



(43) International Publication Date
26 January 2017 (26.01.2017)

(51) International Patent Classification:

C07K 16/32 (2006.01) C12N 15/62 (2006.01)
C07K 14/47 (2006.01)

(21) International Application Number:

PCT/EP2016/067216

(22) International Filing Date:

19 July 2016 (19.07.2016)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

EP15177545.9 20 July 2015 (20.07.2015) EP

(71) Applicant: SCIL PROTEINS GMBH [DE/DE]; Heinrich-Damerow-Str. 1, 06120 Halle/Saale (DE).

(72) Inventors: FIEDLER, Erik; c/o Scil Proteins GmbH, Heinrich Damerow Str. 1, 06120 Halle/Saale (DE). MEY-SING, Maren; c/o Scil Proteins GmbH, Heinrich-Damerow-Str. 1, 06120 Halle/Saale (DE).

(74) Agent: HAUCKE, Kerstin; Scil Proteins GmbH, Heinrich Damerow Str. 1, 06120 Halle/Saale (DE).

(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, 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.

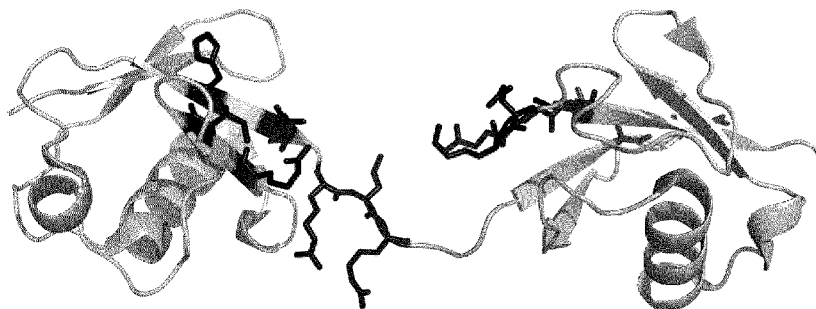
(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, 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))

(54) Title: NOVEL BINDING PROTEINS BASED ON DI-UBIQUITIN MUTEINS AND METHODS FOR GENERATION

FIG. 1



(57) Abstract: The present invention refers to new binding proteins based on di-ubiquitin modified in 12, 13, or 14 positions selected from positions R42, I44, H68, V70, R72, L73, R74, K82, L84, Q138, K139, E140, S141, and T142 of di-ubiquitin and being able to bind specifically with high affinity to selected targets. Furthermore, the invention provides a method for the generation of said binding proteins. In addition, libraries encoding for said binding proteins based on di-ubiquitin muteins modified in 12, 13, or 14 positions selected from R42, I44, H68, V70, R72, L73, R74, K82, L84, Q138, K139, E140, S141, and T142 are provided.

NOVEL BINDING PROTEINS BASED ON DI-UBIQUITIN MUTEINS AND METHODS FOR GENERATION**FIELD OF THE INVENTION**

The present invention refers to new binding proteins based on di-ubiquitin modified in 12, 13, or 14 positions selected from R42, I44, H68, V70, R72, L73, R74, K82, L84, Q138, K139, E140, S141, and T142 and being able to bind specifically with high affinity to selected targets. The binding protein is based on two ubiquitin muteins moieties directly linked without any linker. Furthermore, the invention provides a method for the generation of said binding proteins. In addition, libraries encoding for said binding proteins based on di-ubiquitin muteins with modifications in 12, 13, or 14 positions selected from R42, I44, H68, V70, R72, L73, R74, K82, L84, Q138, K139, E140, S141, and T142 are provided.

BACKGROUND OF THE INVENTION

Non-immunoglobulin based binding agents can be beneficially used in the medical fields of diagnosis, prophylaxis and treatment of diseases. A solution to the disadvantages resulting from antibodies in diagnosis, prophylaxis and treatment of diseases is to provide polypeptides with comparable or even better affinity and specificity towards the specific targets combined with smaller molecular size enabling an improved tissue penetration and thus having better biodistribution properties.

Among non-immunoglobulin-derived small proteins, molecules based on modified ubiquitin are particularly interesting because these molecules promise alternative therapeutic and diagnostic possibilities compared to antibodies. Ubiquitin is a highly conserved, small, single domain protein present in all known eukaryotic cells and is 100 % conserved amongst all vertebrates. In addition, ubiquitin naturally occurs in serum lowering the immunogenic potential. This facilitates preclinical development in different species required for toxicological and efficacy studies.

Target-specific ubiquitin muteins are known as Affilin® (registered trademark of Scil Proteins GmbH) proteins(see, for example, EP1626985B1, EP2379581B1, EP2094845B1, and WO2012/172055). Affilin proteins are engineered to generate *de novo* binding affinity towards desired targets making them ideal for applications that demand targeting and capturing of biomolecules. In addition, Affilin-based molecules can be conjugated with other functional molecules for targeting approaches as well as chemical coupling for half-life modulation (see, e.g. WO2012/172058, WO2014/094799).

It is an objective of the present invention to provide novel binding proteins and methods for the generation and identification of di-ubiquitin muteins with high binding capabilities to a target. It is a further object of the present invention to provide libraries for the identification of novel binding proteins based on di-ubiquitin muteins.

The above-described objectives and advantages are achieved by the subject-matters of the enclosed independent claims. Preferred embodiments of the invention are included in the dependent claims as well as in the following description, examples and figures. The above overview does not necessarily describe all problems solved by the present invention.

SUMMARY OF THE INVENTION

An aspect of the invention relates to a method for the generation of a binding protein which binds to a given target, preferably a protein or peptide or domain of a protein, with a dissociation constant (K_D) of at least 1000 nM, said method comprising subjecting nucleotide triplets of a nucleic acid molecule of di-ubiquitin (SEQ ID NO: 4) to mutagenesis in at least 12, 13, or 14 positions coding for amino acid positions R42, I44, H68, V70, R72, L73,

R74, K82, L84, Q138, K139, E140, S141, and T142 of the linear polypeptide sequence of di-ubiquitin, resulting in one or more di-ubiquitin mutein nucleic acid molecule(s), expressing the one or more di-ubiquitin mutein nucleic acid molecule(s) obtained in (b) in a suitable expression system, enriching the one or more di-ubiquitin mutein(s) having a dissociation constant (K_D) of at least 1000 nM for a given target by means of selection and/or isolation, wherein said binding protein has between 85 % and 92 % sequence identity to di-ubiquitin (SEQ ID NO: 4), and wherein the affinity to the target is at least 10fold higher than for di-ubiquitin.

A further aspect relates to a nucleic acid library encoding muteins of di-ubiquitin, wherein the amino acid residues selected from 12, 13, or 14 positions of sequence positions R42, I44, H68, V70, R72, L73, R74, K82, L84, Q138, K139, E140, S141, and T142 of di-ubiquitin are substituted in the muteins.

Another aspect relates to a method for identifying a binding protein with the affinity to the target of at least 10fold higher than for di-ubiquitin comprising screening said library to identify a polypeptide that binds to a given target.

In another aspect, the present invention relates to a binding protein with an affinity to a target comprising an amino acid sequence wherein at least 12, 13, or 14 amino acids selected from positions R42, I44, H68, V70, R72, L73, R74, K82, L84, Q138, K139, E140, S141, and T142 of di-ubiquitin (for example, SEQ ID NO: 4 or SEQ ID NO: 43) are substituted and wherein the binding protein has between 85 % and 92 % sequence identity to di-ubiquitin (for example, SEQ ID NO: 4 or a sequence with at least 95 % identity, such as SEQ ID NO: 43) and wherein the affinity to the target is at least 10fold higher than for di-ubiquitin (for example, SEQ ID NO: 4 or a sequence with at least 95 % identity, such as SEQ ID NO: 43). The binding protein optionally comprises 1, 2, 3, 4, 5, or 6 further amino acid substitutions of di-ubiquitin. Further the binding protein may have a specific binding affinity (K_D) of at least 1000 nM, preferably at least 500 nM, preferably at least 100 nM, preferably at least 10 nM, to a target, preferably a peptide, a protein, or a domain of a protein, wherein the affinity to said target is at least 10fold higher than for di-ubiquitin.

In a further aspect the present invention relates to a binding protein comprising at least one additional molecule, preferably selected from at least one member of the groups (i), (ii) and (iii) consisting of (i) a pharmacokinetic moiety modulating pharmacokinetic behavior selected from a polyethylene glycol (PEG), a human serum albumin (HSA), an albumin-binding peptide, or an immunoglobulin or immunoglobulin fragments, a polysaccharide, and, (ii) a therapeutically active component, optionally selected from a monoclonal antibody or a fragment, cytokine, a chemokine, a cytotoxic compound, an enzyme, or derivatives thereof, or a radionuclide, and (iii) a diagnostic component, optionally selected from a fluorescent compound, a photosensitizer, a tag, an enzyme, or a radionuclide.

The present invention also provides, in further aspects, a nucleic acid or nucleic acids encoding the binding protein comprising or essentially consisting or consisting of a mutein of the present invention, as well as a vector or vectors comprising said nucleic acid or nucleic acids, and a host cell or host cells comprising said vector or vectors.

This summary of the invention does not necessarily describe all features of the present invention. Other embodiments will become apparent from a review of the ensuing detailed description.

Brief description of the Figures

The Figures show:

FIG. 1. Model of di-ubiquitin. Amino acid residues selected for generation of the SPIF library are shown in black.

FIG. 2. Biochemical properties of the SPIF library

FIG. 2 A compares the solubility of proteins encoded by five different libraries based on di-ubiquitin muteins directly connected (i.e. without a peptide linker between two ubiquitin moieties) (referred to as SPIF, SPKK,

SPGO, SPOO) or joined via a peptide linker (SPVF). Libraries SPIF, SPKK, SPGO, and SPOO have at least 14 randomized positions starting between positions 42-74 and 122-142 of di-ubiquitin (SEQ ID NO: 4; see Examples for more details), except SPVF (randomized at positions 6-8, 62-66 at both ubiquitin moieties (SEQ ID NO: 1) linked via an amino-acid peptide linker). The columns show the values for more than 70 % solubility of proteins encoded by the library. Surprisingly, the proteins encoded by SPIF library are highly soluble compared to all other libraries tested. The level of aggregation of proteins is very low (SPIF) compared to proteins encoded by other libraries.

FIG. 2 B compares the expression of the proteins encoded by the five different libraries mentioned above. Surprisingly, the proteins encoded by SPIF library show strong total expression values.

FIG. 3 shows binding molecules of the invention.

FIG. 3 A lists positions of di-ubiquitin (SEQ ID NO: 4) that are substituted in order to generate a Her2 binding protein. In the first row, the corresponding amino acid position is given. All Her2 binding proteins (for example, SEQ ID NOs: 5-38) are modified at least in 12, 13, or 14 positions selected from positions R42, I44, H68, V70, R72, L73, R74, K82, L84, Q138, K139, E140, S141, and T142 of SEQ ID NO: 4. A "." in the table refers to a wildtype position (unchanged); for example, in SEQ ID Nos: 6, 31, 33, 34, 35, 36, 37.

FIG. 3 B shows the same amino acid exchanges as FIG. 3 A, however, the exchanges are translated according to the following code which groups amino acids with similar biophysical properties. A waved line "~" is the symbol for polar amino acids (T, S, N, or Q) "H" is the symbol for hydrophobic amino acids (e.g. A, M, L, V, I) "O" is the symbol for aromatic amino acids (e.g. F, W, Y), "+" the symbol for basic amino acids (e.g. K, R, H), "-" the symbol for acidic amino acids (e.g. D, E), and "G" corresponds to Glycine.

FIG. 3 C. Further modifications are listed; all Her2 binding proteins have 0, 1, 2, 3, 4, 5, or 6 further modifications in addition to at least 12 positions selected from positions 42, 44, 68, 70, 72, 73, 74, 82, 84, 138, 139, 140, 141, and 142 of SEQ ID NO: 4.

FIG. 3 D shows the same amino acid exchanges as FIG. 3 C, however, the exchanges are translated according to the code which groups amino acids with similar biophysical properties as described in FIG. 3 B.

FIG. 4 shows results of an analysis of binding proteins for protein targets of di-ubiquitin muteins via label-free interaction assays using SPR. Different concentrations of Affilin proteins were analyzed for binding to protein targets of di-ubiquitin (SEQ ID NO: 4) immobilized on a chip (Biacore) to analyze the interaction between Affilin and protein targets of di-ubiquitin.

FIG. 4 A shows the binding kinetics of Affilin-142628 to Her2. Concentrations: 0 nM, 0.14 nM, 0.41 nM, 1.23 nM, 3.70 nM, 11.11 nM, and 33.33 nM.

FIG. 4 B shows the binding kinetics of Affilin-143488 to EpCAM. Concentrations were: Long dash (15.6 nM), medium dash (31.25 nM), short dash (6.5 nM), dotted (125 nM), dash dot (250 nM), dash dot dot (500 nM) and medium medium (1000 nM). Following a 1:1 binding mode (Langmuir) a K_D of 150 nM vs epCAM was determined.

FIG. 4 C shows the binding kinetics of Affilin-143283 to EpCAM Concentrations were: Long dash (3.13 nM), medium dash (6.25 nM), solid (12.5 nM), short dash (25 nM), dotted (50 nM), dash dot (100 nM) and dash dot dot (200 nM). Following a 1:1 binding mode (Langmuir) a K_D of 32 nM vs epCAM was determined.

FIG. 4 D shows the binding kinetics of Affilin-144160 to CD3. Concentrations were: Long dash (2 nM), medium dash (4 nM), short dash (8 nM), dotted (16 nM), dash dot (32 nM), dash dot dot (64 nM), medium medium (125 nM), short short short (250 nM) and short short (500 nM). Following a 1:1 binding mode (Langmuir) a K_D of 46 nM vs CD3 was determined.

FIG. 5 shows a functional characterization of binding molecules of the invention.

FIG. 5 A shows the affinity of Affilin-142265 to EGFR. The figure shows binding of Affilin-142265 (dark grey; black line) to exogenously EGFR expressing CHO-K1 cells (light grey) as determined by FACS analysis.

FIG. 5 B shows the high affinity of Her2 binding proteins to exogenously Her2 expressing CHO K1 cells at concentrations of 50 nM, 5 nM, and 0.5 nM. No binding was observed for di-ubiquitin or on control cells (empty vector control cells)

FIG. 5 C shows the affinity of Her2 binding proteins to exogenously Her2 expressing SkBr cells as determined by FACS analysis. Cellular Her2 binding at 50 nM was confirmed for Affilin-proteins. No binding was observed to non-expressing HEK/293 control cells or for di-ubiquitin (139090).

FIG. 6. Binding analysis of Her2 binding proteins on Her2-overexpressing SkBr3-cells by immunofluorescence cell staining. All Affilin proteins show a strong binding of 50 nM on the SkBr3-cells, whereas di-ubiquitin shows no binding.

FIG. 7. Binding analysis of Her2 binding proteins on SKOV-3 xenograft tumor tissue cells (immunohistological staining). The figure confirms that Her2 binding proteins bind to SKOV-3 tumor tissue whereas di-ubiquitin 139090 shows no binding on SKOV-3 tissue.

FIG. 8 confirms that Affilin proteins bind to extracellular EGFR expressed on tumor cells. Shown are immunofluorescence images of EGFR expressing A431 tumor cells. The staining of A431 tumor cells expressing EGFR confirms binding of Affilin-142265 whereas di-ubiquitin (139090) shows no binding.

Detailed Description of the Invention

Before the present invention is described in more detail below, it is to be understood that this invention is not limited to the particular methodology, protocols and reagents described herein as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims. Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs.

Preferably, the terms used herein are defined as described in "A multilingual glossary of biotechnological terms: (IUPAC Recommendations)", Leuenberger, H.G.W, Nagel, B. and Kölbl, H. eds. (1995), Helvetica Chimica Acta, CH-4010 Basel, Switzerland).

Throughout this specification and the claims which follow, unless the context requires otherwise, the word "comprise", and variants such as "comprises" and "comprising", will be understood to imply the inclusion of a stated integer or step or group of integers or steps but not the exclusion of any other integer or step or group of integers or steps.

Several documents (for example: patents, patent applications, scientific publications, manufacturer's specifications, instructions, GenBank Accession Number sequence submissions etc.) are cited throughout the text of this application. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention. Some of the documents cited herein are characterized as being "incorporated by reference". In the event of a conflict between the definitions or teachings of such incorporated references and definitions or teachings recited in the present specification, the text of the present specification takes precedence.

All sequences referred to herein are disclosed in the attached sequence listing that, with its whole content and disclosure, is a part of this specification.

