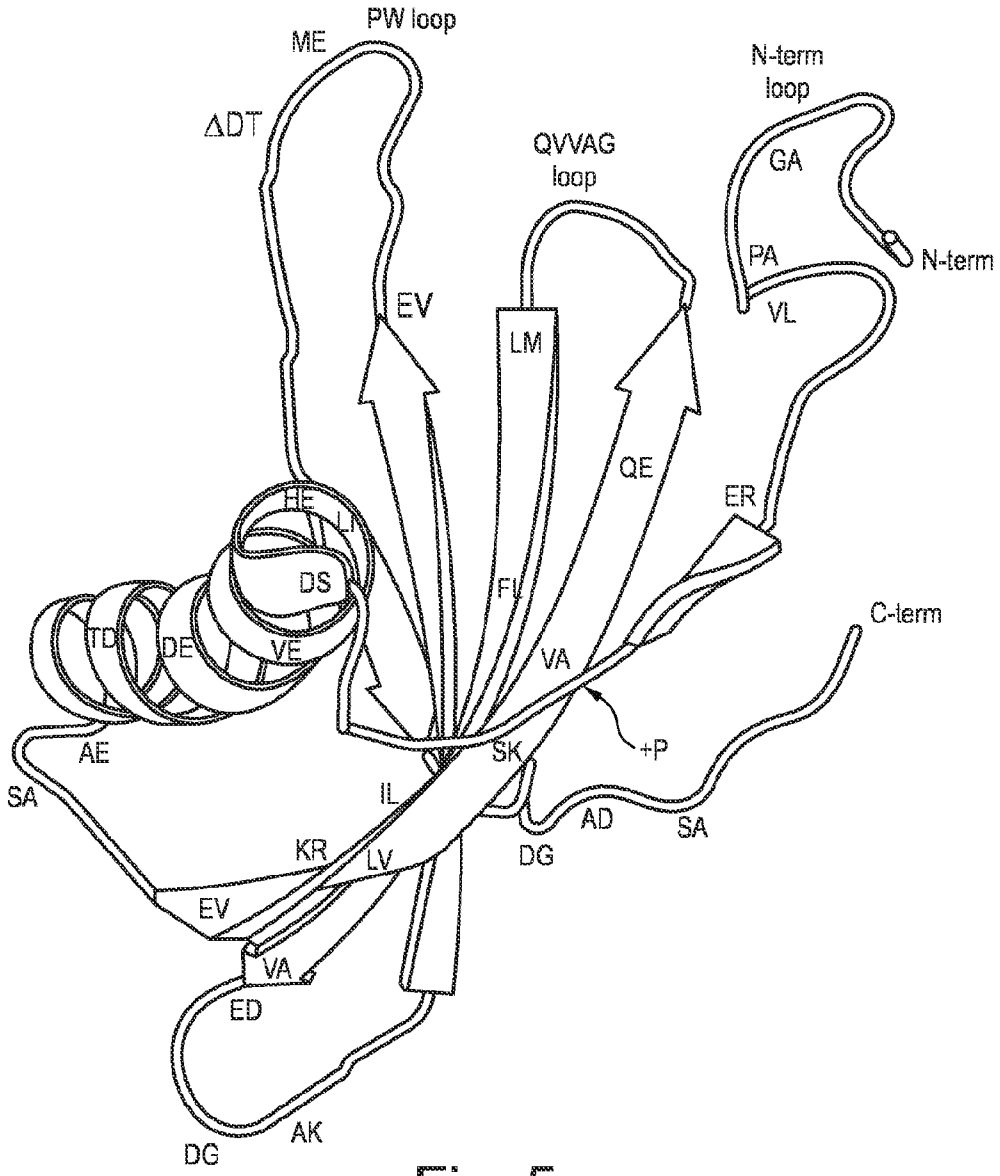


Fig. 5

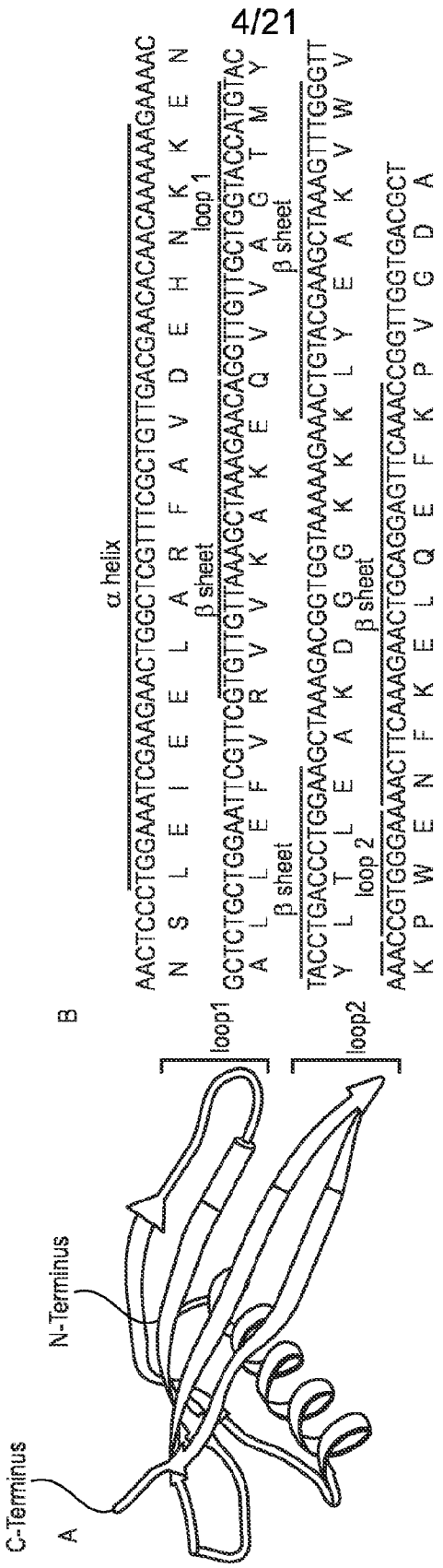
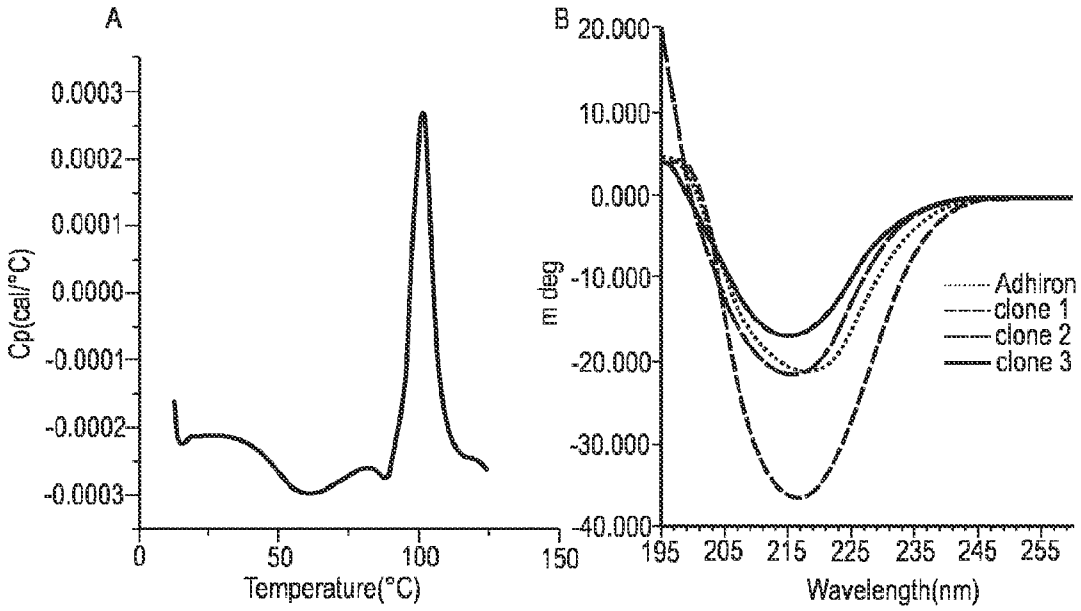

α helix
AACTCCCTGGAAATCGAAGAACTGGCTCGTTTCGCTGTTGACGAACACAACAAAAAAGAAAAC
N S L E I E E L A R F A V D E H N K K E N
β sheet
loop 1
GCTCTGCTGGAATTCGTTCGTGTTGTTAAAGCTAAAGAACAGGTTGTTGCTGGTACCATGTAC
A L L E F V R V V K A K E Q V V A G T M Y
β sheet
β sheet
TACCTGACCCTGGAAGCTAAAGACGGTGGTAAAAAGAACTGTACGAAGCTAAAGTTTGGGTT
Y L T L E A K D G G K K K L Y E A K V W V
loop 2
β sheet
AAACCGTGGGAAAACTTCAAAGAACTGCAGGAGTTCAAACCGGTTGGTGACGCT
K P W E N F K E L Q E F K P V G D A
B
loop1
loop2
N-Terminus
C-Terminus
A

Fig. 6


A
Cp(cal/°C)
0.0003
0.0002
0.0001
0.0000
-0.0001
-0.0002
-0.0003
0
50
100
150
Temperature(°C)
B
m deg
20.000
10.000
0.000
-10.000
-20.000
-30.000
-40.000
195
205
215
225
235
245
255
Wavelength(nm)
Adhiron
clone 1
clone 2
clone 3