Definitions of important terms used in the application

The terms "protein" and "polypeptide" refer to any chain of two or more amino acids linked by peptide bonds, and does not refer to a specific length of the product. Thus, "peptides", "protein", "amino acid chain," or any other term

used to refer to a chain of two or more amino acids, are included within the definition of "polypeptide," and the term "polypeptide" may be used instead of, or interchangeably with any of these terms. The term "polypeptide" is also intended to refer to the products of post-translational modifications of the polypeptide, including without limitation glycosylation, acetylation, phosphorylation, amidation, proteolytic cleavage, modifications by non-naturally occurring amino acids and similar modifications which are well known in the art. Thus, binding proteins comprising two or more protein moieties also fall under the definition of the term "protein" or "polypeptides".

The term "binding protein" as used herein refers to a protein which either comprises or essentially consists of or consists of two ubiquitin moieties (Affilin) with specific binding affinity (K_D) of at least 1000 nM, preferably at least 500 nM, preferably at least 100 nM, preferably at least 10 nM, for example, when determined by surface plasmon resonance, to a target wherein the affinity to the target is at least 10fold higher than for di-ubiquitin (for example, SEQ ID NO: 4 or at least 95 % identical sequences such as SEQ ID NO: 43).

The terms "protein capable of binding" or "binding protein" or "binding affinity for" according to this invention refer to a protein comprising a binding capability to a defined target. The term "Her2 binding protein" refers to a protein with high affinity binding capability to Her2, the term "EGFR binding protein" refers to a protein with high affinity binding capability to EGFR, the term "CD3 binding protein" refers to a protein with high affinity binding capability to CD3, and the term "epCAM binding protein" refers to a protein with high affinity binding capability to epCAM.

In the present specification, the terms "target", "antigen", "target antigen", and "binding partner" are all used synonymously and can be exchanged. Preferably the target is one of the targets defined herein below. The term "antigen", as used herein, is to be interpreted in a broad sense and includes any target moiety that is bound by the binding moieties of the binding proteins of the present invention. Targets can be inorganic or organic substances. The term includes, for example, small molecules in body fluid such as drugs, toxins, autoantibodies, autoantigens, proteins, polypeptides, carbohydrates, nucleic acids and other molecules. Specific examples are a cancer target antigen, a receptor target antigen on immune cells, and a soluble target antigen selected e.g. from hormones, and cytokines. Summarizing, as target antigen according to the invention all biologically and medically active and relevant molecules can be employed. Possible binding partners will be described in the following by way of example. It should be noted, however, that a plurality of other possible targets can be added to this list.

The term "is not a natural binding target" as used herein refers to a target with affinity of at least 10fold higher to the binding protein of the invention than for di-ubiquitin. "not a natural binding target" further implies that the binding activity of the binding molecules of the invention is created *de novo*. In other words, a "natural binding target" would bind to unmodified ubiquitin (for example, SEQ ID NO: 1 or at least 95 % identical sequences such as SEQ ID NO: 2) or unmodified di-ubiquitin (for example, SEQ ID NO: 4 or at least 95 % identical sequences such as SEQ ID NO: 43) with at least 10fold lower affinity than the binding proteins of the invention.

The term "dissociation constant" or " K_D " defines the specific binding affinity. A high affinity corresponds to a low value of K_D . Thus, the expression "a K_D of at least e.g. 10^{-7} M" means a value of 10^{-7} M or lower (binding more tightly). 1×10^{-7} M corresponds to 100 nM. A value of 10^{-5} M and below down to 10^{-12} M can be considered as a quantifiable binding affinity. Depending on the application a value of 10^{-7} to 10^{-12} M is preferred for e.g. chromatographic applications or for e.g. diagnostic or therapeutic applications. In accordance with the invention the affinity for the target binding should be in the range of less than 1×10^{-6} M (1000 nM). Final target binding affinity should be ideally below 10^{-9} M (1 nM).

The term "ubiquitin" or "unmodified ubiquitin" refers to ubiquitin in accordance with SEQ ID NO: 1 (wild type ubiquitin), according to SEQ ID NO: 2 (point mutations in positions F45W, G75A, G76A which do not influence binding to a target), according to SEQ ID NO: 4 (point mutations for example in positions G75A, G76A, F121W, G151A, G152A; also referred to as "di-ubiquitin" herein; corresponds to clone 139090) according to SEQ ID NO: 43 (two wild type ubiquitin moieties directly fused together without additional peptide linker; also referred to as "di-ubiquitin" herein) or which do not influence binding to a target) and according to the following definition.

Particularly preferred are ubiquitins from mammals, e.g. humans, primates, pigs, and rodents. On the other hand, the ubiquitin origin is not relevant since according to the art all eukaryotic ubiquitins are highly conserved and the mammalian ubiquitins examined up to now are even identical with respect to their amino acid sequence. In addition, ubiquitin from any other eukaryotic source can be used. For instance ubiquitin of yeast differs only in

three amino acids from the wild-type human ubiquitin (SEQ ID NO: 1). The term "di-ubiquitin" refers to a linear protein wherein two ubiquitin moieties are directly fused to each other in head to tail orientation. The term "di-ubiquitin" refers to according to SEQ ID NO: 4 (point mutations in positions G75A, G76A, F121W, G151A, G152A; corresponds to clone 139090; two ubiquitin moieties directly fused together) or to proteins with at least 95 % amino acids identity to SEQ ID NO: 4, for example, according to SEQ ID NO: 43 (two wild type ubiquitin moieties directly fused together). A "di-ubiquitin mutein" according to this invention refers to a di-ubiquitin of SEQ ID NO: 43 or SEQ ID NO: 4 having modifications in at least 12, 13, or 14 amino acid positions, preferably in positions R42, I44, H68, V70, R72, L73, R74, K82, L84, Q138, K139, E140, S141, and T142, and is capable of binding to selected targets.

The terms "Affilin" and "ubiquitin mutein" and "modified ubiquitin" are all used synonymously and can be exchanged. The term "ubiquitin mutein" and "Affilin" and "modified ubiquitin" as used herein refers to derivatives of ubiquitin which differ from said unmodified ubiquitin or di-ubiquitin by amino acid exchanges, insertions, deletions or any combination thereof, provided that the modified ubiquitin or ubiquitin mutein or di-ubiquitin mutein has a specific binding affinity to a target which is at least 10fold lower or absent in unmodified ubiquitin or di-ubiquitin. This functional property of a di-ubiquitin mutein (Affilin) is a *de novo* created function.

The term "Affilin®" (registered trademark of Scil Proteins GmbH) refers to non-immunoglobulin derived binding proteins based on ubiquitin muteins. An Affilin is not a natural ubiquitin existing in or isolated from nature. The scope of the invention excludes unmodified ubiquitin. An Affilin molecule according to this invention comprises or essentially consists of two differently modified ubiquitin moieties linked together in a head-to-tail fusion (di-ubiquitin mutein). A "head-to-tail fusion" is to be understood as fusing two proteins together by connecting them in the direction (head) N-C-N-C- (tail) (tandem molecule), as described for example in EP2379581B1 which is incorporated herein by reference. The head part is designated as the first moiety and the tail part as the second moiety. In this head-to-tail fusion, the ubiquitin moieties may be connected directly without any linker, as shown in this invention. Alternatively, the fusion of ubiquitin moieties can be performed via linkers, for example, a polypeptide linker (e.g. for library SPVF, see Table 2).

The term "substitution" includes both conservative and non-conservative substitutions. "Conservative substitutions" may be made, for instance, on the basis of similarity in polarity, charge, size, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the amino acid residues involved. Amino acids can be grouped into the following standard amino acid groups: (1) hydrophobic side chains: Ala (A), Met (M), Leu (L), Val (V), Ile (I), (2) acidic polar side chain: Asp (D), Glu (E), (3) basic side chain polarity: Lys (K), Arg (R), His (H), (4) aromatic amino acids: Trp (W), Tyr (Y), Phe (F), (5) polar amino acids: Thr (T), Ser (S), Asn (N), Gln (Q), (6) residues that influence chain orientation: Gly (G), Pro (P), and (7) Cys (C). As used herein, "conservative substitutions" are defined as exchanges of an amino acid by another amino acid listed within the same group of the standard amino acid groups shown above. For example, the exchange of Asp by Glu retains one negative charge in the so modified polypeptide. In addition, Gly and Pro may be substituted for one another based on their ability to disrupt α -helices. Some preferred conservative substitutions within the above groups are exchanges within the following sub-groups: (i) Ala, Val, Leu and Ile; (ii) Ser and Thr; (iii) Asn and Gln; (iv) Lys and Arg; and (v) Tyr and Phe. Given the known genetic code, and recombinant and synthetic DNA techniques, the skilled scientist can readily construct DNAs encoding the conservative amino acid variants. As used herein, "non-conservative substitutions" or "non-conservative amino acid exchanges" are defined as exchanges of an amino acid by another amino acid listed in a different group of the standard amino acid groups (1) to (7) shown above.

The term "fused" means that the components (for example, two ubiquitin moieties or two ubiquitin mutants) are linked by peptide bonds, either directly or via peptide linkers.

The term "fusion protein" relates to a protein comprising at least a first protein joined genetically to at least a second protein. A fusion protein is created through joining of two or more genes that originally coded for separate proteins. Thus, a fusion protein may comprise a multimer of different or identical binding proteins which are expressed as a single, linear polypeptide. Such fusion proteins may further comprise additional domains that are not involved in binding of the target, such as but not limited to, for example, multimerization moieties, polypeptide tags, polypeptide linkers.

The term "conjugate" as used herein relates to a protein comprising or essentially consisting of at least a first protein attached chemically to other substances such as to a second protein or a non-proteinaceous moiety. The conjugation can be performed by means of organic synthesis or by use of enzymes including natural processes of enzymatic post-translational modifications. Examples for protein conjugates are glycoproteins (conjugated protein with carbohydrate component) or lipoproteins (conjugated protein with lipid component). The molecule can be attached e.g. at one or several sites through any form of a linker. Chemical coupling can be performed by chemistry well known to someone skilled in the art, including substitution (e.g. N-succinimidyl chemistry), addition or cycloaddition (e.g. maleimide chemistry or click chemistry) or oxidation chemistry (e.g. disulfide formation). Some examples of non-proteinaceous polymer molecules which are chemically attached to protein of the invention are hydroxyethyl starch, polyethylene glycol, polypropylene glycol, dendritic polymers, or polyoxyalkylene and others.

A fusion protein or protein conjugate may further comprise one or more reactive groups or peptidic or non-peptidic moieties such as targets or therapeutically or diagnostically relevant molecules such as radionuclides or toxins. It may also comprise small organic or non-amino acid based compounds, e.g. a sugar, oligo- or polysaccharide, fatty acid, etc. Methods for attaching a protein of interest to such non-proteinaceous components are well known in the art, and are thus not described in further detail here.

The term "amino acid sequence identity" refers to a quantitative comparison of the identity (or differences) of the amino acid sequences of two or more proteins. "Percent (%) amino acid sequence identity" with respect to a reference polypeptide sequence is defined as the percentage of amino acid residues in a sequence that are identical with the amino acid residues in the reference polypeptide sequence, after aligning the sequences and introducing gaps, if necessary, to achieve the maximum percent sequence identity.

To determine the sequence identity, the sequence of a query protein is aligned to the sequence of a reference protein, for example, to SEQ ID NO: 4. Methods for alignment are well known in the art. For example, for determining the extent of an amino acid sequence identity of an arbitrary polypeptide relative to the amino acid sequence of SEQ ID NO: 4, the SIM Local similarity program is preferably employed (Xiaoquin Huang and Webb Miller (1991), *Advances in Applied Mathematics*, vol. 12: 337-357), that is freely available (see also: <http://www.expasy.org/tools/sim-prot.html>). For multiple alignment analysis, ClustalW (Thompson et al. (1994) *Nucleic Acids Res.*, 22(22): 4673-4680) or Clone Manager (Scientific & Educational Software) are preferably used.

In the context of the present invention, the extent of sequence identity between a modified sequence and the sequence from which it is derived (also termed: "parental sequence") is generally calculated with respect to the total length of the unmodified sequence, if not explicitly stated otherwise. Each amino acid of the query sequence that differs from the reference amino acid sequence at a given position is counted as one difference. An insertion or deletion in the query sequence is also counted as one difference. For example, an insertion of a linker between two ubiquitin moieties is counted as one difference compared to the reference sequence. The sum of differences is then related to the length of the reference sequence to yield a percentage of non-identity. The quantitative percentage of identity is calculated as 100 minus the percentage of non-identity. In this invention, in specific

cases of determining the identity of the ubiquitin mutein aligned against di-ubiquitin (for example, SEQ ID NO: 4 or SEQ ID NO: 43), differences in positions 45, 75, 76, 121, 151, and/or 152 of di-ubiquitin are not counted, in particular, because they are not relevant for the novel binding capability of the ubiquitin mutein. The ubiquitin moiety can be modified in amino acid residues 45, 75, 76, 121, 151, and/or 152 of di-ubiquitin without affecting its binding capability; said modifications might, however, be relevant for achieving modifications in the biochemical properties of the ubiquitin mutein. Generally, the di-ubiquitin sequence *used as starting material* for the modifications has an amino acid identity of at least 95 % amino acids, of at least 96 %, or of at least 97 % to SEQ ID NO: 4. Thus, a polypeptide which is, for example, 95 % "identical" to a reference sequence may comprise, for example, five point mutations, e.g. substitutions, or four point mutations and one insertion, per 100 amino acids, compared to the reference sequence of SEQ ID NO: 4. An example is given in SEQ ID NO: 43.

Binding proteins of the invention comprise di-ubiquitin muteins wherein the two ubiquitin mutein moieties are **linked directly without any linker** to result in unique and high affinity binding proteins with substitutions at least in 12, 13, or 14 positions selected from R42, I44, H68, V70, R72, L73, R74, K82, L84, Q138, K139, E140, S141, and T142 of di-ubiquitin with between 85 % and 92 % sequence identity to di-ubiquitin. In addition, binding proteins of the invention can be genetically fused to other functional protein moieties. In the context of such fusion proteins of the invention the term "linker" refers to a single amino acid or a polypeptide that joins at least two other protein molecules covalently. The linker is genetically fused to the first and second protein or protein moieties to generate a single, linear polypeptide chain. The length and composition of a linker may vary between at least one and up to about 50 amino acids. Preferably, the linker length is between one and 30 amino acids. More preferably, the peptide linker has a length of between 1 and 20 amino acids; e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 amino acids.

It is preferred that the amino acid sequence of the peptide linker is not immunogenic to human beings, stable against proteases and optionally does not form a secondary structure. An example is a linker comprised of small amino acids such as glycine, serine or alanine. The linkers can be glycine-rich (e.g., more than 50 % of the residues in the linker can be glycine residues). Preferred are glycine-serine-linker of variable length consisting of glycine and serine residues only. In general, linkers of the structure (SGGG)_n or permutations of SGGG, e.g. (GGGS)_n, can be used wherein n can be any number between 1 and 6, preferably 1 or 2 or 3. Also preferred are linkers comprising further amino acids. Other linkers for the genetic fusion of proteins are known in the art and can be used.

In case of chemical conjugates of the binding proteins of the invention, the term "linker" refers to any chemical moiety which connects the binding protein with other proteinaceous or non-proteinaceous moieties either covalently or non-covalently, e.g., through hydrogen bonds, ionic or van der Waals interactions, such as two complementary nucleic acid molecules attached to two different moieties that hybridize to each other, or chemical polymers such as polyethylene glycol or others. Such linkers may comprise reactive groups which enable chemical attachment to the protein through amino acid side chains. Such linkers and reactive groups are well-known to those skilled in the art and not described further.

Advantages of the invention

Compared to conventional monoclonal antibodies, a significant advantage of the binding proteins of the invention is the reduced complexity in terms of (i) reduced size (e.g. of maximal about 152 amino acids), (ii) simple molecular structure (e.g., one chain compared to four chains of an antibody), and (iii) no posttranslational modifications required for full functionality. These factors contribute to an easy handling of the molecules including simple genetic engineering as well as easy production and purification methods.

Although binding molecules based on antibody formats are well-known, there are still major drawbacks, for example suboptimal physicochemical properties or low production yields of those molecules. The binding proteins

of the invention provide simple molecular formats which are easy to engineer and the resulting molecules have favorable physicochemical properties (such as solubility and stability), high-level expression and allow easy production methods.

Methods for the generation of ubiquitin muteins having modifications selected from positions 2, 4, 6, 62, 63, 64, 65, 66 of one ubiquitin moiety (SEQ ID NO: 1) were described (see EP1626985B1). Another method for identifying hetero-multimeric modified ubiquitin molecules relates to hetero-multimeric modified ubiquitin molecules with substitutions selected from positions 2, 4, 6, 8, 62, 63, 64, 65, 66, 68 of two ubiquitin moieties fused with a peptide linker (see EP2379581B1).

Although these ubiquitin mutein proteins represent high affinity binding proteins and efficient methods for their production and identification, there are disadvantages such as the low solubility of the libraries required for the identification of such binding molecules. The libraries known in the art are prone to low solubility and low expression rates, resulting in inefficient screening procedures and adding complexity to the process, lowering yields, and increasing costs.