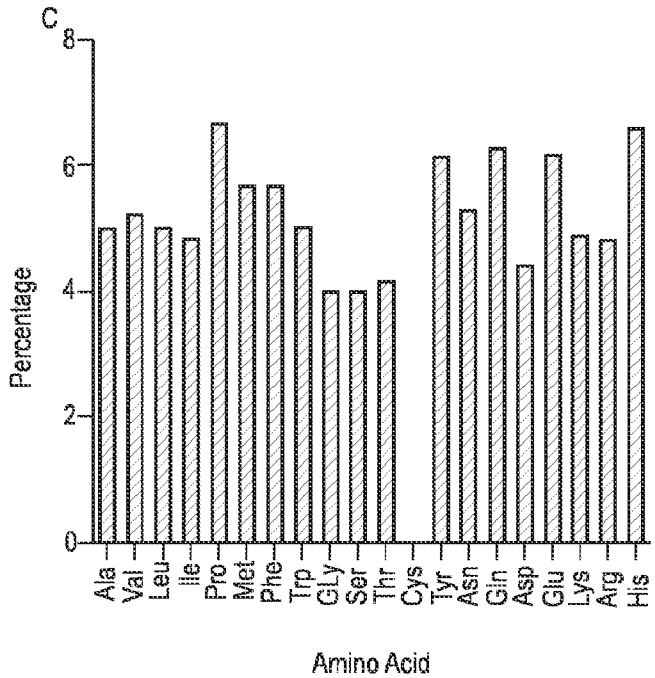

C
Percentage
8
6
4
2
0
Ala
Val
Leu
Ile
Pro
Met
Phe
Trp
GLy
Ser
Thr
Cys
Tyr
Asn
Gln
Asp
Glu
Lys
Arg
His
Amino Acid

Fig. 7

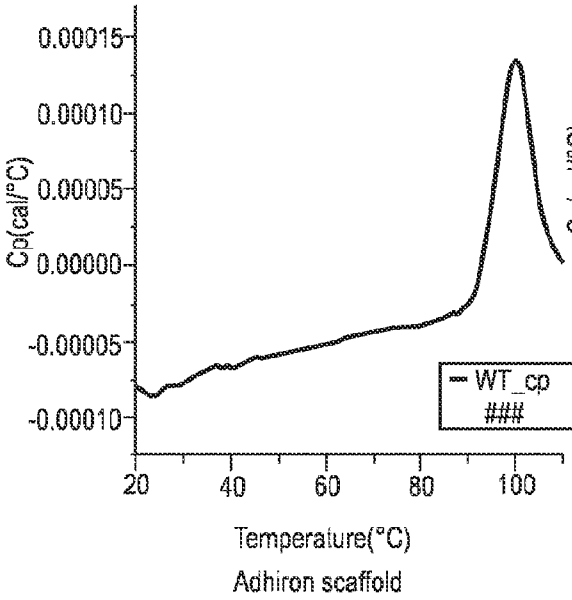


Fig. 8a

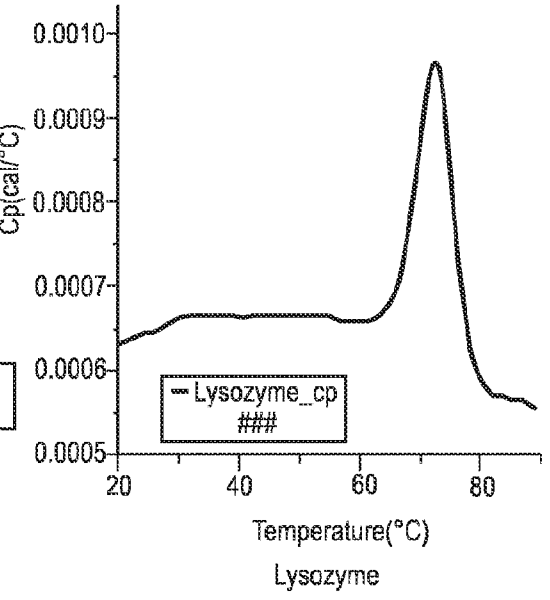


Fig. 8b

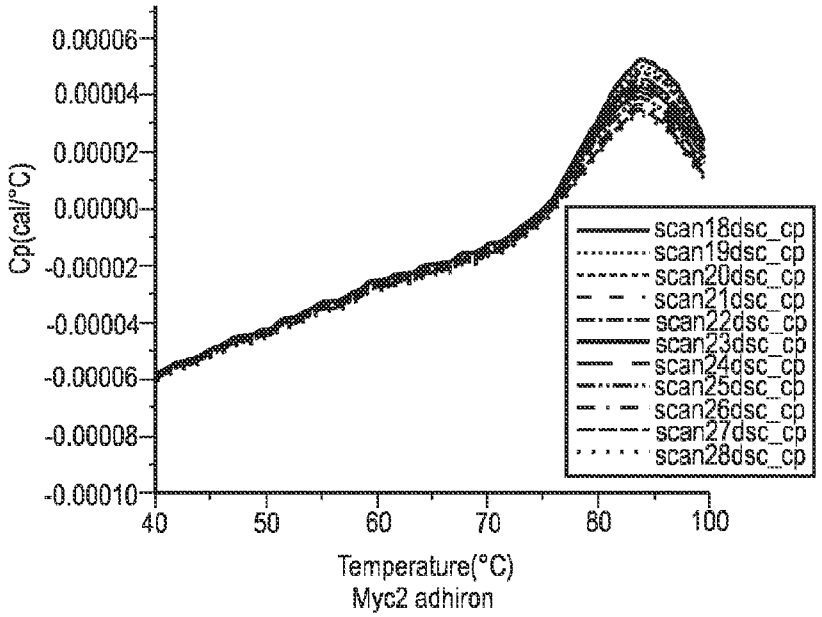


Fig. 8c

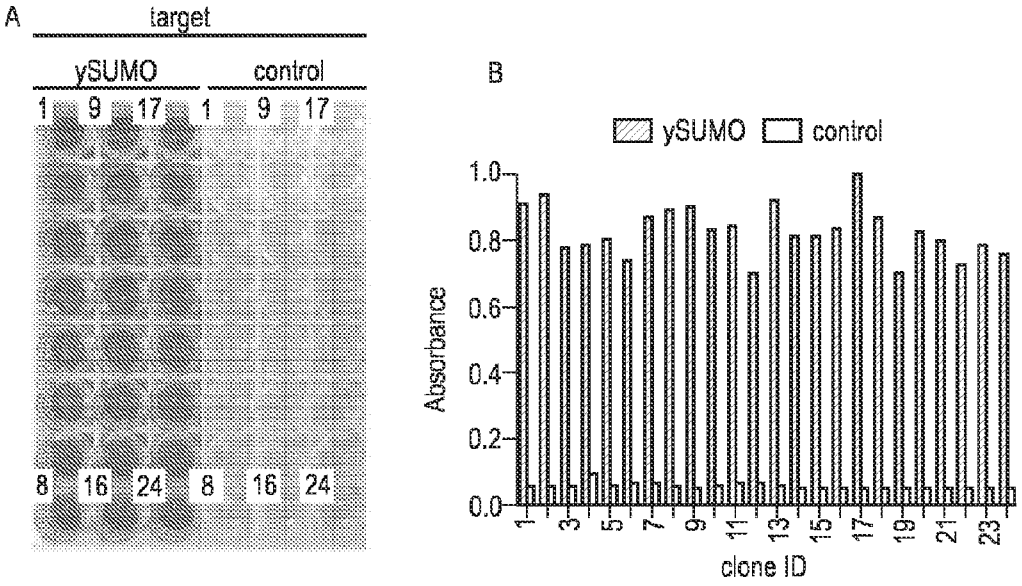
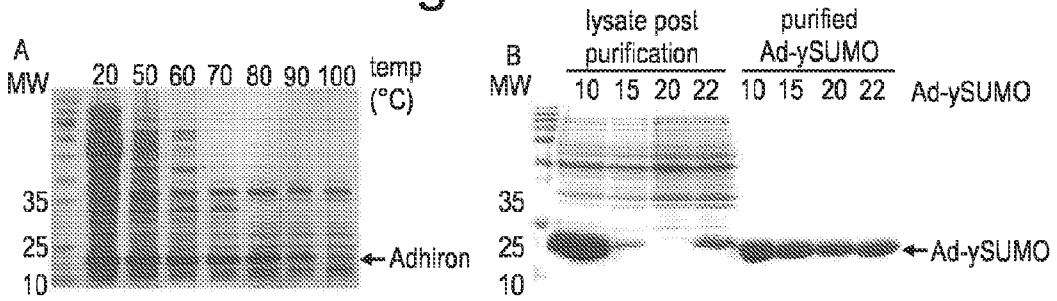

A
target
ySUMO
control
1 9 17 1 9 17
8 16 24 8 16 24
B
ySUMO
control
Absorbance
1.0
0.8
0.6
0.4
0.2
0.0
1 3 5 7 9 11 13 15 17 19 21 23
clone ID

Fig. 9


A
MW
20 50 60 70 80 90 100
temp (°C)
35
25
10
Adhiron
B
MW
lysate post purification
purified Ad-ySUMO
10 15 20 22 10 15 20 22
Ad-ySUMO
35
25
10
Ad-ySUMO

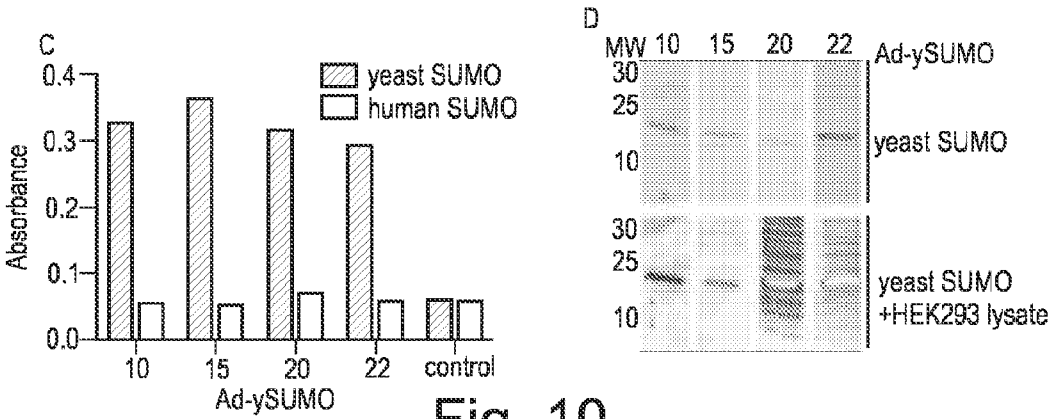

C
yeast SUMO
human SUMO
Absorbance
0.4
0.3
0.2
0.1
0.0
10 15 20 22 control
Ad-ySUMO
D
MW 10 15 20 22
Ad-ySUMO
30
25
10
yeast SUMO
30
25
10
yeast SUMO
+HEK293 lysate

Fig. 10

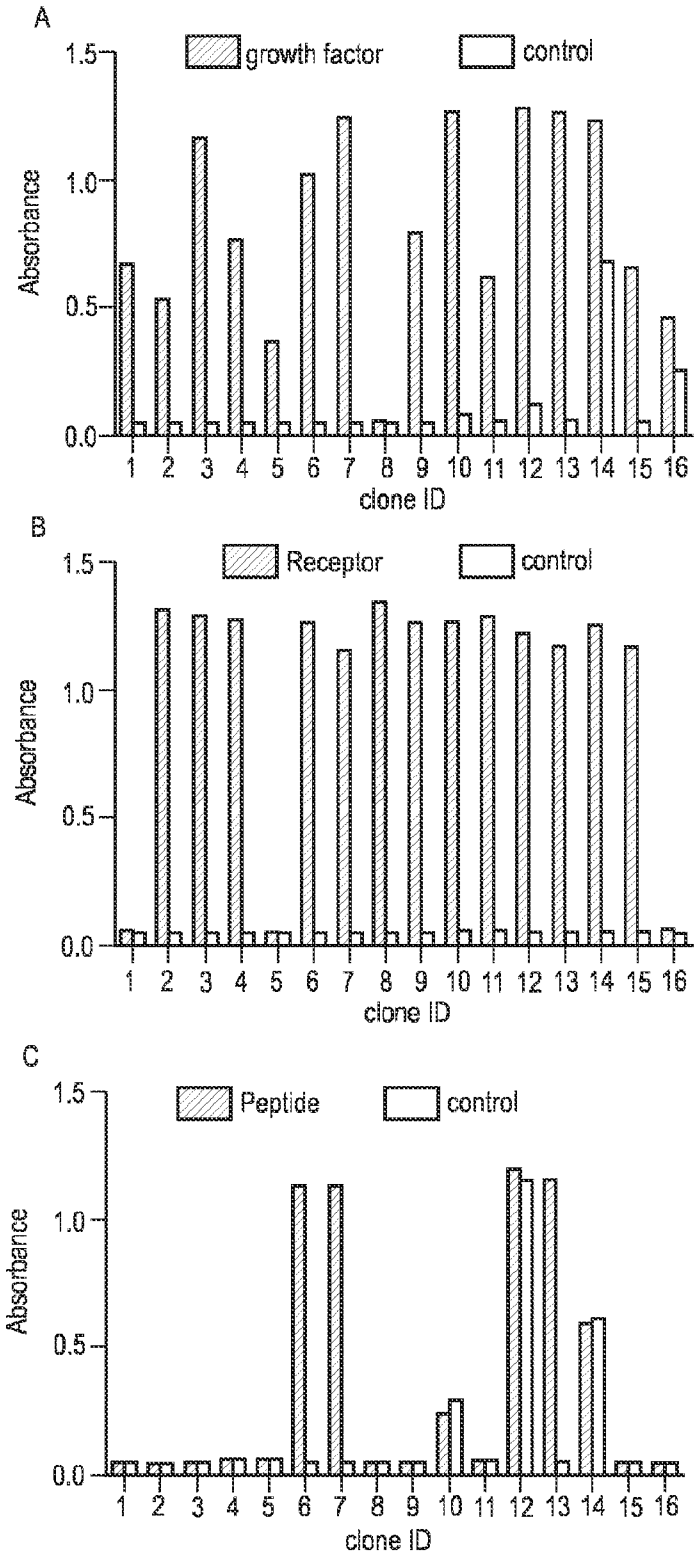

A
growth factor
control
Absorbance
1.5
1.0
0.5
0.0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
clone ID
B
Receptor
control
Absorbance
1.5
1.0
0.5
0.0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
clone ID
C
Peptide
control
Absorbance
1.5
1.0
0.5
0.0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
clone ID

Fig. 11

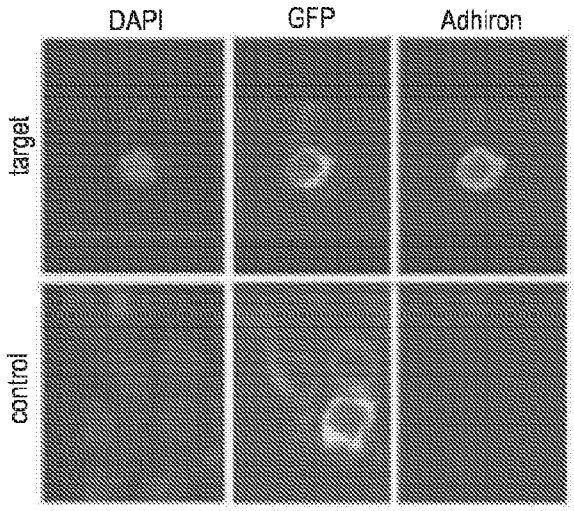

DAPI
GFP
Adhiron
target
control

Fig. 12

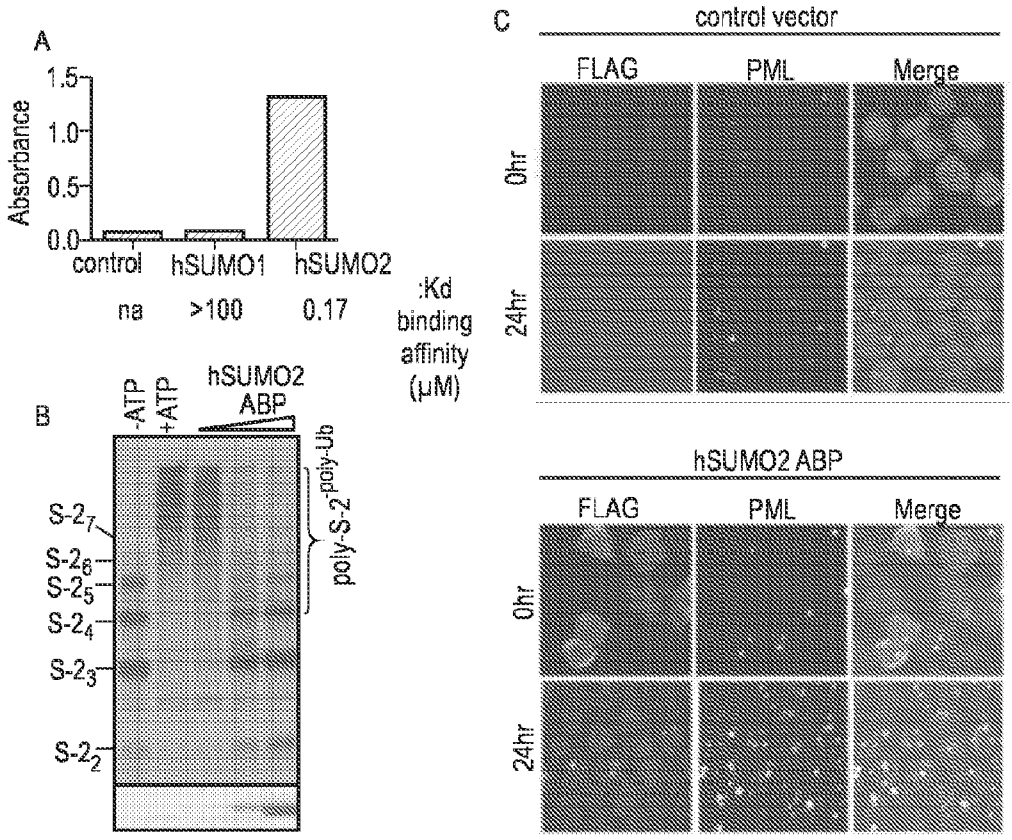

A
Absorbance
1.5
1.0
0.5
0.0
control
hSUMO1
hSUMO2
na
>100
0.17
:Kd binding affinity (μM)
B
-ATP
+ATP
hSUMO2 ABP
poly-S-2-poly-Ub
S-2$_7$
S-2$_6$
S-2$_5$
S-2$_4$
S-2$_3$
S-2$_2$
C
control vector
FLAG
PML
Merge
0hr
24hr
hSUMO2 ABP
FLAG
PML
Merge
0hr
24hr