Taken together, due to several limitations of known binding proteins and their methods of generation and identification there was a strong need in the prior art to provide novel binding proteins with improved properties. In this context, there was a need for novel generation and identification methods, in particular for highly soluble and expressible libraries encoding binding proteins based on di-ubiquitin muteins.

An advantage of the invention is the method of generation and identification of novel binding proteins by providing libraries with enhanced solubility and expression rates. An advantage of the novel binding proteins of the invention is that they bind with high efficiency to targets.

The present invention provides libraries with a combination of defined positions for randomization for identifying artificial binding proteins based on di-ubiquitin muteins that are particularly well-suited for binding targets but overcome the disadvantages of the prior art.

Description of preferred embodiments of the invention

One aspect of the invention relates to a binding protein of the invention having binding affinity (K_D) of less than 1000 nM to a target wherein the affinity of the binding protein to the target is at least 10fold higher than for SEQ ID NO: 4. The binding protein comprises or essentially consists of or consists of an amino acid sequence wherein at least 12 amino acids selected from positions R42, I44, H68, V70, R72, L73, R74, K82, L84, Q138, K139, E140, S141, and T142 of di-ubiquitin are substituted wherein the binding protein has at least 85 % sequence identity to di-ubiquitin. Thus, the binding proteins show maximal 92 % sequence identity to di-ubiquitin. It is unexpected and surprising that the specific combination of at least said 12 amino acids in said positions is generating a high affinity binding protein for a target wherein said target affinity is at least 10fold higher than for SEQ ID NO: 4.

The preferred binding proteins comprise 152 amino acids with at least 85 %, preferably at least 87 % identity to di-ubiquitin (SEQ ID NO: 4), provided that at least 12 positions selected from R42, I44, H68, V70, R72, L73, R74, K82, L84, Q138, K139, E140, S141, and T142 are substituted. Surprisingly, the specific combination of substitutions in said 12, 13, or 14 positions of di-ubiquitin (SEQ ID NO: 4) result in high affinity binding proteins.

These proteins are artificial proteins that are created *de novo*. The binding proteins of the invention do not exist in nature. Examples for *de novo* created binding proteins are provided in SEQ ID NOs: 5-42.

It is preferred that a binding protein of the invention has additionally 0, 1, 2, 3, 4, 5, or 6 substitutions. In other words, it is preferred that a binding proteins of the invention has at least 85 %, 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, preferably at least 85 % or 87 % sequence identity to maximal 92 % sequence identity to di-ubiquitin (SEQ ID NO: 4). The binding protein of the invention is substituted in positions R42, I44, H68, V70, R72, L73, R74, K82, L84, Q138, K139, E140, S141, and T142 of di-ubiquitin and has no further substitution (for example, SEQ ID

NOs: 29, 33, 39, 40, 41, 42), one additional substitution (for example, SEQ ID NOs: 27, 28, 31, 32), two additional substitutions (for example, SEQ ID NOs: 14, 16, 21, 25, 26, 30, 35), three additional substitutions (for example, SEQ ID NOs: 6, 12, 13, 15, 17, 18, 20, 34, 36), four additional substitutions (for example, SEQ ID NOs: 10, 11, 19, 22, 23, 24), five additional substitutions (for example, SEQ ID NOs: 5, 7, 8, 9), or six additional substitutions (for example, SEQ ID NO: 37). For example, further 1, 2, 3, 4, 5, or 6 substitutions in addition to substitutions in positions R42, I44, H68, V70, R72, L73, R74, K82, L84, Q138, K139, E140, S141, and T142 of di-ubiquitin may be preferably selected from positions 6, 10, 11, 15, 20, 21, 23, 28, 31, 34, 36, 40, 46, 48, 49, 52, 58, 62, 63, 75, 78, 88, 92, 95, 96, 98, 114, 120, 124, 131, 133, 144, 147 of di-ubiquitin, more preferably selected from positions 46, 78, 92, 23, 52.

A binding protein of the invention is having a specific binding affinity (K_D) of at least 1000 nM, preferably at least 100 nM, preferably at least 10 nM, when determined by surface plasmon resonance, to a target wherein said target is not a natural binding target of di-ubiquitin (i.e. affinity of the binding protein to the target is at least 10fold higher than for di-ubiquitin. For example, a binding protein of the invention is having a specific binding affinity of at least 700 nM, preferably at least 100 nM, preferably at least 10 nM, to peptide targets as for example Her2, CD3, EGFR, or EpCAM, when determined by surface plasmon resonance.

The present invention has been successfully established on the following representative protein targets: Her2, CD3, EpCAM, and EGFR. It is to be understood that these protein targets have only been selected to show that the presently described methods can be successfully carried out by a person skilled in the art without undue burden after having received the information provided herein. The invention is not restricted to these specific protein targets but can be performed on all or most of target molecules known in the art. Those targets can be selected by the skilled artisan within his general knowledge of the art.

Her2 (Human Epidermal Growth Factor Receptor 2; synonym names are ErbB-2, Neu, CD340 or p185) is a 185-kDa receptor first described in 1984 (Schlechter et al (1984) Nature 312:513-516). Amplification or over-expression of this gene has been shown to play an important role in the pathogenesis and progression of certain aggressive types of breast cancer, and Her2 is known as an important biomarker and target of therapy for the disease. Other tumors where Her2 plays a role include ovarian cancer and gastric cancer.

The epidermal growth factor receptor (EGFR; synonym names are HER1 or ErbB1) is the cell-surface receptor for members of the epidermal growth factor family (EGF-family). EGFR is known for its role in lung cancer, head and neck cancer and colorectal cancer. The involvement in many cancers validates EGFR as a useful therapeutic target and supports the search for improved understanding of receptor biology and the development of improved therapies. Potential causes of the modest efficacy of current EGFR antagonists include the inability to effectively compete with target, especially in the presence of autocrine signaling; insufficient down-regulation of receptor; lack of inhibition of constitutively active EGFRvIII; and mutational escape. Thus, novel binders capable of downregulation and/or inhibition via different modes of action would be beneficial and multivalent and/or multispecific binders against EGFR hold the potential to be more effective in this respect.

epCAM (epithelial cell adhesion molecule) is a 30 to 40 kDa transmembrane glycoprotein involved in cell signaling, migration, proliferation, and differentiation. The protein is expressed in neoplasms derived from epithelia and is thus a marker for the diagnosis of various cancers.

CD3 (cluster of differentiation 3) is a cell-surface co-receptor on T-cells. The protein complex CD3 is composed of four distinct chains that associate with a T-cell receptor (TCR) to generate an activation signal in T lymphocytes. Drugs binding to CD3 might be useful in immunosuppressant therapies for example for autoimmune diseases such as diabetes type 1.

The term „Her2” or “EGFR” or “EpCAM” or “CD3” comprises all polypeptides which show a sequence identity of at least 70 %, 80 %, 85 %, 90 %, 95 %, 96 % or 97 % or more, or 100 % to accession number NP_004439 (Her2), to accession number NP_005219 (EGFR), to accession number NP_002354.2 (EpCAM), or to accession number

P077661, respectively, and have the functionality of Her2, EGFR, EpCAM, or CD3, respectively. The extracellular domain of Her2 is represented by uniprot accession number P04626 (residues 1-652). The extracellular domain of human EGFR is represented by uniprot accession number P00522 (residues 1-645). The extracellular domain of human EpCAM is represented by uniprot accession number P16422 (residues 1-265).

Many selected examples of binding proteins with at least 12 substitutions in positions 42, 44, 68, 70, 72, 73, 74, 82, 84, 138, 139, 140, 141, and 142 of di-ubiquitin are provided in this invention. The binding proteins of the invention bind to the non-natural targets of di-ubiquitin with measurable binding affinity of less than 1000 nM, less than 500 nM, less than 100 nM, less than 20 nM, less than 10 nM (for example, SEQ ID NOs: 6, 14, 15, 18, 22, 24, 25, 26, 28, 35, 36, 38), and more preferred less than 1 nM (for example, SEQ ID NOs: 7, 8, 9, 10, 11, 12, 13, 16, 17, 19, 20, 23). The unmodified di-ubiquitin (SEQ ID NO: 4) does not naturally bind to protein targets for example Her2, epCAM, CD3, or EGFR with any measurable binding affinity. All binding proteins of the invention show *de novo* created binding to targets with high affinity, i.e. wherein the affinity of the binding protein to the target is at least 10fold higher than for di-ubiquitin. Specific examples for binding proteins are provided in FIG. 3, Table 1, and Table 4. Examples for Her2 binding proteins are given in SEQ ID NOs: 5-38, for CD3 binding proteins in SEQ ID NO: 39, for EpCAM binding proteins in SEQ ID NO: 40 and 41, and for EGFR binding proteins in SEQ ID NO: 42.

Another preferred binding protein (Affilin-141975) is substituted in 20 amino acids of di-ubiquitin (substitutions in amino acids in positions 42, 44, 68, 70, 72, 73, 74, 82, 84, 138, 139, 140, 141, 142 and in positions 52, 75, 78, 96, 114, 131). The Her2 binding protein Affilin-141884 shows 87 % identity to di-ubiquitin (SEQ ID NO: 4). The Her2 binding capability was confirmed by FACS analysis (SK-Br-3 cells).

For example, a preferred binding protein (Affilin-142628) is substituted in 18 amino acids of di-ubiquitin (substitutions in amino acids in positions 42, 44, 68, 70, 72, 73, 74, 82, 84, 138, 139, 140, 141, 142 and in positions 23, 78, 92, 46). The Her2 binding protein Affilin-142628 shows 88 % identity to di-ubiquitin (SEQ ID NO: 4). The Her2 binding protein shows strong binding to tumor tissues expressing the target.

Another preferred binding protein (Affilin-141884) is substituted in 16 amino acids of SEQ ID NO: 24 (substitutions in amino acids in positions 42, 44, 68, 70, 72, 73, 74, 82, 84, 138, 139, 140, 141, 142 and in positions 27 and 78). The Her2 binding protein Affilin-141884 shows 90.5 % identity to di-ubiquitin (SEQ ID NO: 4). The Her2 binding protein shows strong binding to tumor tissues expressing the target.

Another preferred binding protein (Affilin-141926) is substituted in 15 amino acids of di-ubiquitin (substitutions in amino acids in positions 42, 44, 68, 70, 72, 73, 74, 82, 84, 138, 139, 140, 141, 142 and in position 78). The Her2 binding protein Affilin-141884 shows 90 % identity to di-ubiquitin (SEQ ID NO: 4).

Other preferred binding proteins (for example, Affilin-144160, Affilin 143488, Affilin-143283, Affilin-142265) are substituted in 14 amino acids of di-ubiquitin (substitutions in amino acids corresponding to positions 42, 44, 68, 70, 72, 73, 74, 82, 84, 138, 139, 140, 141, 142). The binding proteins shows 91 % identity to di-ubiquitin (SEQ ID NO: 4).

Preferred substitutions of the Her2 binding protein based on di-ubiquitin are substitutions of amino acids selected from position 70 and 140 by aromatic amino acids. Further preferred substitutions of the Her2 binding protein based on di-ubiquitin are substitutions of amino acids selected from position 42 by a polar amino acid, position 44 is substituted by a hydrophobic or polar amino acid, position 68 is substituted by an aromatic amino acid, position 72 is substituted by a polar or aromatic amino acid, position 73 is substituted by any amino acid but not basic or acidic amino acid, position 74 is substituted by an aromatic, basic or polar amino acid, position 82 is substituted by any amino acid but not basic or acidic amino acid, position 84 is substituted by a basic or acidic amino acid, position 138 is substituted by a basic or acidic or polar amino acid, position 139 is substituted by acidic or hydrophobic amino acid or Glycine, position 141 is substituted by hydrophobic or polar or basic amino acid, and/or position 142 is substituted by a hydrophobic or polar amino acid. Preferred substitutions of the Her2 binding

protein based on di-ubiquitin are selected from R42T, R42S, R42L, I44A, I44V, I44S, I44T, H68W, H68Y, H68F, V70Y, V70W, R72T, R72F, R72G, R72Y, L73W, L73S, L73V, L73I, R74Y, R74S, R74N, R74K, K82T, K82L, K82N, K82I, K82Y, L84H, L84D, L84E, L84S, Q138S, Q138R, Q138E, K139E, K139G, K139L, E140W, S141A, S141R, T142I, T142L, and/or T142N. Further preferred are Her2 binding proteins with a specific combination of amino acid substitutions in SEQ ID NO: 4, for example, at least R42T, I44A, H68W, V70Y, R72T, L73W, R74Y, K82T, L84H (for example, SEQ ID NOs: 7-29). Other preferred Her2 binding proteins with a specific combination of amino acid substitutions in di-ubiquitin, are for example at least R42S, I44V, H68Y, V70Y, R72F, L73S, K82L, L84D (for example, SEQ ID NOs: 34, 35, 36, and 36). Further preferred are Her2 binding proteins with a specific combination of amino acid substitutions in di-ubiquitin, for example, Q138S, K139E, E140W, S141A, T142I (for example, in SEQ ID NOs: 5, 7-29, 33, 36, 37), or Q138R, K139G, E140W, T142L (for example, in SEQ ID NOs: 6, 34, 35), or Q138 E, K139L, E140W, S141R, T142N (for example, in SEQ ID NOs: 30, 31, 32).

Table 1 lists examples for substitutions of di-ubiquitin in order to generate a novel binding protein (in addition to FIG. 1). All binding proteins (for example, SEQ ID NOs: 39-42) are substituted in 14 positions selected from positions 42, 44, 68, 70, 72, 73, 74, 82, 84, 138, 139, 140, 141, and 142 of di-ubiquitin. SEQ ID NO: 39 binds to the protein target CD3, SEQ ID NOs: 40-41 bind to the protein target epCAM, and SEQ ID NO: 42 binds to the protein target EGFR.

Table 1.

SEQ	Affilin	target	42	44	68	70	72	73	74	82	84	138	139	140	141	142
42	142265	EGFR	E	T	W	E	Y	A	K	Q	I	T	R	L	V	W
41	143283	epCAM	R	Q	G	L	T	Q	R	Y	N	F	R	F	Q	H
40	143488	epCAM	Y	Y	K	Y	W	R	V	A	G	G	S	Y	G	Y
39	144160	CD3	W	W	K	W	V	D	K	Y	R	A	L	E	S	G

It is preferred that the binding proteins of the invention comprise amino acid sequences that exhibits at least 85 % or at least 87 % or at least 91 % or at least 94 % or at least 96 % sequence identity to one or more of the amino acid sequences of SEQ ID NO: 5-42.

The further characterization of binding proteins can be performed in the form of soluble proteins. The appropriate methods are known to those skilled in the art or described in the literature. The methods for determining the binding affinities are known *per se* and can be selected for instance from the following methods known in the art: Surface Plasmon Resonance (SPR) based technology, Bio-layer interferometry (BLI), enzyme-linked immunosorbent assay (ELISA), flow cytometry, fluorescence spectroscopy techniques, isothermal titration calorimetry (ITC), analytical ultracentrifugation, radioimmunoassay (RIA or IRMA) and enhanced chemiluminescence (ECL). Some of the methods are described in the Examples below.

For stability analysis, for example spectroscopic or fluorescence-based methods in connection with chemical or physical unfolding are known to those skilled in the art. Exemplary methods for characterization of Her2 binding proteins are outlined in the Examples section of this invention.

A biochemical binding analysis of binding proteins of the invention to given targets is summarized in Table 4 and further described in the Examples. In an embodiment of this aspect, the Her2-binding protein has a dissociation constant K_D to human Her2 in the range between between 0.01 nM and 700 nM, more preferably between 0.05 nM and 500 nM, more preferably between 0.1 nM and 100 nM, more preferably between 0.1 nM and 20 nM, more preferably between 0.1 nM and 10 nM. The dissociation constant K_D can be determined for example by ELISA or by surface plasmon resonance assays. Typically, the dissociation constant K_D is determined at 20°C, 25°C or

30°C. If not specifically indicated otherwise, the K_D values recited herein are preferably determined at 25°C by surface plasmon resonance.

Another aspect of the invention covers a binding protein of the invention and further at least one additional protein or molecule. The additional protein can be a second binding protein with identical or different specificity for an antigen as the first binding protein. One embodiment of the invention covers a fusion protein or a conjugate comprising an Affilin-antibody fusion protein or conjugate, optionally further fused with or conjugated to a moiety preferably selected from at least one member of the groups (i), (ii) and (iii) consisting of (i) a moiety modulating pharmacokinetics selected from a PEG, a HSA, an albumin-binding peptide, or an immunoglobulin or immunoglobulin fragments, a polysaccharide, and, (ii) a therapeutically active component, optionally selected from a monoclonal antibody or a fragment thereof, a cytokine, a chemokine, a cytotoxic compound, an enzyme, or derivatives thereof, or a radionuclide, and (iii) a diagnostic component, optionally selected from a fluorescent compound, a photosensitizer, a tag, an enzyme or a radionuclide. The conjugate molecule can be attached e.g. at one or several sites through a peptide linker sequence or a carrier molecule.