Fig. 13

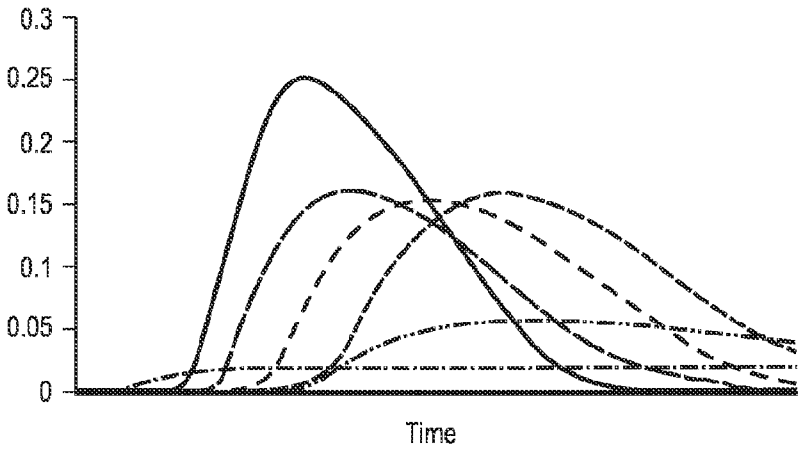

0.3
0.25
0.2
0.15
0.1
0.05
0
Time

Fig. 14

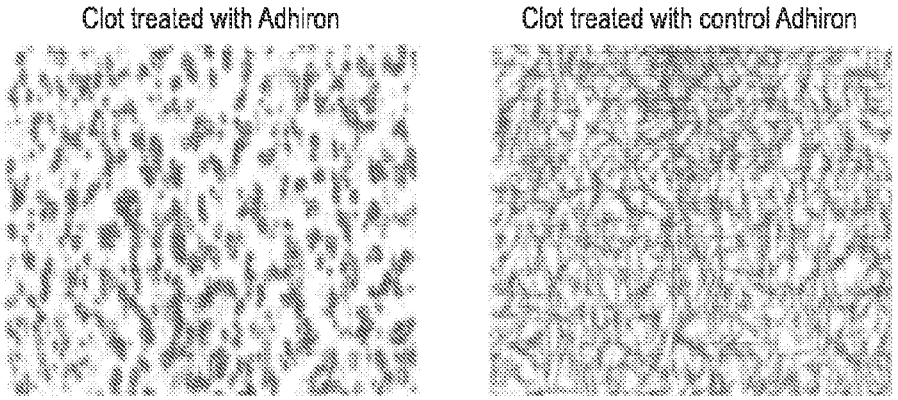

Clot treated with Adhiron
Clot treated with control Adhiron

Fig. 15

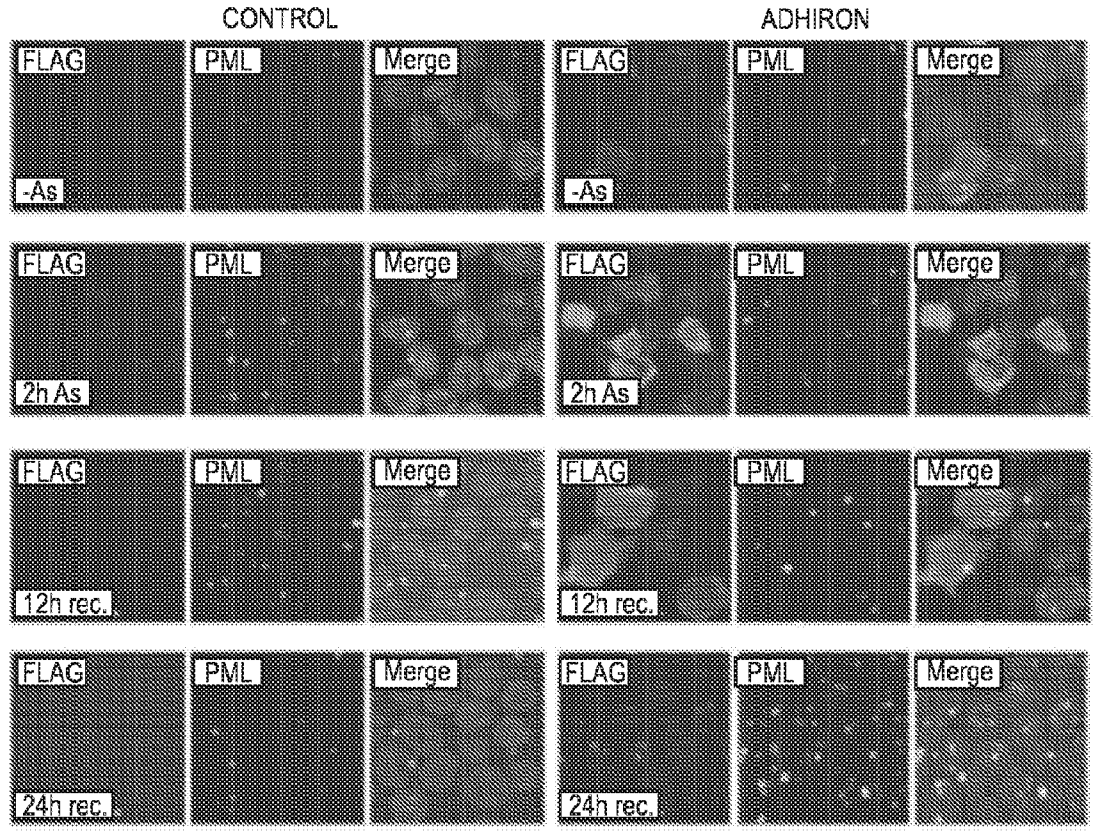


Fig. 16

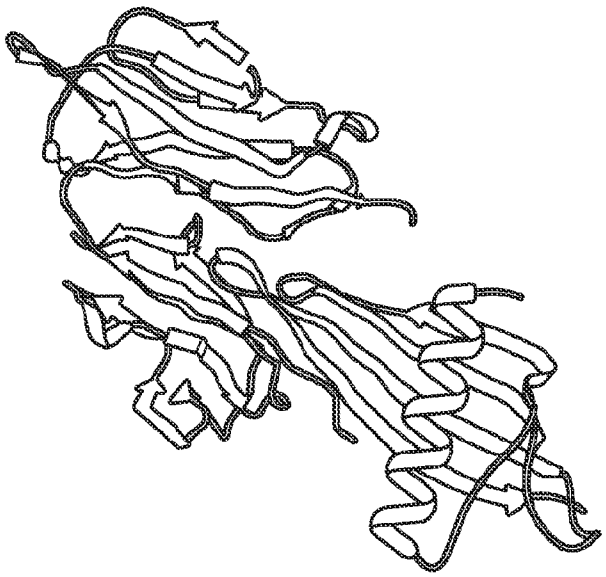


Fig. 17A

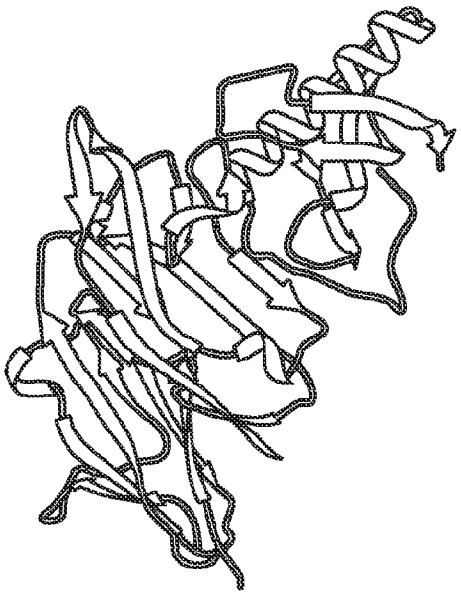


Fig. 17B

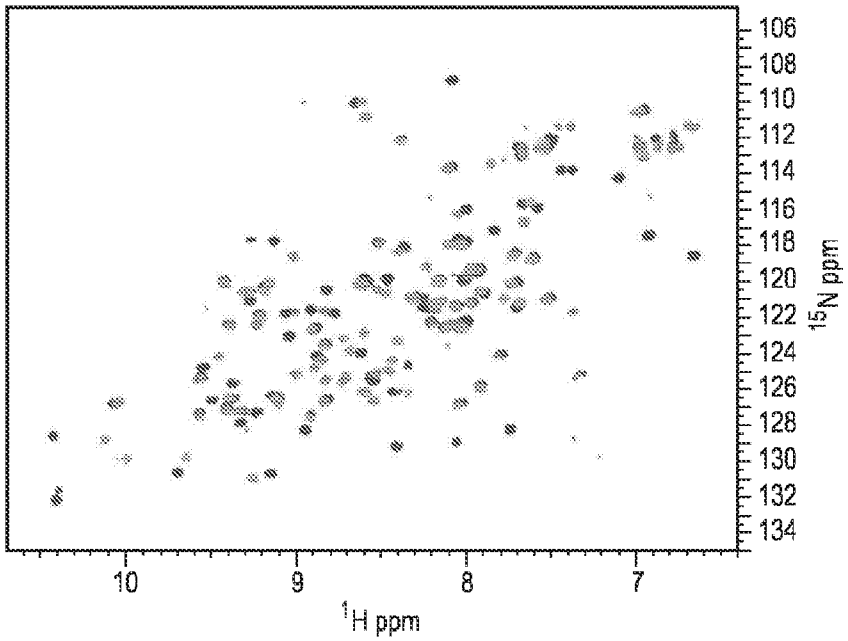


Fig. 18A

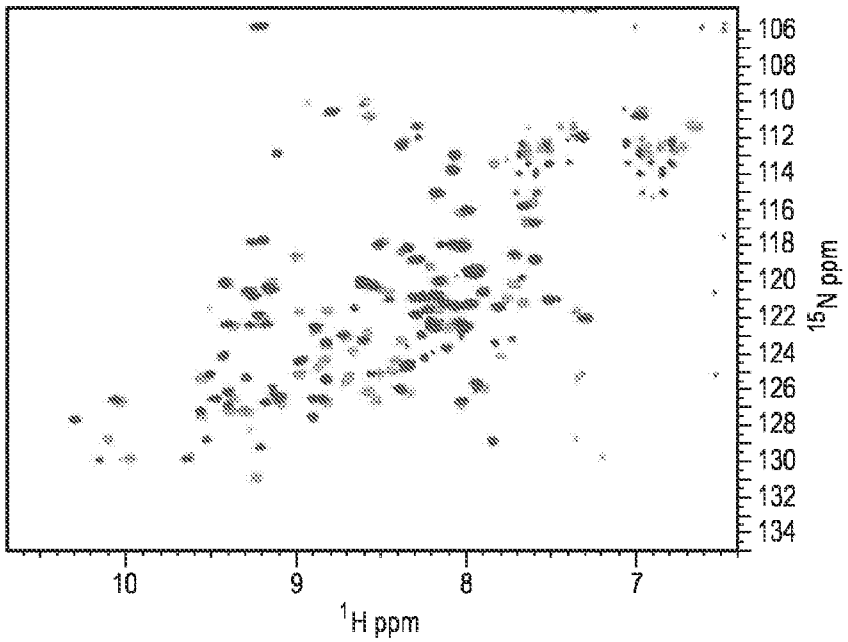


Fig. 18B

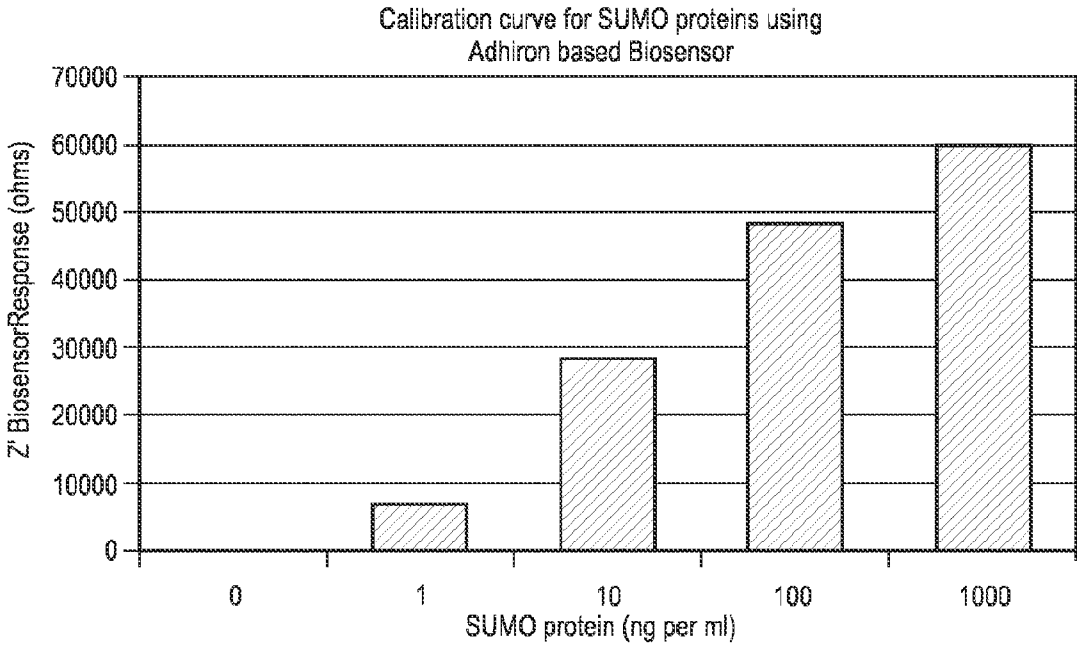

Calibration curve for SUMO proteins using
Adhiron based Biosensor
Z' BiosensorResponse (ohms)
70000
60000
50000
40000
30000
20000
10000
0
0
1
10
100
1000
SUMO protein (ng per ml)