Further conjugation with proteinaceous or non-proteinaceous moieties to generate protein conjugates according to the invention can be performed applying chemical methods well-known in the art. In particular, coupling chemistry specific for derivatization of cysteine or lysine residues is applicable. In case of the introduction of non-natural amino acids further routes of chemical synthesis are possible, e.g. "click chemistry" or aldehyde specific chemistry and others.

Conjugates thus obtained can be selected from one or more of the following examples: (i) conjugation of the protein via lysine residues; (ii) conjugation of the protein via cysteine residues via maleimide chemistry; in particular, cysteine residues can be specifically introduced and can be located at any position suitable for conjugation of further moieties, (iii) peptidic or proteinogenic conjugations. These and other methods for covalently and non-covalently attaching a protein of interest to other functional components are well known in the art, and are thus not described in further detail here.

A further embodiment relates to binding proteins according to the invention, further comprising a moiety modulating pharmacokinetics or biodistribution, preferably selected from PEG, HSA, or an immunoglobulin or immunoglobulin fragments, for example an Fc fragment. Several techniques for producing proteins with extended half-life are known in the art.

In a further aspect of the invention, a binding protein of the invention or fusion protein or conjugate is used in medicine, in particular in a method of medical treatment or diagnosis, preferably in cancer or autoimmune disorders.

For example, the membrane proteins Her2, EGFR, and epCAM are known to be upregulated in tumor cells, resulting in uncontrolled growth of tumor cells and in the formation of metastases. New therapies for cancer patients include an inhibition of EGFR by targeted therapeutics such as for example the monoclonal antibodies Cetuximab (Erbix®) or Panitumumab (Vectibix®) or an inhibition of Her2 by for example Trastuzumab (Herceptin®) or Pertuzumab (Perjeta®). T-DM1, an antibody-drug conjugate, is highly effective against breast, uterine and ovarian carcinosarcomas overexpressing Her2. For EGFR, further monoclonal antibodies are in development, for example, Zalutumumab (HuMax-EGFr) and Nimotuzumab (h-R3 or Theraloc). For EpCAM, monoclonal antibodies such as Edrecolomab or Catumaxomab are in development.

Overexpression of Her2 and EGFR was described in a wide variety of cancers. For example, overexpression of Her2 occurs in approximately 15–30 % of breast cancers and 10–30 % of gastric/gastroesophageal cancers, and has also been observed in other cancers like ovary, endometrium, bladder, lung colon, head and neck. Thus, the pharmaceutical composition comprising the binding protein of the invention, can be used for treatment of cancer in which the target is relevant for the development of the disease including but not limited to particularly breast, ovarian, gastric, but also in lung, head and neck, cervical, prostate, pancreas, and others.

EpCAM plays different roles and is heterogeneous overexpressed in carcinomas, but also in cancer stem cells. EpCAM is found not only in tumors but also in normal epithelium. Further EpCAM plays a role in a genetic disorder (Lynch syndrome) leading to an increased cancer risk.

CD3 is expressed in stem cells from and is required for T-cell activation. Antibodies such as Otelixizumab (TRX4) are in development for immunosuppressant therapies for autoimmune diseases (e.g. diabetes type 1).

The compositions contain a therapeutically or diagnostically effective dose of the binding protein of the invention. The amount of protein to be administered depends on the organism to be treated, the type of disease, the age and weight of the patient and further factors known *per se*.

The invention covers a pharmaceutical composition comprising the binding protein, fusion protein or conjugate or the nucleic acid molecule of the invention, the vector of the invention, and/or the host cell or virus and a pharmaceutically acceptable carrier. The invention further covers a diagnostic agent comprising the binding protein or conjugate or the nucleic acid molecule of the invention, the vector of the invention, and/or the host cell or non-human host with a diagnostically acceptable carrier. The compositions contain a pharmaceutically or diagnostically acceptable carrier and optionally can contain further auxiliary agents and excipients known *per se*.

These include for example but are not limited to stabilizing agents, surface-active agents, salts, buffers, coloring agents etc.

The pharmaceutical composition comprising the binding protein can be in the form of a liquid preparation, a lyophilisate, a cream, a lotion for topical application, an aerosol, in the form of powders, granules, in the form of an emulsion or a liposomal preparation. The compositions are preferably sterile, non-pyrogenic and isotonic and contain the pharmaceutically conventional and acceptable additives known *per se*. In addition, reference is made to the regulations of the U.S. Pharmacopoeia or Remington's Pharmaceutical Sciences, Mac Publishing Company (1990).

In the field of human and veterinary medical therapy and prophylaxis pharmaceutically effective medicaments containing at least one binding protein in accordance with the invention can be prepared by methods known *per se*. Depending on the galenic preparation these compositions can be administered parentally by injection or infusion, systemically, intraperitoneally, intramuscularly, subcutaneously, transdermally or by other conventionally employed methods of application. The type of pharmaceutical preparation depends on the type of disease to be treated, the route of administration, the severity of the disease, the patient to be treated and other factors known to those skilled in the art of medicine.

In a still further aspect the invention discloses diagnostic compositions comprising binding protein according to the invention specifically binding specific targets or its isoforms together with diagnostically acceptable carriers.

Since enhanced Her2, EGFR, and epCAM expression is correlated with tumor malignancy, it is desirable to develop diagnostics for non-invasive imaging in order to gain information about the Her2, EGFR, and epCAM expression in patients. Furthermore, Her2, EGFR, and epCAM imaging could be useful for the assessment of the response of a patient to a therapeutic treatment. For example, using a protein of the invention labelled with a suitable radioisotope or fluorophore can be used for non-invasive imaging to determine the location of tumors and metastasis (for review see for example Milenic et al. 2008 Cancer Biotherapy & Radiopharmaceuticals 23: 619-631; Hoeben et al. 2011, Int. Journal Cancer 129: 870-878). Due to their pharmacokinetic characteristics, intact antibodies are not suitable for routine imaging. Due to their small size and high affinity, labelled fusion proteins of the invention are expected to be much better suited for use as diagnostics for imaging.

It is expected that a protein of the invention can be advantageously applied in therapy. In particular, the molecules are expected to show superior tumor targeting effect and desired biodistribution and thus, reduced side effects.

Pharmaceutical compositions of the invention may be manufactured in any conventional manner.

In one embodiment of the invention, novel binding proteins given targets, preferably protein or peptide targets, are provided wherein the binding proteins comprise substitutions in 12, 13, or 14 positions corresponding to

positions 42, 44, 68, 70, 72, 73, 74, 82, 84, 138, 139, 140, 141, and 142 of di-ubiquitin. Further key aspects of the invention cover the production of the binding proteins and the selection of the muteins of di-ubiquitin.

Binding molecules of the invention may be prepared by any of the many conventional and well known techniques such as plain organic synthetic strategies, solid phase-assisted synthesis techniques or by commercially available automated synthesizers. On the other hand, they may also be prepared by conventional recombinant techniques alone or in combination with conventional synthetic techniques. Conjugates according to the present invention may be obtained by combining compounds by chemical methods, e.g. lysine or cysteine-based chemistry, as described herein above.

According to another aspect of the invention, an isolated polynucleotide encoding a binding protein of the invention is provided.

In a further embodiment the invention relates to a vector comprising the nucleic acid molecule of the invention. A vector means any molecule or entity (e.g., nucleic acid, plasmid, bacteriophage or virus) that can be used to transfer protein coding information into a host cell.

The present invention furthermore relates to an isolated cell comprising the nucleic acid molecule of the invention or the vector of the invention. Suitable host cells include prokaryotes or eukaryotes. Various mammalian or insect cell culture systems can also be employed to express recombinant proteins. The invention also relates to a host cell carrying the vector of the invention. A host cell is a cell that has been transformed, or is capable of being transformed, with a nucleic acid sequence and thereby expresses a gene of interest. The term includes the progeny of the parent cell, whether or not the progeny is identical in morphology or in genetic make-up to the original parent cell, so long as the gene of interest is present.

In another aspect is provided a method of producing the binding protein of the invention, comprising the steps of a) culturing the host cell of the invention under conditions suitable for the expression of the binding protein and b) isolating the produced binding protein.

One embodiment of the present invention is directed to a method for the preparation of a binding protein according to the invention as detailed above, said method comprising the following steps: (a) preparing a nucleic acid encoding a binding protein as defined above, (b) introducing said nucleic acid into an expression vector; (c) introducing said expression vector into a host cell; (d) cultivating the host cell; (e) subjecting the host cell to culturing conditions under which a binding protein is expressed, thereby producing a binding protein as described above; (f) optionally isolating the binding protein produced in step (e).

Suitable conditions for culturing a prokaryotic or eukaryotic host are well known to the person skilled in the art. Cultivation of cells and protein expression for the purpose of protein production can be performed at any scale, starting from small volume shaker flasks to large fermenters, applying technologies well-known to any skilled in the art.

Following the expression of the di-ubiquitin mutein according to the invention, the binding protein is optionally isolated. Methods of isolation of the polypeptide produced are well-known in the art and comprise without limitation method steps such as ion exchange chromatography, gel filtration chromatography (size exclusion chromatography), or affinity chromatography. Further purification of binding proteins comprises methods known to someone skilled in the art. The selected methods depend on several factors known *per se* to those skilled in the art, for example the expression vector used, the host organism, the intended field of use, the size of the protein and other factors. For simplified purification the protein modified according to the invention can be fused to other peptide sequences having an increased affinity to separation materials. Preferably, such fusions are selected that do not have a detrimental effect on the functionality of the di-ubiquitin muteins or can be separated after the purification due to the introduction of specific protease cleavage sites. Such methods are also known to those skilled in the art.

Isolation of purified protein can be performed applying conventional methods and technologies well known in the art, such as centrifugation, precipitation, flocculation, different embodiments of chromatography, filtration, dialysis, concentration and combinations thereof, and others.

A further aspect of the invention relates to a method for the generation of a binding protein of the invention which binds to a given target with a dissociation constant (K_D) of 1000 nM or lower, said method comprising (a) 5
subjecting nucleotide triplets of a nucleic acid molecule of di-ubiquitin to mutagenesis in 12, 13, or 14 positions coding for amino acid positions corresponding to positions R42, I44, H68, V70, R72, L73, R74, K82, L84, Q138, K139, E140, S141, and T142 of the linear polypeptide sequence of di-ubiquitin (SEQ ID NO: 4), (b) resulting in one or more di-ubiquitin mutein nucleic acid molecule(s), (c) expressing the one or more di-ubiquitin mutein 10
nucleic acid molecule(s) obtained in (b) in a suitable expression system, (d) enriching the one or more di-ubiquitin mutein(s) having a dissociation constant (K_D) of 1000 nM or lower for the given target by means of selection and/or isolation, (e) wherein the affinity to the target is at least 10fold higher than for di-ubiquitin. Different procedures known *per se* are available for mutagenesis, such as methods for site-specific mutagenesis, methods for random mutagenesis, mutagenesis using PCR or similar methods. All methods are known to those skilled in 15
the art.

In one preferred embodiment of the invention, two ubiquitin moieties are directly linked in head to tail orientation to result a protein of 152 amino acids. Each mutein is based on a common scaffold represented by the amino acid sequence

MQIFVKLTGKITLEVEPSDTIENVKAKIQDKEGIPPDQQXLXFAGKQLEDGRTLSDYNIQKESTLXLXLLXXAAMQ 20
IFVXTXTGKITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIWAGKQLEDGRTLSDYNIXXXXXLHLVRLRLAA (SEQ ID NO: 3), wherein X individually corresponds to an amino acid residue selected from S, Y, W, V, H, R, D, N, E, Q, G, K, F, T, A, L, I, or F (not allowed are M, C, and P).

As a state-of-the-art method for library synthesis the favorable triplet technology (Morphosys Slonomics) is capable of synthesizing random libraries with an even (or uneven, if wanted) distribution of e.g. the 20 natural 25
amino acids. Assuming a random distribution of the 20 natural amino acids at e.g. 14 positions generates a pool of 20 to the power of 14 (20^{14}) theoretical di-ubiquitin muteins, each with a different amino acid composition and potentially different binding properties. Preferably, insertion of amino acid cysteine is avoided at all randomized positions resulting in 19^{14} theoretical di-ubiquitin muteins. This large pool of genes constitutes a library of different Affilin molecules.

A "randomly modified nucleotide or amino acid sequence" is a nucleotide or amino acid sequence which in a number of positions has been subjected to substitution, insertion, or deletion by nucleotides or amino acids, the nature of which cannot be predicted. In many cases the random nucleotides (amino acids) or nucleotide (amino acid) sequences inserted will be "completely random" (e. g. as a consequence of randomized synthesis or PCR-mediated mutagenesis). However, the random sequences can also include sequences which have a common 30
functional feature (e. g. reactivity with a target of the expression product) or the random sequences can be random in the sense that the ultimate expression product is of completely random sequence with e. g. an even distribution of the different amino acids.

The mutagenized nucleic acid molecule comprising randomized nucleotide triplets coding 12, 13, or 14 positions selected from sequence positions R42, I44, H68, V70, R72, L73, R74, K82, L84, Q138, K139, E140, S141, and 40
T142 is resulting in one or more mutein nucleic acid molecule(s) which will be expressing the one or more mutein nucleic acid molecule(s) obtained in (b) in a suitable expression system.

In some embodiments, the application provides libraries comprising a plurality of the non-naturally occurring binding polypeptides based on di-ubiquitin muteins as described herein. The libraries provided herein may comprise, for example, a sequence diversity of between at least about 10^{12} , about 10^{13} , or about 10^{14} , or more 45
polypeptides, each optional comprising a different amino acid sequence. In one embodiment of the invention, a

nucleic acid library is provided encoding muteins of di-ubiquitin, wherein the amino acid residues selected from 12, 13, or 14 positions corresponding to amino acid positions 42, 44, 68, 70, 72, 73, 74, 82, 84, 138, 139, 140, 141, and 142 of SEQ ID NO: 3 (SEQ ID NO: 4) are substituted in the muteins by any amino acid except are M, C, or P. It is most preferred that the library comprises at least 19¹² polypeptides each comprising a different amino acid sequence. Sequence differences between library members are responsible for the diversity present in the library. The library may take the form of a simple mixture (population; plurality) of polypeptides or nucleic acids. Advantageously, the nucleic acids are incorporated into expression vectors, in order to allow production of the polypeptides encoded by the nucleic acids.

The library can be cloned into a phagemid vector (e.g. pCD87SA (Paschke, M. and W. Hohne (2005). " Gene 350(1): 79-88)). The library may be displayed on the surfaces of ribosomes, bacteriophage, viruses, bacteria, or yeast cells, preferably displayed on bacteriophage, and subjected to repeated rounds of panning against the respective target. Di-ubiquitin muteins from enriched phage pools are cloned into expression vectors for individual protein expression. Preferably, expression of the di-ubiquitin mutein enables screening for specific binding proteins by established techniques, such as ELISA on automated high-throughput screening platforms. Identified clones with desired binding properties are then sequenced to reveal the amino acid sequences of Affilin molecules. The identified binding protein may be subjected to further maturation steps, e.g. by generating additional libraries based on alterations of the identified sequences and repeated phage display, ribosomal display, panning and screening steps as described above.

The gene pool libraries obtained as described above can be combined with appropriate functional genetic elements which enable expression of proteins for selection methods such as display methods. A method for identifying a binding protein is provided with the affinity to the target of at least 10fold higher than for di-ubiquitin comprising screening the library to identify a polypeptide that binds to the given target. A method for identifying a binding protein is thus comprising the following steps: a. providing a library wherein the amino acid residues selected from 12, 13, or 14 positions corresponding to 42, 44, 68, 70, 72, 73, 74, 82, 84, 138, 139, 140, 141, and 142 of SEQ ID NO: 3 (SEQ ID NO: 4) are randomized, b. providing a potential target to said library (for example, wherein the target is a protein or a peptide or a domain of a protein); c. contacting said library with said target; and d. identifying a binding protein by a screening process, wherein said binding protein binds to said target with a specific binding affinity of at least 1000 nM, and optionally isolating said binding protein with said binding affinity.

The expressed proteins are contacted according to the invention with a target molecule to enable binding of the partners to each other if a binding affinity exists. This process enables identification of those proteins which have a binding activity to the target molecule. See, for example, EP2379581B1, which is herewith incorporated by reference.

Contacting according to the invention is preferably performed by means of a suitable presentation and selection method such as the phage display, ribosomal display, mRNA display or cell surface display, yeast surface display or bacterial surface display methods, preferably by means of the phage display method. For complete disclosure, reference is made also to the following references: Hoess, Curr. Opin. Struct. Biol. 3 (1993), 572-579; Wells and Lowmann, Curr. Opin. Struct. Biol. 2 (1992), 597-604; Kay et al., Phage Display of Peptides and Proteins-A Laboratory Manual (1996), Academic Press. The methods mentioned above are known to those skilled in the art. In the phage display procedure described herein recombinant variations of ubiquitin are presented on a filamentous phage while the coding DNA of the presented variation is present at the same time packed in a single-stranded form in the phage envelope. Thus, in the frame of an affinity enrichment variations having certain properties can be selected from a library and their genetic information can be amplified by infection of suitable bacteria or added to another cycle of enrichment, respectively. Presentation of the di-ubiquitin mutein on the

phage surface is achieved by genetic fusion to a signal sequence and a capsid or surface protein of the phage. Furthermore, the encoded protein can contain further functional elements such as an affinity tag or an antibody epitope for detection and/or purification by affinity chromatography or a protease recognition sequence for specific cleavage of the protein in the course of the affinity enrichment. Furthermore, an amber stop codon can be present

5 between the gene for the di-ubiquitin mutein and the coding region of the phage capsid protein which is not recognized during translation in a suitable suppressor strain.

The bacterial vector suitable for the selection procedure in the context of the isolation of di-ubiquitin muteins and into which the gene cassette for the fusion protein described is inserted is referred to as phagemid. Among others, it contains the intergenic region of a filamentous phage (e.g. M13 or f1) or a portion thereof which in the

10 case of a superinfection of the bacterial cell carrying the phagemid by means of helper phages results in the packaging of a covalently closed strand of phagemid DNA into a phage capsid.

Phage particles obtained can be selected with respect to the binding of the di-ubiquitin muteins presented thereon to any target by means of methods known to those skilled in the art. For this purpose, the presented di-ubiquitin muteins can be transiently immobilized to target substance and can be specifically eluted after non-binding

15 variations have been separated. The phage particles obtained in this manner can be re-amplified and enriched by successive cycles of selection and amplification of di-ubiquitin muteins with binding properties to selected targets. The further biochemical and functional characterization of binding proteins was described above.

The invention permits the person skilled in the art to remove, from a chosen repertoire of polypeptides, those polypeptides which are incapable of binding to the target with the affinity specified in the claims. The invention

20 permits the person skilled in the art to enrich a chosen repertoire of di-ubiquitin muteins which are functional and capable of binding to a given target.

EXAMPLES

The following Examples are provided for further illustration of the invention. The invention is particularly exemplified by particular ubiquitin muteins with binding capability to protein targets Her2, CD3, epCAM, or EGFR.

25 The invention, however, is not limited thereto, and the following Examples merely show the practicability of the invention on the basis of the above description. For a complete disclosure of the invention reference is made also to the literature cited in the application which is incorporated completely into the application by reference.

Example 1. Identification of binding proteins

Library Construction and Cloning Two ubiquitin moieties each comprising seven randomized amino acid positions were synthesized by triplet technology (MorphoSys Slonomics, Germany) to achieve a well-balanced amino acid distribution. A mixture of 19-amino acid coding premade double-stranded triplets (excluding cysteine) was used for the synthesis of the library. Both ubiquitin moieties were directly linked in head to tail orientation to result a

30 protein of 152 amino acids with 14 randomized amino acid positions. The sequence of di-ubiquitin with 14

35 randomized positions is shown in SEQ ID NO: 3:

MQIFVKLTGKITLEVEPSDTIENVKAKIQDKEGIPPDQQXLXFAGKQLEDGRTLSDYNIQKESTLXLXAXXAAMQ
IFVXTXTGKITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIWAGKQLEDGRTLSDYNIXXXXXLHLVRLRAA
wherein X individually corresponds to an amino acid residue selected from S, Y, W, V, H, R, D, N, E, Q, G, K, F,
T, A, L, I, or F.

40 The 14 randomized amino acids in the mutein correspond to positions R42, I44, H68, V70, R72, L73, R74, K82, L84, Q138, K139, E140, S141, and T142 of di-ubiquitin. The sequence of di-ubiquitin is shown in SEQ ID NO: 4:

MQIFVKLTGTGKITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTLSDYNIQKESTLHLVLRRLRAAMQI
FVKLTGTGKITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIWAGKQLEDGRTLSDYNIQKESTLHLVLRRLRAA.

- 5 The construct was ligated with a modified pCD87SA phagemid (herein referred to as pCD12) using standard methods known to a skilled person. The pCD12 phagemid comprises a modified *torA* leader sequence (deletion of amino acid sequence QPAMA) to achieve protein processing without additional amino acids at the N terminus. Aliquots of the ligation mixture were used for electroporation of *Escherichia coli* ER2738 (Lucigen). The library is referred to as SPIF. Unless otherwise indicated, established recombinant genetic methods were used, for
- 10 example as described in Sambrook *et al.*
- Targets.* Recombinant human target-Fc Chimera were purchased from R&D Systems. A DNA sequence encoding the extracellular domain of human EGFR (Uniprot Accession Number P00533; residues 1-645) or encoding the extracellular domain of human Her2 (Uniprot Accession Number P04626; residues 1-652) or encoding the extracellular domain of human epCAM (Uniprot Accession Number P16422; residues 1-265) or encoding human
- 15 CD3 epsilon (Uniprot Accession Number P16422; all residues) was genetically fused with the Fc region of human IgG1 at the C-terminus.
- TAT Phage Display Selection.* The SPIF library was enriched against the given protein target using TAT phage display as selection system. After transformation of competent bacterial ER2738 cells (Lucigen) with phagemid pCD12 carrying the SPIF library, phage amplification and purification was carried out using standard methods
- 20 known to a skilled person. For selection the target protein was provided as Fc-fusion protein (Her2-Fc or EGFR-Fc or epCAM-Fc or CD3-Fc) immobilized on Dynabeads® Protein A or G. The target concentration during phage incubation varied from 200 nM (first round) to 50 nM (third round). Target phage complexes were magnetically separated from solution and washed several times. Target bound phages were eluted by trypsin. To deplete the phage library for Fc-binding variants a preselection of phages with immobilized Fc-fragment of IgG1 (Athens
- 25 Research & Technology) was performed prior to round two and three.
- To identify target specific phage pools, eluted and reamplified phages of each selection round were analysed by phage pool ELISA. Wells of a medium binding microtiter plate (Greiner bio-one) were coated at 2,5 µg/ml with Her2-Fc or EGFR-Fc or epCAM-Fc or CD3-Fc and Fc-fragment of IgG1 (5 µg/ml), respectively. Bound phages were detected using α-M13 HRP-conjugated antibody (GE Healthcare). Eluted and reamplified phages of round
- 30 three showed specific binding to the target and were used subsequently for pool maturation by error prone PCR. Isolated phagemid pools served as template for error prone PCR (GeneMorph II Random Mutagenesis Kit, Agilent Technologies). The amplified pool of SPIF variants carrying now additional substitutions compared to the library positions was recloned into phagemid pCD12 and transformed into ER2738 for phage amplification and purification. The phages were again subjected to two rounds of panning as described above. The target was
- 35 employed at a concentration of 5 nM and 1 nM in round one and two, respectively. For both rounds a preselection with Fc-fragment of IgG1 was performed. To analyze the matured and selected pools for specific target binding a phage pool ELISA was performed as described above.
- Cloning of Target Binding Phage Pools into an Expression Vector.* Upon completion of the selection procedure the target specific DNA pools of maturation selection round one and two were amplified by PCR according to
- 40 methods known in the art, cut with appropriate restriction nucleases and ligated into a derivative of the expression vector pET-28a (Merck, Germany) comprising a Strep-Tag II (IBA GmbH).
- Single Colony Hit Analysis.* After transformation into BL21 (DE3) cells (Merck, Germany) kanamycin-resistant single colonies were grown. Expression of the target-binding modified ubiquitin was achieved by cultivation in 96 deep well plates (Genetix, UK) using auto induction medium (Studier, 2005). Cells were harvested and
- 45 subsequently lysed. After centrifugation the resulting supernatants were screened by ELISA on plates coated with

target and detected with *Strep-Tactin*® HRP Conjugate (IBA GmbH). As detecting reagent TMB-Plus (Biotrend, Germany) was used and the yellow colour was developed using 0.2 M H₂SO₄ solution and measured in a plate reader at 450 nm versus 620 nm.

5 Example 2. Solubility analysis of the proteins encoded by the library

The SPIF library was cloned into an expression vector using standard methods known to a skilled person. Expression of single clones was carried out in 96 deep well plates. Cells were cultivated over 16-18 h at 37 °C in auto induction medium. The pelleted cells were lysed using lysozyme and freeze-thaw-cycles. Soluble and insoluble fractions were separated by centrifugation. Proteins were recovered from pellets by addition of 8 M urea. Supernatants and resuspended pellets were analyzed by Criterion™ precast gels and the Criterion Stain Free imager (BioRad). For solubility analysis the Image Lab software from BioRad was used. The solubility of the proteins encoded by the SPIF library and the expression level of all proteins encoded by the SPIF library was compared to the solubility and expression of the proteins encoded by four different di-ubiquitin based libraries. One library referred to as SPVF with randomized positions 6, 8, 62-66, 85, 87, 141-145 comprises two ubiquitin moieties (SEQ ID NO: 1) linked by a 3 amino acid linker (GIG). Further libraries with 14 different randomized positions between amino acids 46-66 and 122-142 of di-ubiquitin (SEQ ID NO: 4) and no linker between the two ubiquitin moieties (referred to as SPGO, SPOO, SPKK) were compared to SPIF.

Table 2: Libraries based on two ubiquitin moieties

Library name	Linker	Positions substituted in di-ubiquitin (SEQ ID NO: 4)
SPIF	-	42, 44, 68, 70, 72, 73, 74, 82, 84, 138-142
SPVF	GIG	6, 8, 62-66, 85, 87, 141-145
SPGO	-	14 selected from 53-63 and 123-137
SPOO	-	14 selected from 47-60 and 123-137
SPKK	-	14 selected from 46-66 and 122-142

The solubility of the proteins encoded by the SPIF library was unexpected significantly enhanced compared to the SPVF library. All other libraries having different randomized positions than SPIF showed significantly lower levels of solubility of the proteins encoded by SPVF library, SPGO library, SPOO library, or SPKK library than those encoded by the SPIF library. "Enhanced solubility" may mean a higher proportion of a protein solute in a solvent as compared to another protein. It may also mean decreased protein aggregation or lower proportion of aggregated protein at a given protein concentration in a solvent, as compared to other proteins. Surprisingly, the expression level of all proteins encoded by the SPIF library is improved compared to the SPVF library, SPGO library, SPOO library, or SPKK library. All other libraries tested showed lower expression rates of encoded proteins (see FIG. 2 and Table 3).

Table 3: Comparison of solubility and expression level of different di-ubiquitin based libraries

solubility and expression level	SPIF library	SPVF library	SPGO library	SPOO library	SPKK library
at least 70 % soluble	50	13	3	2	5
at least 30 % soluble	24	14	2	1	7
maximum 30 % soluble	25	61	35	9	39

not soluble	1	12	60	88	49
Expression strong	90	84	65	22	42
Expression middle	8	10	25	45	34
Expression weak	1	5	6	27	16
no expression	1	1	4	6	8

Example 3. Expression and purification of ubiquitin muteins

Affilin molecules were subcloned to an expression vector using standard methods known to a skilled person, purified and analyzed as described below. All Affilin proteins were expressed and highly purified by affinity chromatography and gel filtration. After affinity chromatography purification a size exclusion chromatography (SE HPLC or SEC) has been performed using an Äkta system and a Superdex™ 200 HiLoad 16/600 column (GE Healthcare). The column has a volume of 120 ml and was equilibrated with 2 CV. The samples were applied with a flow rate of 1 ml/min purification buffer B. Fraction collection starts as the signal intensity reaches 10 mAU. Following SDS-PAGE analysis positive fractions were pooled and their protein concentrations were measured. Further analysis included SDS-PAGE, SE-HPLC and RP-HPLC. Protein concentrations were determined by absorbance measurement at 280 nm using the molar absorbent coefficient. RP chromatography (RP HPLC) has been performed using a Dionex HPLC system and a Vydac 214MS54 C4 (4,6 x 250 mm, 5µm, 300 Å) column (GE Healthcare).

Example 4. Solubility analysis of binding proteins

Supernatants and resuspended pellets were analyzed by NuPage Novex 4-12 % Bis-Tris SDS gels and stained with Coomassie. Proteins were recovered from the pellets by addition of 8 M urea. All Affilin proteins tested displayed a high solubility of at least 80 % solubility.

Example 5. Binding proteins are stable at high temperatures

Thermal stability of the binding proteins of the invention was determined by Differential Scanning Fluorimetry (DSF). Each probe was transferred at concentrations of 0,1 µg/µL to a MicroAmp® Optical 384-well plate well plate, and SYPRO Orange dye was added at suitable dilution. A temperature ramp from 25 to 95°C was programmed with a heating rate of 1°C per minute (ViiA-7 Applied Biosystems). Fluorescence was constantly measured at an excitation wavelength of 520 nm and the emission wavelength at 623 nm (ViiA-7, Applied Biosystems). The midpoints of transition for the thermal unfolding (T_m, melting points) are shown for selected variants in Table 4.

Example 6. Analysis of affinity to targets

A CM5 sensor chip (GE Healthcare) was equilibrated with Surface Plasmon Resonance (SPR) running buffer. Surface-exposed carboxylic groups were activated by passing a mixture of EDC and NHS to yield reactive ester groups. 700-1500 RU "target"-Fc (*on*-target) were immobilized on a flow cell, IgG-Fc (*off*-target) was immobilized on another flow cell at a ratio of 1:3 (hIgG-Fc:Target) to the target. Injection of ethanolamine after target immobilization removes non-covalently bound target. Upon target binding, protein analyte was accumulated on the surface increasing the refractive index. This change in the refractive index was measured in real time and plotted as response or resonance units (RU) versus time. The analytes were applied to the chip in serial dilutions with a flow rate of 30 µl/min. The association was performed for 30 seconds and the dissociation for 60 seconds. After each run, the chip surface was regenerated with 30 µl regeneration buffer and equilibrated with running

buffer. A dilution series of Trastuzumab served as positive control, whereas a dilution series of unmodified ubiquitin represents the negative control. The control samples were applied to the matrix with a flow rate of 30 $\mu\text{l}/\text{min}$, while they associate for 60 seconds and dissociate for 120 seconds. Regeneration and re-equilibration were performed as previously mentioned. Binding studies were carried out by the use of the Biacore® 3000 (GE Healthcare); data evaluation was operated via the BIAevaluation 3.0 software, provided by the manufacturer, by the use of the Langmuir 1:1 model ($RI=0$). Results of binding to specific targets are shown in FIG. 4 and in Table 4. Evaluated dissociation constants (K_D) were standardized against *off*-target and indicated.

Table 4 shows biochemical characterization of binding Affilin molecules, in particular binding affinities (K_D) for the binding proteins, for example, Her2, EpCAM, EGFR, and CD3, as obtained from SPR (Affinity to target in nM; fourth column of the table) and temperature stability is shown (T_m in $^{\circ}\text{C}$; fifth column of the table).

SEQ ID	Affilin-	Target specificity	Affinity to target (nM)	T_m ($^{\circ}\text{C}$)
5	142437	Her2	29,6	62,7
6	142672	Her2	4,44	68,7
7	144637	Her2	0,35	53,66
8	144636	Her2	0,49	55,12
9	144635	Her2	0,77	46,16
10	144634	Her2	0,72	46,19
11	144633	Her2	0,38	56,25
12	144632	Her2	0,87	60,33
13	144631	Her2	0,64	49,73
14	142679	Her2	1,27	64,3
15	142658	Her2	1,97	65,9
16	142654	Her2	0,65	68,5
17	142653	Her2	0,75	47,6
18	142645	Her2	8,11	48,7
19	142628	Her2	0,39	62,4
20	142620	Her2	0,72	66
21	142618	Her2	75,7	67,4
22	142617	Her2	1,06	43,99
23	142609	Her2	0,91	74,3
24	142451	Her2	1,16	42
25	142441	Her2	3,38	68,5
26	142439	Her2	4,17	62,4
27	141931	Her2	74,2	67,4
28	141926	Her2	8,9	73,8
29	141869	Her2	44	74,3
30	141890	Her2	25,5	64,8
31	141912	Her2	100	72,93
32	141935	Her2	88,9	64,4
33	142465	Her2	21,6	82

34	142655	Her2	510	66,6
35	142418	Her2	1,66	69,2
36	142627	Her2	9,73	N.D.
37	141975	Her2	39,4	72,53
39	144160	CD3	46,5	59,55
40	143488	EpCAM	150	71,4
41	143283	EpCAM	31,8	67,5
42	142265	EGFR	620	66,5

Example 7. Functional characterization: Binding to cell surface expressed targets

Additional functional characterization was performed by target binding analysis with Her2 overexpressing cells or EGFR overexpressing cells. Different concentrations of specific Affilin molecules were tested. Her2 cell target binding or EGFR cell target binding was confirmed, as shown in the Figures. Flow cytometry was used to analyze the interaction of Her2 binding proteins or EGFR binding proteins with surface-exposed Her2. Her2 overexpressing human mammary gland adenocarcinoma-derived SkBr3 cells, Her2 overexpressing transfected CHO-K1 cells or EGFR overexpressing transfected CHO-K1 cells, Her2 non-expressing human embryonic kidney cell line HEK/293 or EGFR non-expressing cell line HEK/293 and empty vector control CHO-K1 cells were used. The anti-Her2 monoclonal antibody Trastuzumab or anti-EGFR monoclonal antibody Cetuximab were used as positive controls, respectively. Results are summarized in FIG. 5 and FIG. 6.