Fig. 19

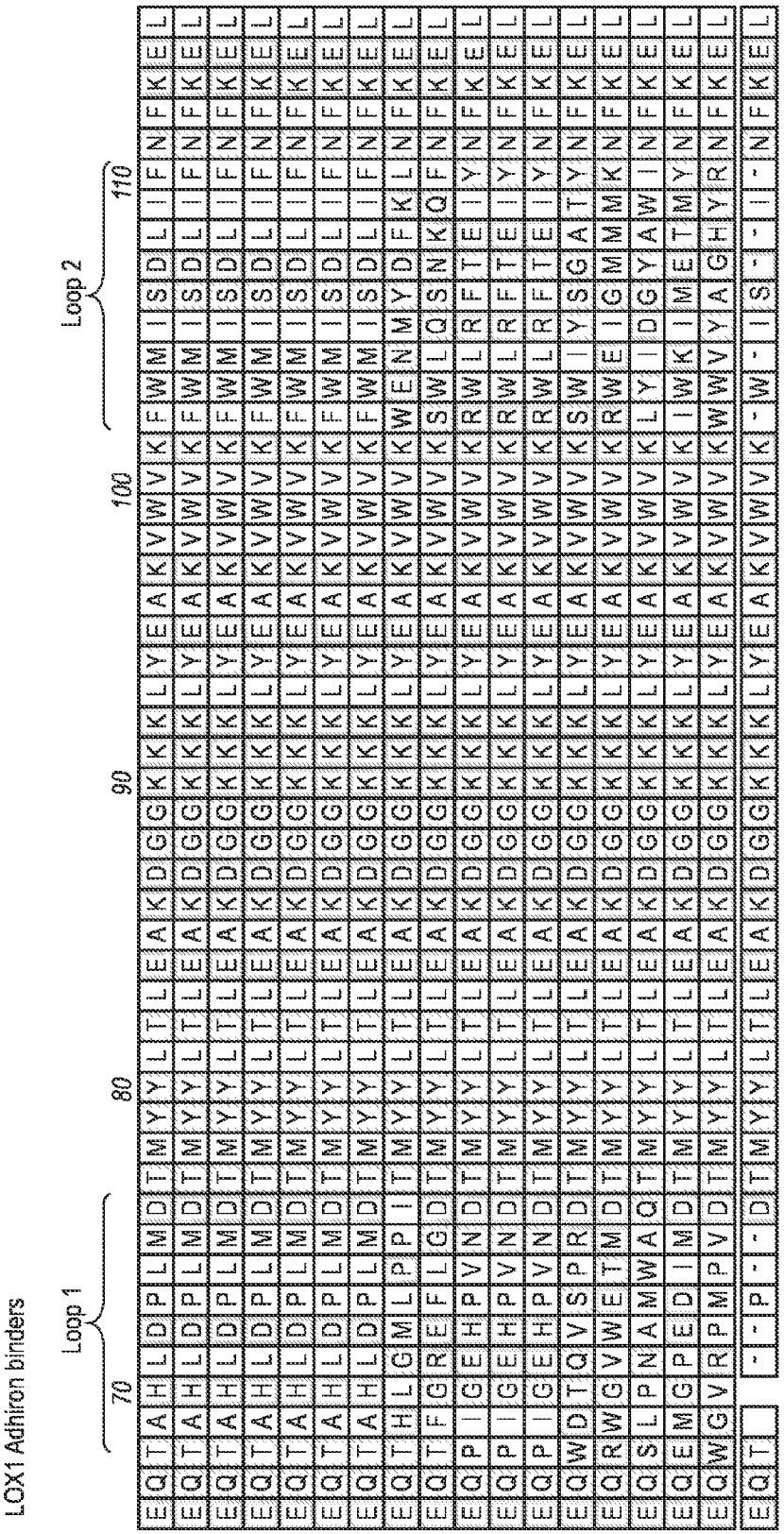


Fig. 20

16/21

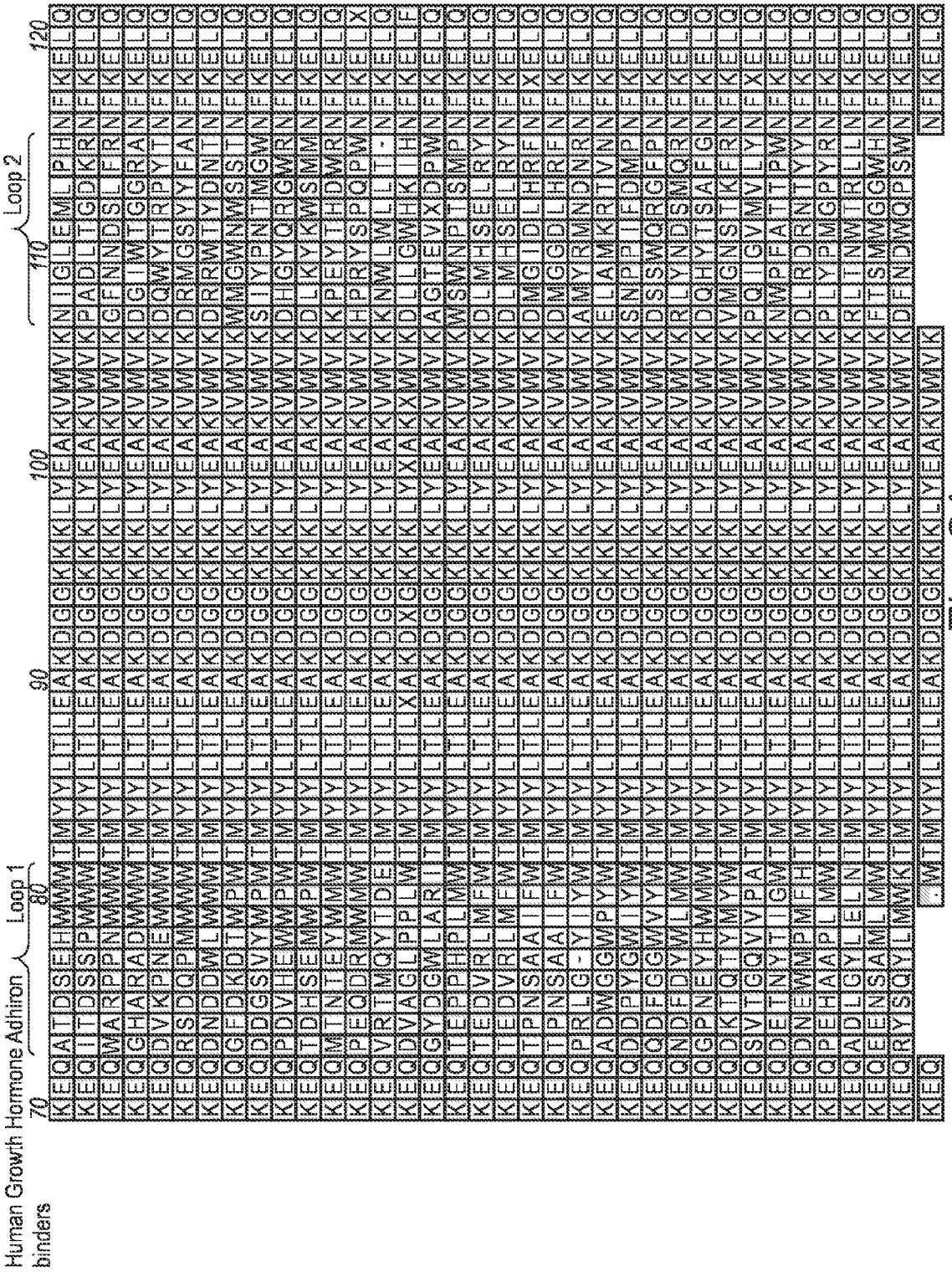

Human Growth Hormone Adhiron binders
Loop 1
Loop 2
70
80
90
100
110
120

Fig. 21

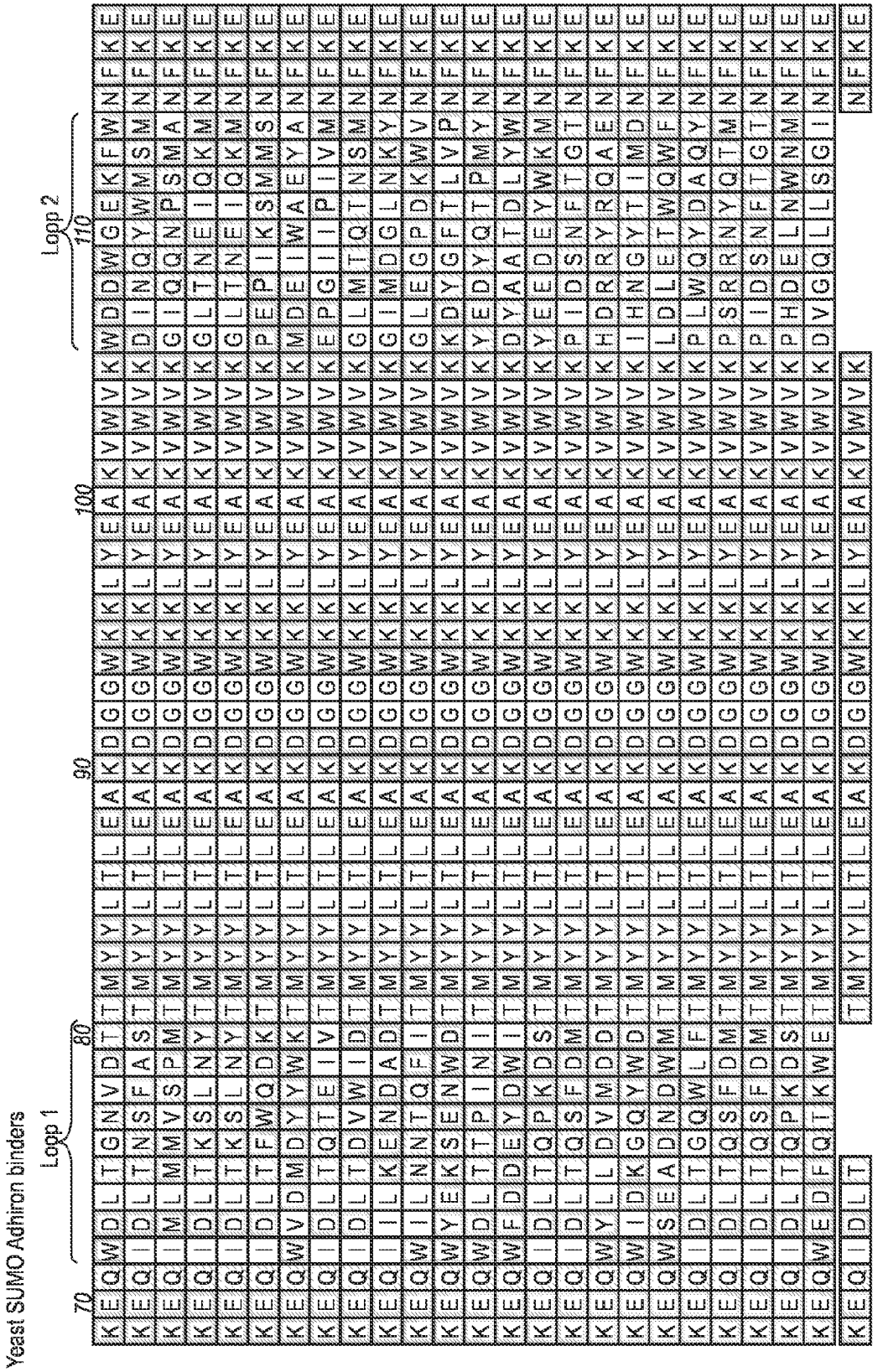


Fig. 22

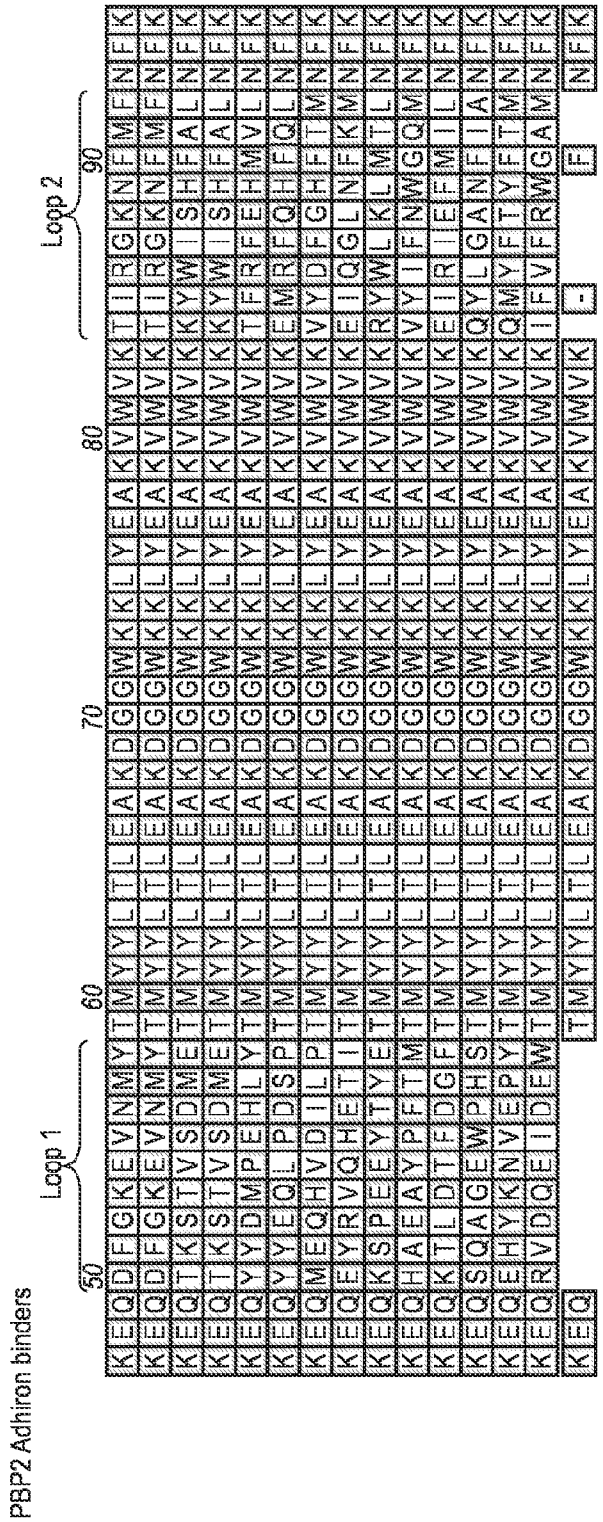

PBP2 Adhiron binders
Loop 1
Loop 2
50
60
70
80
90

Fig. 23

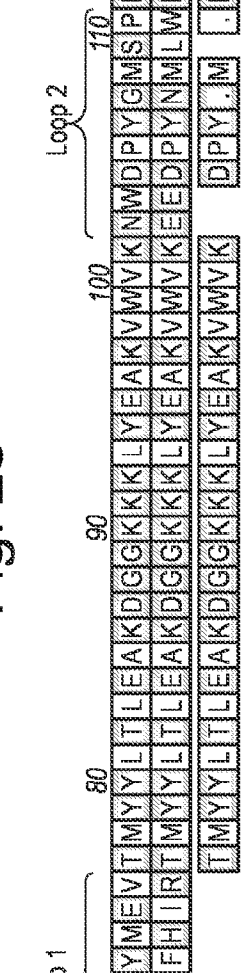

Peptide Adhiron binders
Loop 1