Cells were trypsinized and resuspended in medium containing FCS, washed and stained in pre-cooled FACS blocking buffer. A cell concentration of 2×10^6 cells/ml was prepared for cell staining and filled into a 96 well plate (Greiner) in triplicate for each cell line. Different concentrations of Affilin proteins were added to Her2 or EGFR overexpressing and control cells in several experiments as indicated in the figures. After 45 min the supernatants were removed and 100 μ l/well rabbit anti-Strep antibody (obtained from GenScript; A00626), 1:300 diluted in FACS blocking buffer were added. After removal of the primary antibody goat anti-human IgG Alexa Fluor 488 antibody (obtained from Invitrogen; A11008) was applied in a 1:1000 dilution. Flow cytometry measurement was conducted on the Guava easyCyte 5HT device from Merck-Millipore at excitation wavelength 488 nm and emission wavelength 525/30 nm. Comparable amounts of clone 139090 were used as negative control in the experiments.

Example 8. Functional characterization: Binding to cell surface expressed Her2

Binding of proteins of the invention on cells exogenously expressing human Her2 was confirmed by immunocytochemistry and fluorescence microscopy. 50 nM of Affilin-141884, Affilin-142628, Affilin-141926, Affilin-144637, Affilin-142418 and Affilin-144567 were tested on Her2-expressing SkBr3 cells and the negative control cell line HEK/293. Cells were seeded with a concentration of 1×10^5 cells/ml in Lab-Tek® Chamber-Slides (Sigma-Aldrich). After cultivation over 72h the cells were fixed with methanol (5 min., 20 °C), followed by blocking (5 % Fetal Horse Serum in PBS, 1h) and incubation with 50 nM Affilin for 45 min at rt. Affilin binding was detected by an incubation of rabbit-anti-Strep-Tag-antibody (1:500) for 1h and subsequent incubation with anti-rabbit-IgG-Alexa488-antibody (1:1000) for 1h. The nuclei were stained with 4 μ g/ml DAPI. All incubation steps were at room temperature.

Example 9. Functional characterization: Her2 binding proteins bind to extracellular Her2 expressed on tissue cells derived from tumors

Tissue sections of SKOV-3-tumor, lung, liver, heart muscle and ovary were used to analyze the binding proteins of the invention by immunohistochemistry and fluorescence microscopy. Tissue slices were fixed with ice-cold Acetone for 10 min, followed by blocking and incubation with different concentrations (20 nM and 50 nM) of Affilin-141884 and Affilin-142628 and an equal amount of di-ubiquitin clone 139090 as negative or 10 nM Trastuzumab as positive control for 1h. After washing with PBS the tissue was incubated with rabbit anti-Strep-Tag antibody (1:500) for 1h at room temperature, followed by an incubation with goat anti-rabbit Alexa488 (1:1000) or goat anti-human Alexa594 (1:1000) secondary antibody. Nuclei of cells were visualized with DAPI. Chamber slides were disassembled and the glass slides covered with Mowiol and a cover glass. Slices were imaged at a Zeiss Axio Scope.A1 microscope and images were processed using standard software packages. FIG. 7 show the specific binding of Her2 Affilin molecules on SKOV-3-tumor-slices compared to the non-binding protein clone 139090. No binding on lung, liver, heart muscle and ovarian tissues was observed.

Example 10. Binding proteins bind to extracellular target expressed on tumor cells

A431 cells are derived from epidermoid carcinoma and are known for the high expression levels of EGFR. A431 cells were seeded in Lab-Tek® Chamber Slides (Sigma-Aldrich) and incubated for 2 days at 37°C, 5 % CO₂. Cells were fixed with 3 % paraformaldehyde for 10 minutes at RT, washed with PBS and subsequently blocked with blocking solution (BS, 3 % BSA + 0.1 % Triton-X100) for 30 min at 4°C. Cells were then incubated with 500 nM binding protein Affilin-142265 or 10 nM Cetuximab as control for 1h at 4°C. Unmodified ubiquitin (referred to as clone 139090) served as negative control. After washing with PBS cells were incubated with rabbit anti-Strep-antibody (1:500) for 1h at 4°C. Cells were then incubated with donkey anti-Rabbit Alexa488 (1:1000) or goat anti-Human Alexa 594 (1:1000) secondary antibodies followed by visualization of cell nuclei with DAPI. Chamber slides were disassembled and the glass slide was covered with Mowiol and a cover glass. Cells were imaged at a Zeiss Axio Scope.A1 microscope and images were processed using standard software packages. The staining of A431 tumor cells expressing EGFR confirms binding of Affilin 142265 to extracellular EGFR.

CLAIMS

1. A method for the generation of a binding protein derived from di-ubiquitin (SEQ ID NO: 4) wherein said binding protein binds to a given target with a dissociation constant (KD) of at least 1000 nM, said method comprising
 - a. subjecting nucleotide triplets of a nucleic acid molecule of di-ubiquitin (SEQ ID NO: 4) to mutagenesis in 12, 13, or 14 positions coding for amino acid positions R42, I44, H68, V70, R72, L73, R74, K82, L84, Q138, K139, E140, S141, and T142 of the linear polypeptide sequence of di-ubiquitin (SEQ ID NO: 4),
 - b. resulting in one or more mutein nucleic acid molecule(s),
 - c. expressing the one or more mutein nucleic acid molecule(s) obtained in (b) in a suitable expression system,
 - d. enriching the one or more di-ubiquitin mutein(s) having a dissociation constant (KD) of at least 1000 nM for the given target by means of selection and/or isolation, and
 wherein said binding protein has between 85 % and 92 % sequence identity to di-ubiquitin (SEQ ID NO: 4), and wherein the affinity to the target is at least 10fold higher than for di-ubiquitin (SEQ ID NO: 4).
2. A method according to claim 1 wherein the target is a peptide, a protein or a domain of a protein.
3. A method according to claim 2 wherein the target is selected from Her2 or EGFR or CD3 or epCAM.
4. A nucleic acid library encoding muteins of di-ubiquitin (SEQ ID NO: 4), wherein the amino acid residues selected from 12, 13, or 14 positions R42, I44, H68, V70, R72, L73, R74, K82, L84, Q138, K139, E140, S141, and T142 of SEQ ID NO: 4 are substituted in the muteins.
5. A method for identifying a binding protein with the affinity to the target of at least 10fold higher than for di-ubiquitin comprising screening the library of claim 4 to identify a polypeptide that binds to the given target.
6. The method of claim 5 comprising the following steps: providing a library of claim 4 providing a potential target to said library; contacting said library with said target; and identifying a binding protein by a screening process, wherein said binding protein binds to said target with a specific binding affinity of at least 1000 nM, and optionally isolating said binding protein with said binding affinity.
7. A binding protein with an affinity to a target wherein the binding protein is comprising an amino acid sequence with substitutions in 12, 13, or 14 amino acids selected from positions R42, I44, H68, V70, R72, L73, R74, K82, L84, Q138, K139, E140, S141, and T142 of di-ubiquitin (SEQ ID NO: 4) and wherein the binding protein has between 85 % and 92 % sequence identity to di-ubiquitin (SEQ ID NO: 4) and wherein the affinity to the target is at least 10fold higher than for di-ubiquitin (SEQ ID NO: 4) .
8. The binding protein according to claim 7 comprising 1, 2, 3, 4, 5, or 6 further amino acid substitutions of di-ubiquitin.
9. The binding protein according to claim 7 or 8 having a specific binding affinity of at least 700 nM, preferably at least 100 nM, more preferably at least 10 nM to a target selected from a peptide, a protein, or a domain of a protein, preferably from Her2 or EGFR or CD3 or EpCAM.
10. The binding protein according to claim 9 wherein the amino acid sequence is selected from the amino acid sequences of SEQ ID NOs: 5-42.
11. The binding protein according to any of the preceding claims further comprising at least one additional molecule, preferably selected from at least one member of the groups (i), (ii) and (iii) consisting of (i) a

moiety modulating pharmacokinetics selected from a polyethylene glycol, a human serum albumin, an albumin-binding peptide, or an immunoglobulin or immunoglobulin fragment, a polysaccharide, and, (ii) a therapeutically active component, optionally selected from a monoclonal antibody or a fragment thereof, cytokine, a chemokine, a cytotoxic compound, an enzyme, or derivatives thereof, or a radionuclide, and (iii) a diagnostic component, optionally selected from a fluorescent compound, a photosensitizer, a tag, an enzyme, or a radionuclide.

12. A nucleic acid molecule encoding a binding protein as defined in any one of claims 7 - 11.

13. A vector comprising the nucleic acid molecule of claim 12.

14. A host cell or a non-human host comprising the binding protein as defined in any one of claims 7 - 11, a nucleic acid as defined in claim 12, or a vector of claim 13.

15. The binding protein of any one of claims 7 - 11 or the nucleic acid molecule of claim 12 or the vector of claim 13 or the host cell of claim 14 for use in medicine.

16. A composition comprising the binding protein as defined in any one of claims 7 - 11; the nucleic acid molecule as defined in claim 12; the vector as defined in claim 13; or the host cell as defined claim 14.

.....

FIG. 1

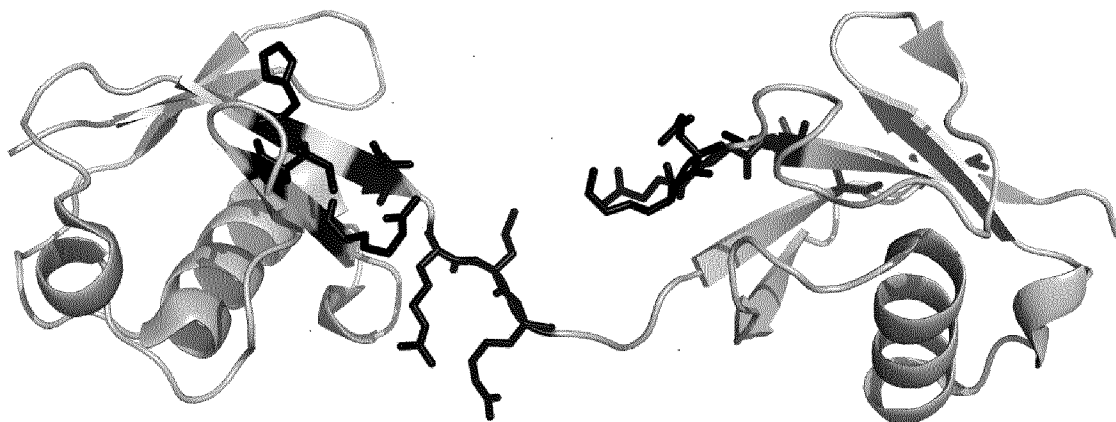


FIG. 2
FIG. 2 A

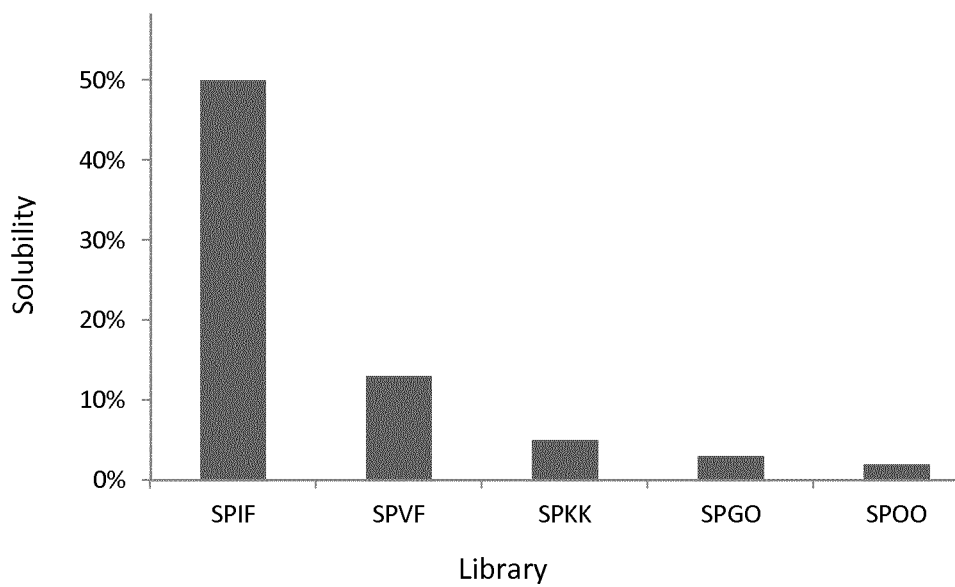


FIG. 2 B

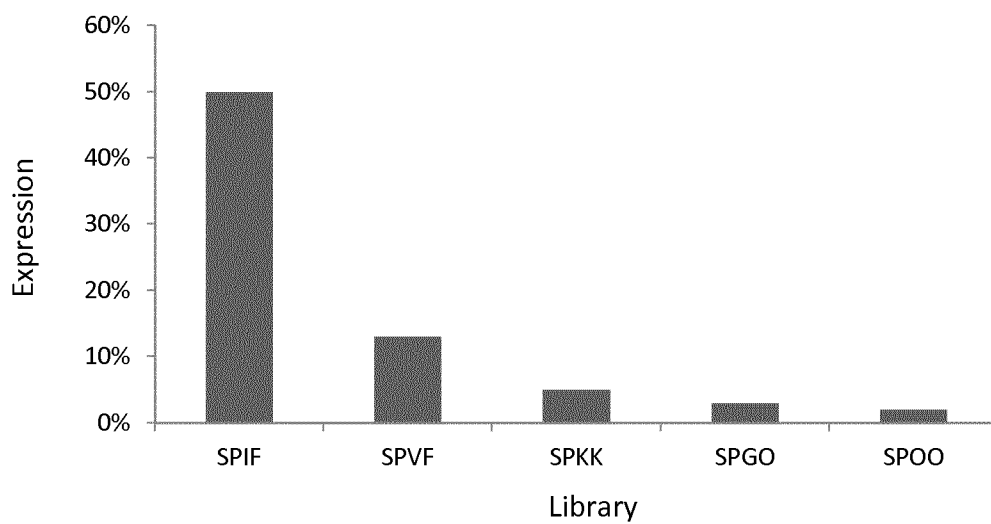


FIG. 3

FIG. 3 A

SEQ ID NO:	Affilin	R42	I44	H68	V70	R72	L73	R74	K82	L84	Q138	K139	E140	S141	T142
5	142437	T	A	W	Y	T	W	N	T	H	S	E	W	A	I
6	142672	T	A	W	Y	T	W	Y	L	D	R	G	W	.	L
7	144637	T	A	W	Y	T	W	Y	T	H	S	E	W	A	I
8	144636	T	A	W	Y	T	W	Y	T	H	S	E	W	A	I
9	144635	T	A	W	Y	T	W	Y	T	H	S	E	W	A	I
10	144634	T	A	W	Y	T	W	Y	T	H	S	E	W	A	I
11	144633	T	A	W	Y	T	W	Y	T	H	S	E	W	A	I
12	144632	T	A	W	Y	T	W	Y	T	H	S	E	W	A	I
13	144631	T	A	W	Y	T	W	Y	T	H	S	E	W	A	I
14	142679	T	A	W	Y	T	W	Y	T	H	S	E	W	A	I
15	142658	T	A	W	Y	T	W	Y	T	H	S	E	W	A	I
16	142654	T	A	W	Y	T	W	Y	T	H	S	E	W	A	I
38	141884	T	A	W	Y	T	W	Y	T	H	S	E	W	A	I
17	142653	T	A	W	Y	T	W	Y	T	H	S	E	W	A	I
18	142645	T	A	W	Y	T	W	Y	T	H	S	E	W	A	I
19	142628	T	A	W	Y	T	W	Y	T	H	S	E	W	A	I
20	142620	T	A	W	Y	T	W	Y	T	H	S	E	W	A	I
21	142618	T	A	W	Y	T	W	Y	T	H	S	E	W	A	I
22	142617	T	A	W	Y	T	W	Y	T	H	S	E	W	A	I
23	142609	T	A	W	Y	T	W	Y	T	H	S	E	W	A	I
24	142451	T	A	W	Y	T	W	Y	T	H	S	E	W	A	I
25	142441	T	A	W	Y	T	W	Y	T	H	S	E	W	A	I
26	142439	T	A	W	Y	T	W	Y	T	H	S	E	W	A	I
27	141931	T	A	W	Y	T	W	Y	T	H	S	E	W	A	I
28	141926	T	A	W	Y	T	W	Y	T	H	S	E	W	A	I
29	141869	T	A	W	Y	T	W	Y	T	H	S	E	W	A	I
30	141890	T	A	Y	Y	T	V	K	Y	H	E	L	W	R	N
31	141912	T	.	F	Y	Y	V	S	L	D	E	L	W	R	N
32	141935	T	T	Y	Y	G	I	S	I	S	E	L	W	R	N
33	142465	L	S	.	W	G	V	.	N	E	S	E	W	A	I
34	142655	S	V	Y	Y	F	S	.	L	D	R	G	W	.	L
35	142418	S	V	Y	Y	F	S	.	L	D	R	G	W	.	L
36	142627	S	V	Y	Y	F	S	.	L	D	S	E	W	A	I
37	141975	S	V	Y	Y	F	S	.	L	D	S	E	W	A	I

FIG. 3 B

SEQ ID NO:	Affilin-	42	44	68	70	72	73	74	82	84	138	139	140	141	142
5	142437	~	H	O	O	~	O	~	~	+	~	-	O	H	H
6	142672	~	H	O	O	~	O	O	H	-	+	G	O	~	H
7	144637	~	H	O	O	~	O	O	~	+	~	-	O	H	H
8	144636	~	H	O	O	~	O	O	~	+	~	-	O	H	H
9	144635	~	H	O	O	~	O	O	~	+	~	-	O	H	H
10	144634	~	H	O	O	~	O	O	~	+	~	-	O	H	H
11	144633	~	H	O	O	~	O	O	~	+	~	-	O	H	H
12	144632	~	H	O	O	~	O	O	~	+	~	-	O	H	H
13	144631	~	H	O	O	~	O	O	~	+	~	-	O	H	H
14	142679	~	H	O	O	~	O	O	~	+	~	-	O	H	H
15	142658	~	H	O	O	~	O	O	~	+	~	-	O	H	H
16	142654	~	H	O	O	~	O	O	~	+	~	-	O	H	H
38	141884	~	H	O	O	~	O	O	~	+	~	-	O	H	H
17	142653	~	H	O	O	~	O	O	~	+	~	-	O	H	H
18	142645	~	H	O	O	~	O	O	~	+	~	-	O	H	H
19	142628	~	H	O	O	~	O	O	~	+	~	-	O	H	H
20	142620	~	H	O	O	~	O	O	~	+	~	-	O	H	H
21	142618	~	H	O	O	~	O	O	~	+	~	-	O	H	H
22	142617	~	H	O	O	~	O	O	~	+	~	-	O	H	H
23	142609	~	H	O	O	~	O	O	~	+	~	-	O	H	H
24	142451	~	H	O	O	~	O	O	~	+	~	-	O	H	H
25	142441	~	H	O	O	~	O	O	~	+	~	-	O	H	H
26	142439	~	H	O	O	~	O	O	~	+	~	-	O	H	H
27	141931	~	H	O	O	~	O	O	~	+	~	-	O	H	H
28	141926	~	H	O	O	~	O	O	~	+	~	-	O	H	H
29	141869	~	H	O	O	~	O	O	~	+	~	-	O	H	H
30	141890	~	H	O	O	~	H	+	O	+	-	H	O	+	~
31	141912	~	H	O	O	O	H	~	H	-	-	H	O	+	~
32	141935	~	~	O	O	G	H	~	H	~	-	H	O	+	~
33	142465	H	~	+	O	G	H	+	~	-	~	-	O	H	H
34	142655	~	H	O	O	O	~	+	H	-	+	G	O	~	H
35	142418	~	H	O	O	O	~	+	H	-	+	G	O	~	H
36	142627	~	H	O	O	O	~	+	H	-	~	-	O	H	H
37	141975	~	H	O	O	O	~	+	H	-	~	-	O	H	H

FIG. 3 C

		K	G	K	L	S	D	I	K	A	Q	E	I	Q	A	K	Q	D	D	Q	K	A	Q	T	E	P	S	T	P	I	K	T	S	H	L
SEQ	Att(j)	6	10	11	15	20	21	23	27	28	31	34	36	40	46	48	49	52	58	62	63	75	78	88	92	95	96	98	114	120	124	131	133	144	147
5	142437	.	.	.	.	.	N	.	.	.	.	.	T	.	.	.	H	.	.	.	.	.	R	N	.	.	.	.	.	.	.	.	.	.	.
6	142672	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	R	.	.	.	.	.	.	.	.	.	.	.	.
7	144637	.	.	.	.	.	.	T	.	.	.	.	.	V	.	.	Y	.	.	.	.	.	R	.	D	.	.	.	.	.	.	.	.	.	.
8	144636	.	.	.	.	.	.	T	.	.	.	.	.	V	.	.	Y	.	.	.	.	.	R	.	.	.	.	.	.	.	.	.	.	.	.
9	144635	.	.	.	.	.	.	.	.	.	.	.	.	V	M	.	.	.	.	.	.	.	R	.	D	.	.	.	.	.	M	.	.	.	.
10	144634	.	.	.	.	.	.	.	.	.	.	.	.	V	.	.	.	.	.	.	.	.	R	.	D	.	.	.	.	M	.	.	.	.	.
11	144633	.	.	.	.	.	.	.	.	.	.	.	.	V	.	.	Y	.	.	.	.	.	R	.	D	.	.	.	.	.	.	.	.	.	.
12	144632	.	.	.	.	.	.	T	.	.	.	.	.	V	.	.	.	.	.	.	.	.	R	.	.	.	.	.	.	.	.	.	.	.	.
13	144631	.	.	.	.	.	.	.	.	.	.	.	.	V	.	.	.	.	.	.	.	.	R	.	D	.	.	.	.	.	.	.	.	.	.
14	142679	R	.	.	.	.	.	.	.	.	.	.	.	V	.	.	.	.	.	.	.	.	R	.	.	.	.	.	.	.	.	.	.	.	.
15	142658	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	R	.	.	.	.	.	.	.	.	.	.	Q	.
16	142654	.	.	.	.	.	.	.	.	.	.	.	.	V	.	.	.	.	.	.	.	.	R	.	.	.	.	.	.	.	.	.	.	.	.
38	141884	.	.	.	.	.	.	.	R	.	.	.	.	.	.	.	.	.	.	.	.	.	R	.	.	.	.	.	.	.	.	.	.	.	.
17	142653	.	.	.	.	.	.	.	.	.	.	.	.	V	.	.	.	.	.	.	.	.	R	.	.	.	S	.	.	.	.	.	.	.	.
18	142645	.	D	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	R	.	.	.	.	.	.	.	.	T	.	.	.
19	142628	.	.	.	.	.	.	T	.	.	.	.	.	V	.	.	.	.	.	.	.	.	R	.	D	.	.	.	.	.	.	.	.	.	.
20	142620	.	.	.	.	.	.	.	.	V	.	.	.	V	.	.	.	.	.	.	.	.	R	.	.	.	.	.	.	.	.	.	.	.	.
21	142618	.	.	.	.	.	.	.	.	.	.	.	.	K	V	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
22	142617	.	.	.	.	.	.	.	.	.	.	.	.	V	.	.	.	.	.	.	.	.	R	.	.	.	.	S	.	.	A	.	.	.	.
23	142609	.	.	N	.	.	.	.	.	.	.	.	.	V	.	.	.	.	.	.	.	.	R	.	.	.	.	.	.	.	.	.	.	.	P
24	142451	.	.	.	.	.	.	.	.	.	.	.	.	V	.	.	.	.	.	.	.	.	R	.	.	.	T	.	.	N	.	.	.	.	.
25	142441	.	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	R	.	.	.	.	.	.	.	.	.	.	.	.
26	142439	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	Y	.	.	.	.	.	R	.	.	.	.	.	.	.	.	.	.	.	.
27	141931	.	.	.	.	.	.	.	.	.	.	.	.	V	.	.	.	.	.	.	.	.	R	.	.	.	.	.	.	.	.	.	.	.	.
28	141926	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	R	.	.	.	.	.	.	.	.	.	.	.	.
29	141869	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
30	141890	R	.	.	.	.	.	.	.	.	.	.	.	V	.	.	.	.	.	R	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
31	141912	.	.	.	.	.	.	.	.	.	.	.	.	V	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
32	141935	.	.	.	.	.	.	.	.	.	.	D	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
33	142465	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
34	142655	.	.	.	.	.	.	.	.	L	.	.	.	.	.	.	.	V	.	.	.	.	R	.	.	S	.	.	.	.	.	.	.	.	.
35	142418	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	I	.	R	.	.	.	.	.	.	.	.	.	.	.	.
36	142627	.	.	.	V	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	R	.	.	.	.	.	.	.	.	.	.	.	.
37	141975	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	Y	.	.	.	V	R	.	.	.	N	.	S	.	.	A	.	.	.	.

FIG. 3 D

		+	G	+	H	~	H	+	H	~	H	~	H	+	~	~	+	H	~	~	P	~	~	P	H	+	~	~	+	H
No.	Affix	6	10	11	15	20	23	27	28	31	36	40	46	48	49	62	63	75	78	88	95	96	98	114	120	124	131	133	144	147
5	142437	.	.	.	.	.	.	.	.	.	~	.	.	.	+	.	.	.	+	~	.	.	.	.	.	.	.	.	.	.
6	142672	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	+	.	.	.	.	.	.	.	.	.	.
7	144637	.	.	.	.	.	~	.	.	.	.	.	H	.	.	.	.	.	+	.	.	.	.	.	.	.	.	.	.	.
8	144636	.	.	.	.	.	~	.	.	.	.	.	H	.	.	.	.	.	+	.	.	.	.	.	.	.	.	.	.	.
9	144635	.	.	.	.	.	.	.	.	.	.	.	H	H	.	.	.	.	+	.	.	.	.	.	.	H	.	.	.	.
10	144634	.	.	.	.	.	.	.	.	.	.	.	H	.	.	.	.	.	+	.	.	.	.	.	.	H	.	.	.	.
11	144633	.	.	.	.	.	.	.	.	.	.	.	H	.	.	.	.	.	+	.	.	.	.	.	.	.	.	.	.	.
12	144632	.	.	.	.	.	~	.	.	.	.	.	H	.	.	.	.	.	+	.	.	.	.	.	.	.	.	.	.	.
13	144631	.	.	.	.	.	.	.	.	.	.	.	H	.	.	.	.	.	+	.	.	.	.	.	.	.	.	.	.	.
14	142679	+	.	.	.	.	.	.	.	.	.	.	H	.	.	.	.	.	+	.	.	.	.	.	.	.	.	.	.	.
15	142658	.	.	.	.	.	.	.	.	.	.	.	~	.	.	.	.	.	+	.	.	.	.	.	.	.	.	.	~	.
16	142654	.	.	.	.	.	.	.	.	.	.	.	H	.	.	.	.	.	+	.	.	.	.	.	.	.	.	.	.	.
38	141884	.	.	.	.	.	.	+	.	.	.	.	.	.	.	.	.	.	+	.	.	.	.	.	.	.	.	.	.	.
17	142653	.	.	.	.	.	.	.	.	.	.	.	H	.	.	.	.	.	+	.	.	.	~	.	.	.	.	.	.	.
18	142645	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	+	.	.	.	.	.	.	.	.	~	.	.
19	142628	.	.	.	.	.	~	.	.	.	.	.	H	.	.	.	.	.	+	.	.	.	.	.	.	.	.	.	.	.
20	142620	.	.	.	.	.	.	.	H	.	.	.	H	.	.	.	.	.	+	.	.	.	.	.	.	.	.	.	.	.
21	142618	.	.	.	.	.	.	.	.	.	.	+	H	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
22	142617	.	.	.	.	.	.	.	.	.	.	.	H	.	.	.	.	.	+	.	.	.	.	~	.	.	H	.	.	.
23	142609	.	.	~	.	.	.	.	.	.	.	.	H	.	.	.	.	.	+	.	.	.	.	.	.	.	.	.	.	P
24	142451	.	.	.	.	.	.	.	.	.	.	.	H	.	.	.	.	.	+	.	.	~	.	.	~	.	.	.	.	.
25	142441	.	~	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	+	.	.	.	.	.	.	.	.	.	.	.
26	142439	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	+	.	.	.	.	.	.	.	.	.	.	.
27	141931	.	.	.	.	.	.	.	.	.	.	.	H	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
28	141926	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	+	.	.	.	.	.	.	.	.	.	.	.
29	141869	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
30	141890	+	.	.	.	.	.	.	.	.	.	.	H	.	.	+	.	.	.	.	.	.	.	.	.	.	.	.	.	.
31	141912	.	.	.	.	.	.	.	.	.	H	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
32	141935	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
33	142465	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
34	142655	.	.	.	.	.	.	.	.	H	.	.	.	.	.	.	.	.	+	.	~	.	.	.	.	.	.	.	.	.
35	142418	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	H	.	+	.	.	.	.	.	.	.	.	.	.	.
36	142627	.	.	.	H	C	.	.	.	.	.	.	.	.	.	.	.	.	+	.	.	.	.	.	.	.	.	.	.	.
37	141975	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	H	+	.	.	~	.	~	.	.	H	.	.	.