Loop 2
70
80
90
100
110
130

Fig. 24

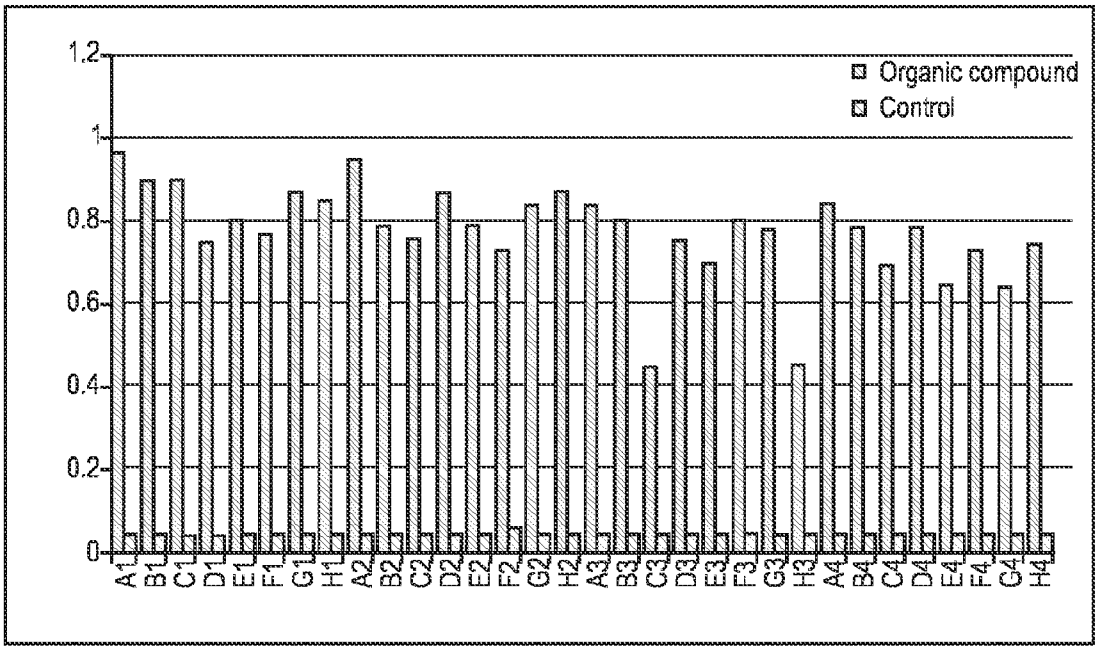

Organic compound
Control
1.2
1
0.8
0.6
0.4
0.2
0
A1 B1 C1 D1 E1 F1 G1 H1 A2 B2 C2 D2 E2 F2 G2 H2 A3 B3 C3 D3 E3 F3 G3 H3 A4 B4 C4 D4 E4 F4 G4 H4

Fig. 25

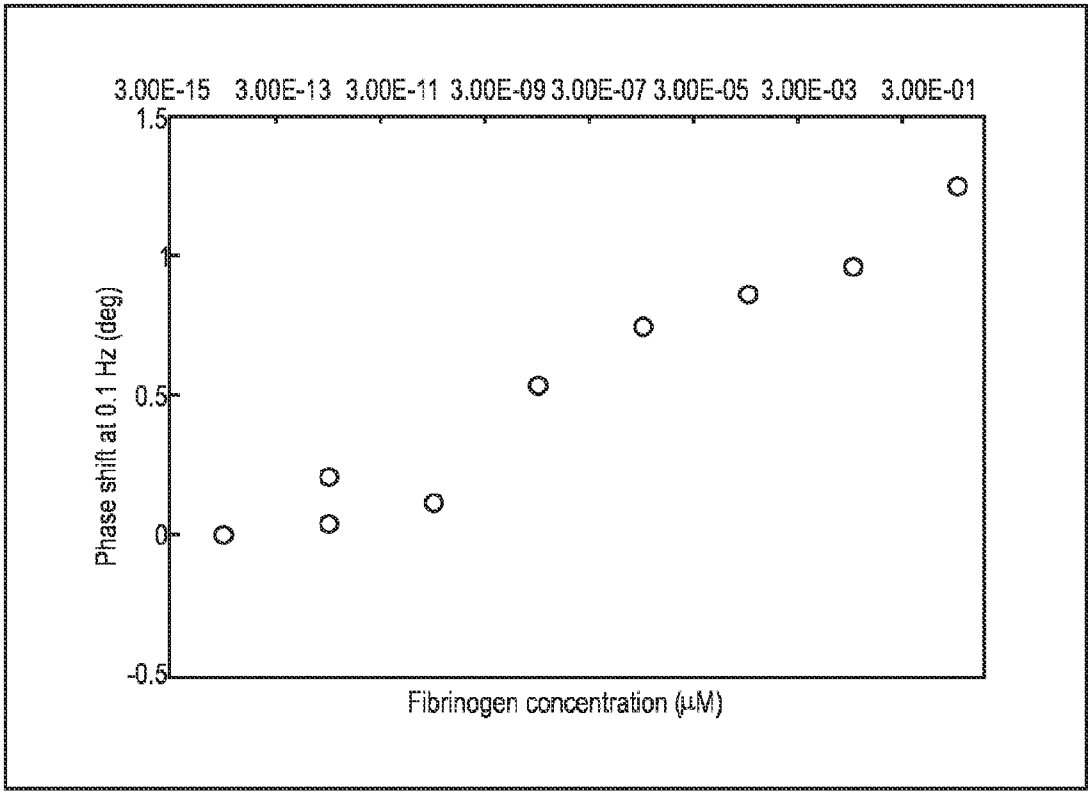

3.00E-15 3.00E-13 3.00E-11 3.00E-09 3.00E-07 3.00E-05 3.00E-03 3.00E-01
1.5
1
0.5
0
-0.5
Phase shift at 0.1 Hz (deg)
Fibrinogen concentration (μM)

Fig. 26

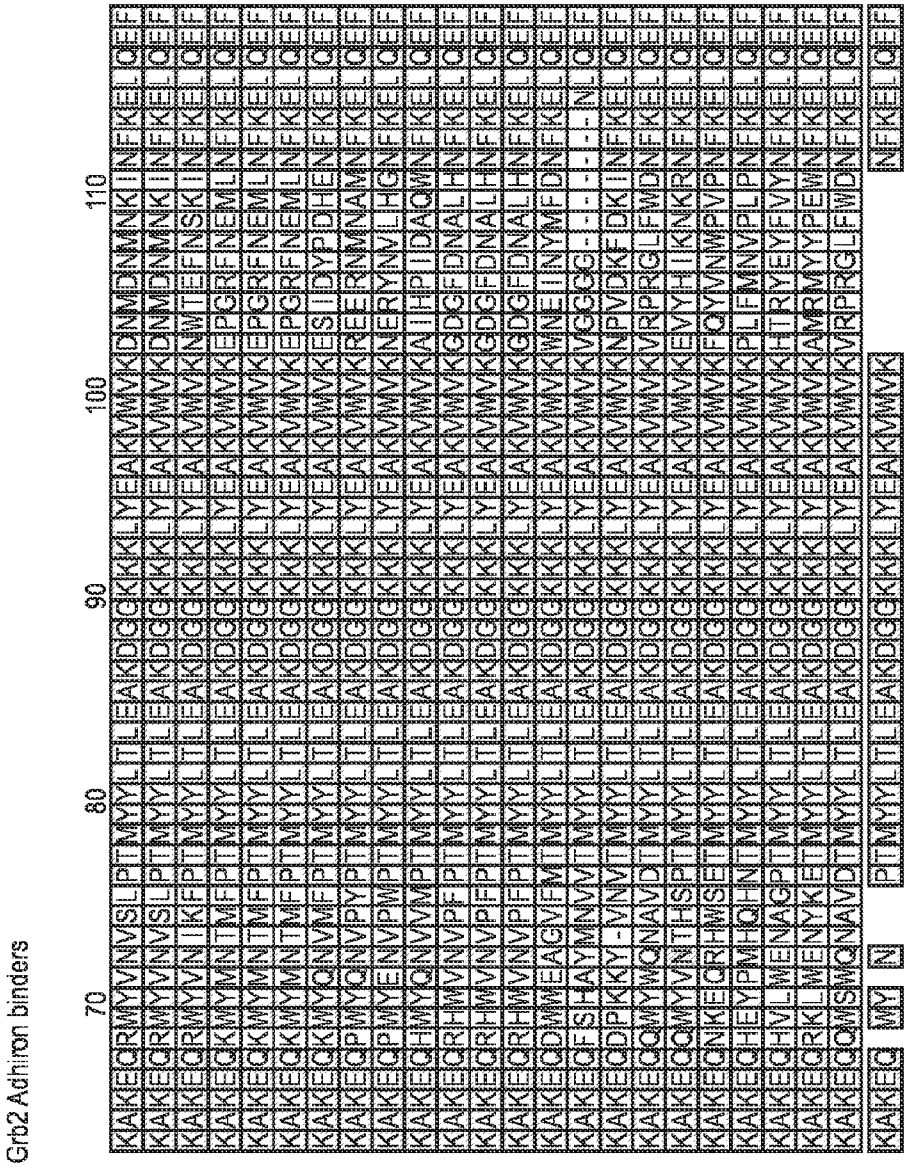


Fig. 27

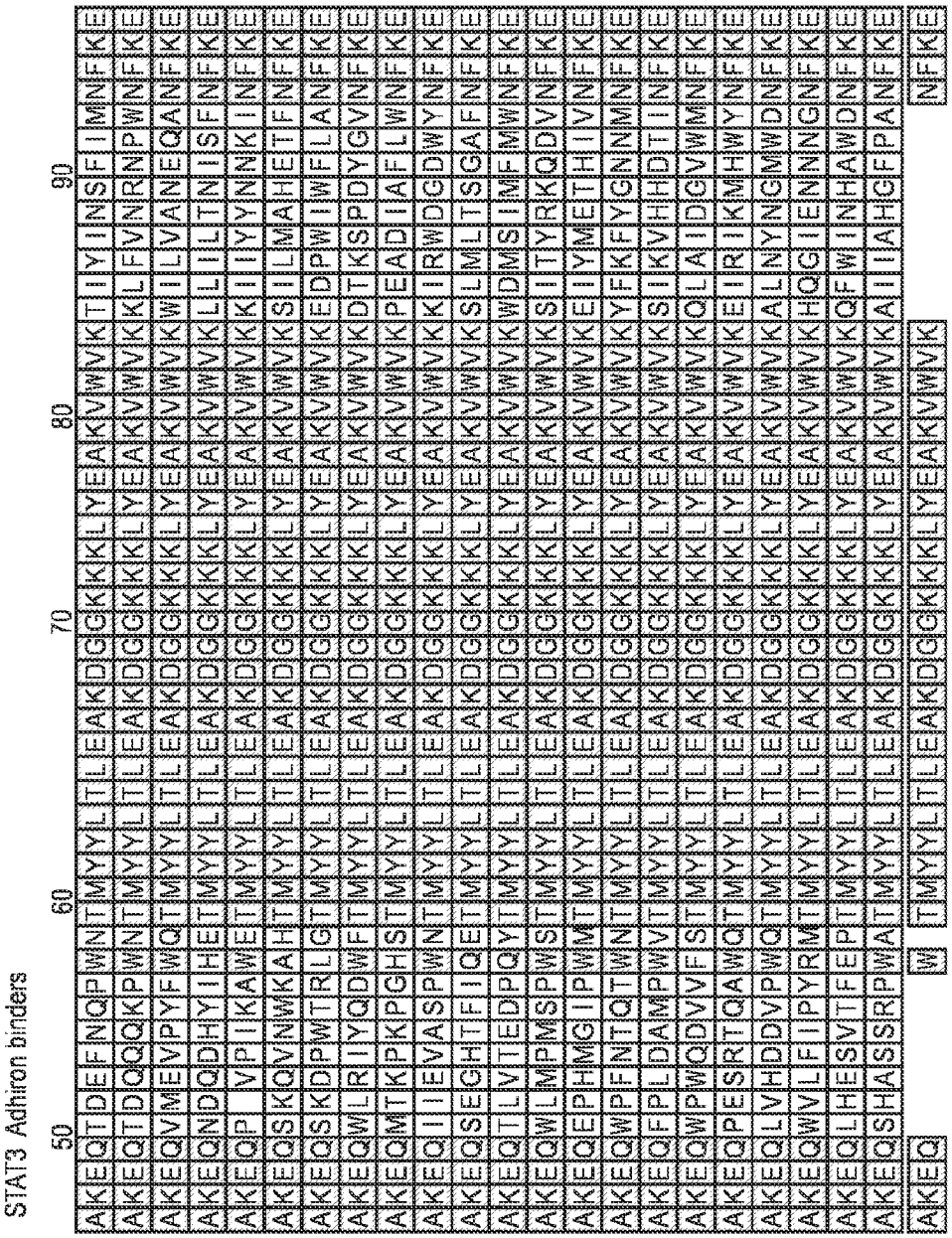


Fig. 28


SfiI
P1
P3
5' ATTAGCGGCCCAGCCGGCCATGGCCGCTCTGCTGGGTGGTGTTCGTGCAGTTCCGGGTAACG3'
5' CGAA
M A A L L G G V R A V P G N E N S L E I E E L A R F A V D E
ATGGCCGCTCTGCTGGGTGGTGTTCGTGCAGTTCCGGGTAACGAAAACTCCCTGGAAATCGAAGAACTGGCTCGTTTCGCTGTTGACGAA
TACCGGCGAGACGACCCACCACAAGCACGTCAAGGCCCATTGCTTTTGAGGGACCTTTAGCTTCTTGACCGAGCAAAGCGACAACTGCTT
3' GGCCCATTGCTTTTGAGGGACCTTTAGCTTCTTGACCGAGCAAAGCGACAACTGCTT
P3
P2
P5
CACAACAAAAAAGAAAACGCTCTGCTGGAATTCGTTCGTGTTGTTAAAGCTAAAGAACAGGTTGTTGC3'
5' GGT
H N K K E N A L E F V R V V K A K E Q V V A G T M Y Y L T L E A K D G G
CACAACAAAAAAGAAAACGCTCTGCTGGAATTCGTTCGTGTTGTTAAAGCTAAAGAACAGGTTGTTGCTGGTACCATGTACTACCTGACCCTGGAAGCTAAAGACGGTGGT
GTGTTGTTTTTTCTTTTGCGAGACGACCTTAAGCAAGCACAACAATTTCGATTTCTTGTCCAACAACGACCATGGTACATGATGGACTGGGACCTTCGATTTCTGCCACCA
GTGTTG5'
P2
P5
3' GTCCAACAACGACCATGGTACATGATGGACTGGGACCTTCGATTTCTGCCACCA
P4
AAAAAGAAACTGTACGAAGCTAAAGTTTGGGTTAAACCGTGGGAAAACTTCAAAGAACTGC3'
K K K L Y E A K V W V K P W E T F K E L Q E F K P V G D A
AAAAAGAAACTGTACGAAGCTAAAGTTTGGGTTAAACCGTGGGAAAACTTCAAAGAACTGCAGGAATTCAAACCGGTTGGTGACGCT
TTTTTCTTTGACATGCTTCGATTTCAAACCCAATTTGGCACCCTTTTGAAGTTTCTTGACGTCCTTAAGTTTGGCCAACCACTGCGA
TTTTTCTTTGAC5'
GTTTCTTGACGTCCTTAAGTTTGGCCAACCACTGCGACCGCCGGCGTTCATGTCGC
P4
P6
NotI
A
M13R
pelB signal sequence
SfiI
NotI
His tag region
PIII coding region
pDHisII
P10
B