FIG. 4
FIG. 4 A

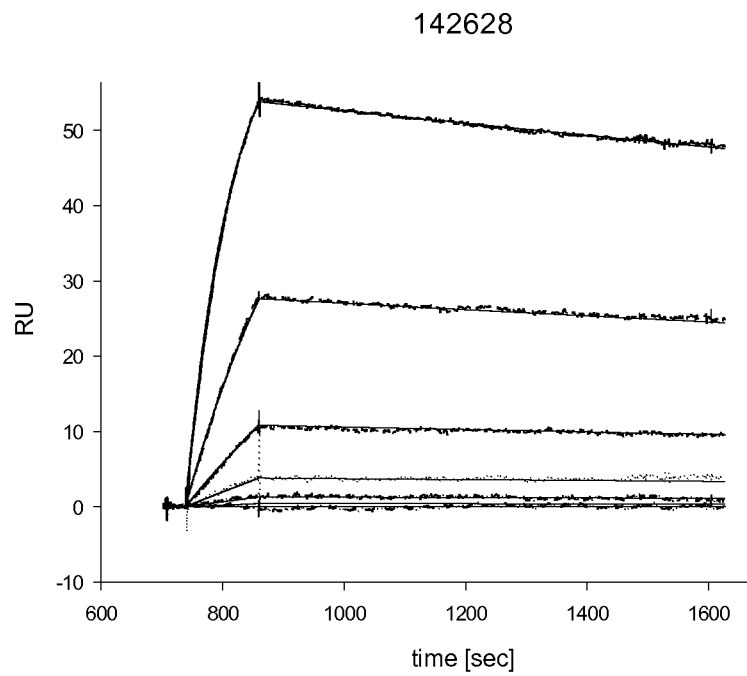


FIG. 4 B

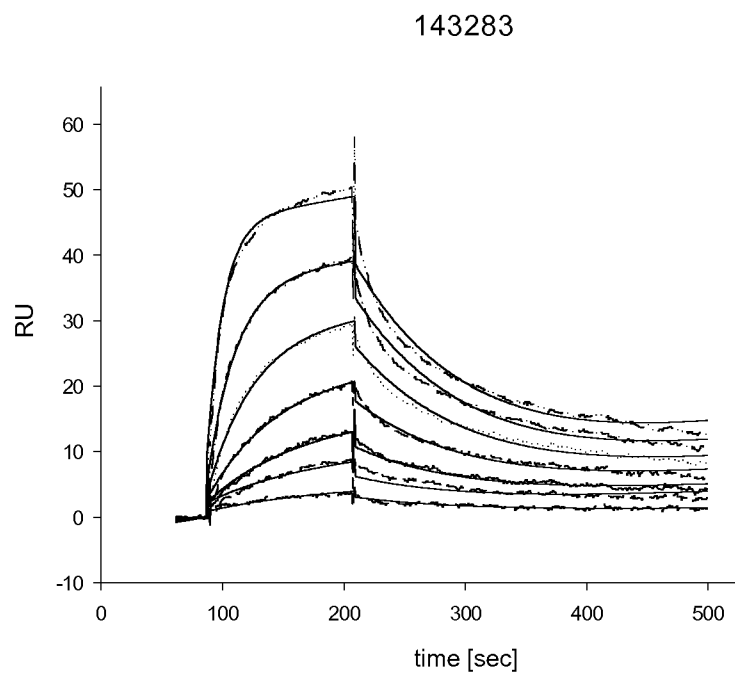


FIG. 4 C

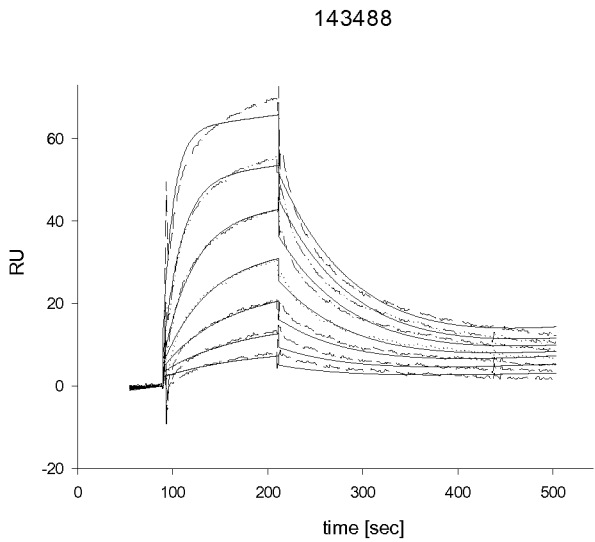


FIG. 4 D

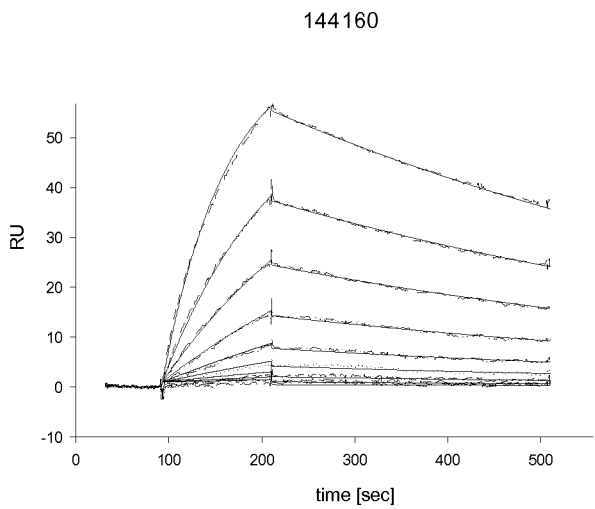


FIG. 5

FIG. 5 A

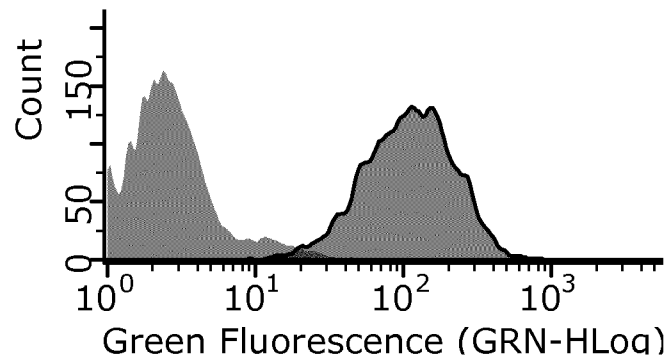


FIG. 5 B

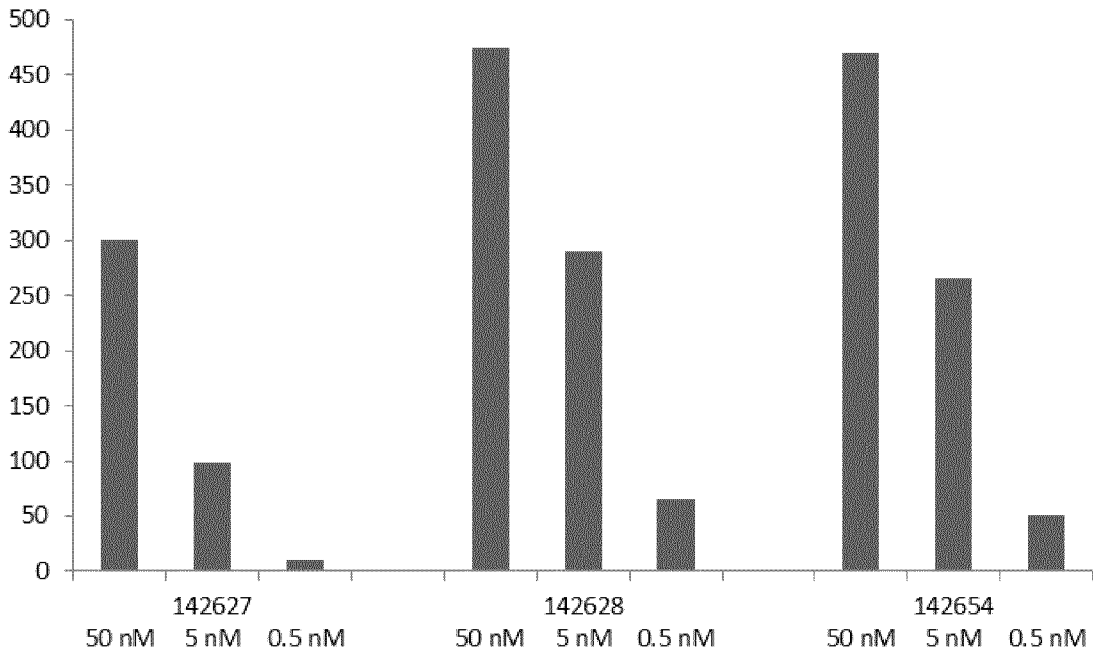


FIG. 5C

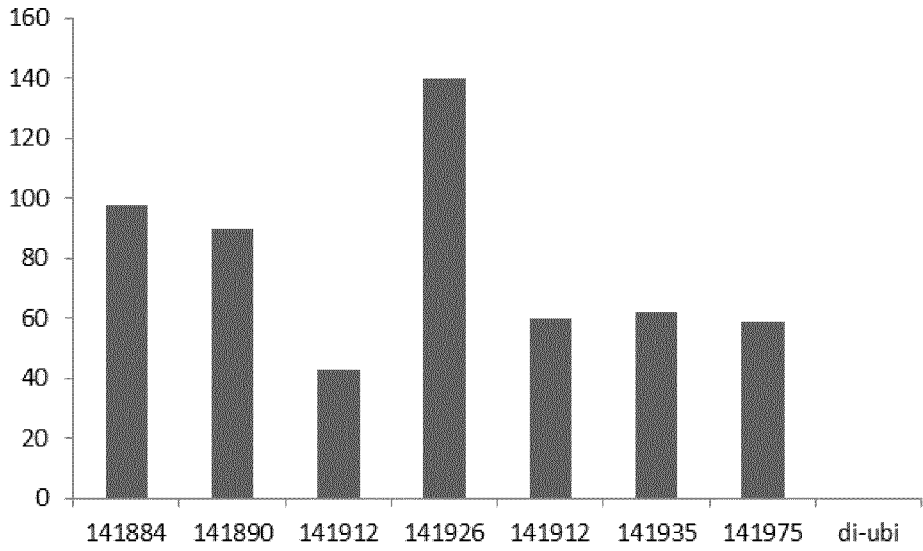


FIG. 6

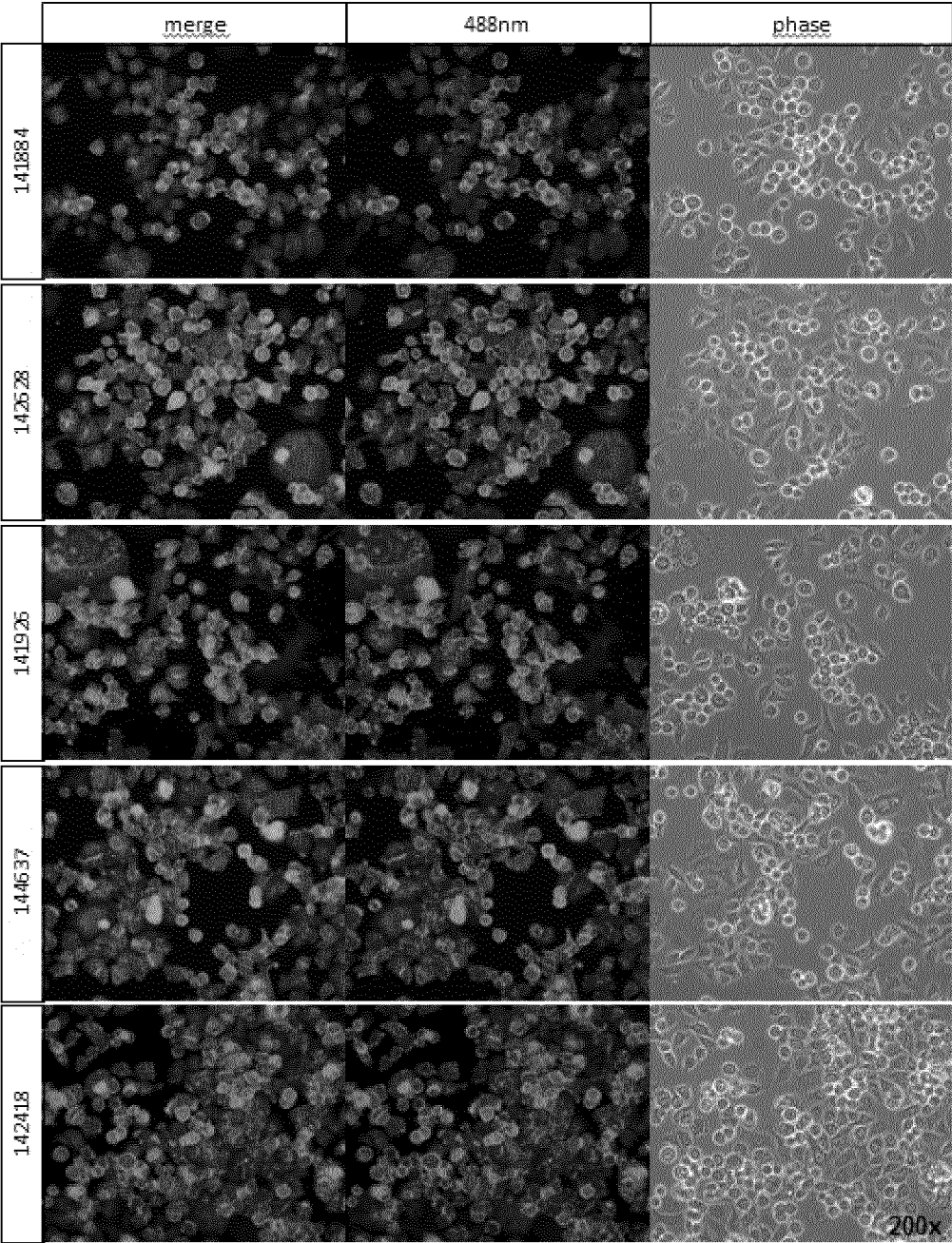


FIG. 7

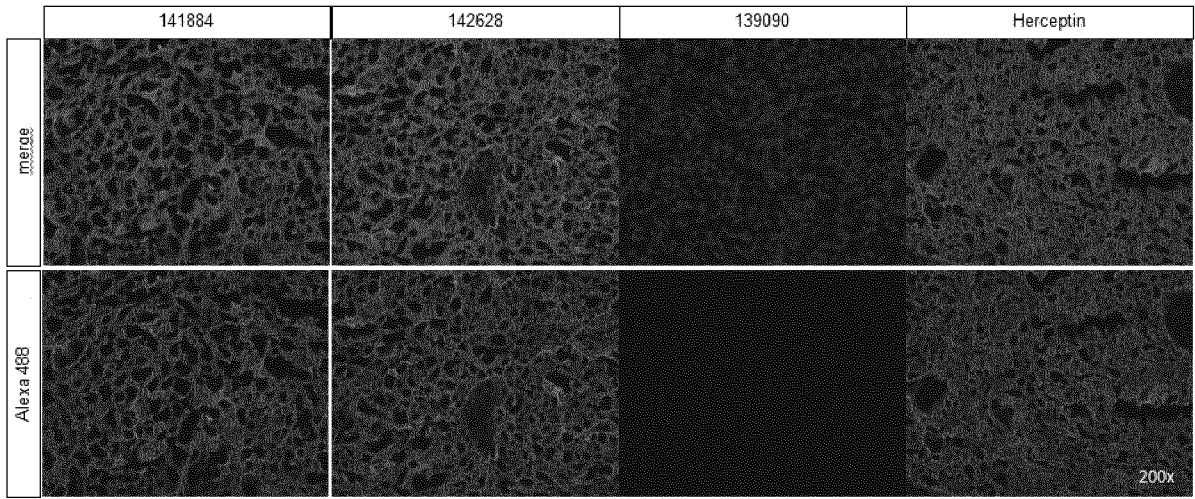
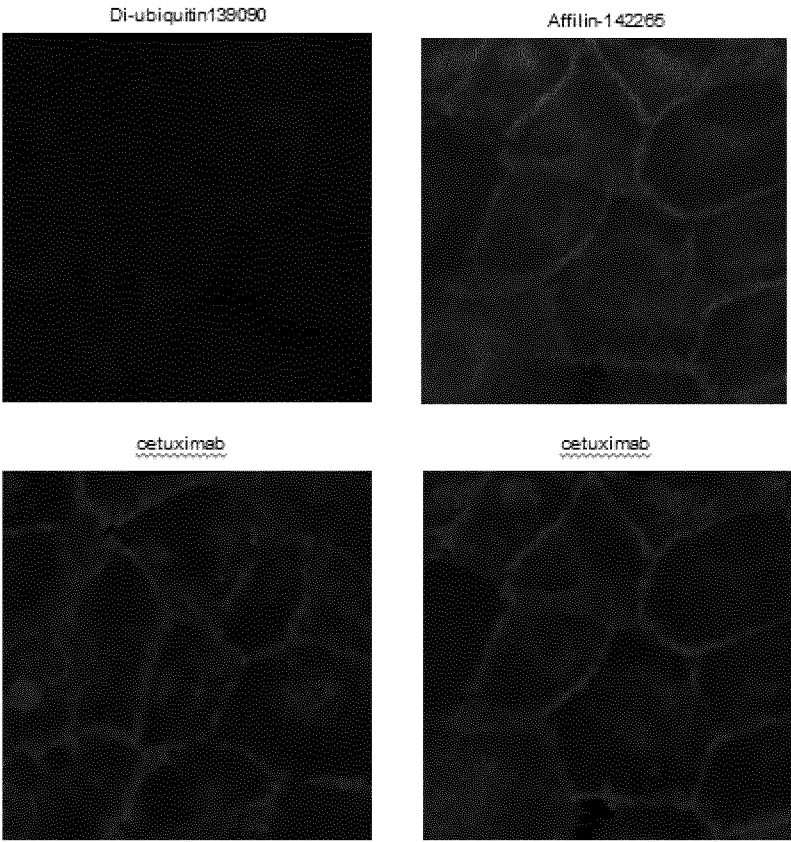


Fig. 8



INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2016/067216

A. CLASSIFICATION OF SUBJECT MATTER
INV. C07K16/32 C07K14/47
ADD. C12N15/62

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
C07K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EPO-Internal, BIOSIS, Sequence Search, EMBASE, WPI Data, CHEM ABS Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	EP 2 738 180 A1 (MOLECULAR PARTNERS AG [CH]) 4 June 2014 (2014-06-04) paragraph [0018] - paragraph [0095] -----	1-16
X	WO 2014/094799 A1 (SCIL PROTEINS GMBH [DE]) 26 June 2014 (2014-06-26) pages 2-8; claims 1-19; sequence 3 ----- -/--	1-16



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

26 September 2016

Date of mailing of the international search report

12/10/2016

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040,
Fax: (+31-70) 340-3016

Authorized officer

Seranski, Peter

INTERNATIONAL SEARCH REPORT

International application No

PCT/EP2016/067216

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	LOREY S ET AL: "Novel ubiquitin-derived high affinity binding proteins with tumor targeting properties", JOURNAL OF BIOLOGICAL CHEMISTRY, AMERICAN SOCIETY FOR BIOCHEMISTRY AND MOLECULAR BIOLOGY, US, vol. 289, no. 12, 21 March 2014 (2014-03-21), pages 8493-8507, XP002742090, ISSN: 0021-9258, DOI: 10.1074/JBC.M113.519884 [retrieved on 2014-01-28] the whole document -----	1-16

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/EP2016/067216

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP 2738180	A1	04-06-2014	
		AU 2013351096 A1	09-07-2015
		CA 2892747 A1	05-06-2014
		CN 104918959 A	16-09-2015
		EP 2738180 A1	04-06-2014
		EP 2925784 A1	07-10-2015
		HK 1211297 A1	20-05-2016
		JP 2015537044 A	24-12-2015
		KR 20150091138 A	07-08-2015
		US 2015299265 A1	22-10-2015
		WO 2014083208 A1	05-06-2014

WO 2014094799	A1	26-06-2014	NONE
